**Chapter - 4**

# CONTINUOUS RANDOM VARIABLES AND PROBABILITY DENSITIES

## Study Strategy and Learning Objectives

Please Remember The Following *study Strategy* and *Learning Objectives*:

### Study Strategy:

1. First, read this section with the limited objective of simply trying to understand what the key terms: *probability density function, probability distribution, cumulative distribution function, mean and variance of continuous random variables, continues probability distributions.*
2. Second, try to understand what they accomplish, and why they are needed and develop the ability to calculate them by selecting proper distribution.
3. Third, learn how to interpret them.
4. Fourth, read the section once again and try to understand the underlying theory.

You will always enjoy much greater success if you understand what you are doing, instead of blindly applying mechanical steps in order to obtain an answer that may or may not make any sense.

### Learning Objectives:

After careful study of this chapter, you should be able to do the following:

1. Determine *probabilities* from probability density functions.
2. Determine *probabilities* from cumulative distribution functions and *cumulative distribution functions* from probability density functions, and the reverse.
3. Calculate *means* and *variances* for continuous random variables.
4. Understand the *assumptions* for each of the continuous probability distributions presented.
5. Select an *appropriate* continuous probability distribution to calculate probabilities in specific applications.
6. Calculate *probabilities*, determine *means* and *variances* for each of the continuous probability distributions presented.
7. *Standardize* normal random variables.

## 4.1 Introduction

Many random variables observed in real life are not discrete random variables because the number of values they can assume is not countable. In contrast to discrete random variables, these variables can take on any value within an interval.

Continuous sample spaces and continuous random variables arise when we deal with quantities that are measured on a continuous scale-for instance, when we measure the speed of a car, the amount of alcohol in a person's blood, the efficiency of a solar collector, the tensile strength of a new alloy, the hardness of steel, the loss of heat in a heating system, the daily rainfall at some location, the strength of a steel bar and the intensity of sunlight at a particular time of day.

In these types of experiments, the measurement of interest can be represented by a random variable. It is reasonable to model the range of possible values of the random variable by an interval (finite or infinite) of real numbers. For example, for the length of a machined part, our model enables the measurement from the experiment to result in any value within an interval of real numbers. Because the

$$= 0 + \int_2^\infty \frac{8}{x^3}\, dx = 8\left[\frac{-1}{2x^2}\right]_2^\infty = 4\left[\frac{-1}{\infty} + \frac{1}{4}\right] = 4\left[0 + \frac{1}{4}\right] = 1$$

Since, $f(x) = \begin{cases} 8/x^3 > 0 & \text{for } x > 2 \\ 0 & \text{for } x \le 2 \end{cases}$

So, $f(x) \ge 0$ for all $x$, and $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

Hence $f(x)$ is *pdf* of rv $X$. Also, $F(x)$ exists for all $x$. Hence $F(x)$ is *cdf* of random variable $X$

Again,

(a) $P(X < 3) = F(3) = \int_{-\infty}^{3} f(x)\, dx = \int_{-\infty}^{2} f(x)\, dx + \int_2^3 f(x)\, dx = 0 + \int_2^3 \frac{8}{x^3} dx = 0.5556$

(b) $P(4 < X < 5) = \int_4^5 f(x)\, dx = \left[\frac{-4}{x^2}\right]_4^5 = \frac{9}{100} = 0.09$

(c) $\mu = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{-\infty}^{2} 0\, dx + \int_2^\infty \frac{8}{x^2} dx = \left[\frac{-8}{x}\right]_2^\infty = \left[0 + \frac{8}{2}\right] = 4.$

$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2 = \int_2^\infty x^2 \times \frac{8}{x^3} dx - 16 = [8 \log x]_2^\infty - 16$ does not exist.

**Example 3:** Determine the value of constant $k$, such that the function $f(x)$ defined by

$f(x) = \begin{cases} k\, x(1-x) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$ is probability density function of some distribution. Also, construct the distribution function $F(x)$ and hence evaluate $P(X > 1/2)$.

[TU, BE, 2061 Ashwin]

**Solution:** Since $f(x)$ is *pdf* of a continuous random variable $X$,

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \Rightarrow \int_{-\infty}^{0} f(x)\, dx + \int_0^1 f(x)\, dx + \int_1^\infty f(x)\, dx = 1$$

or, $0 + \int_0^1 kx(1-x)\, dx + 0 = 1 \quad \Rightarrow k = 6$

For distribution function $F(x) = \int_{-\infty}^{x} f(x)\, dx$;

$\int_{-\infty}^{0} f(x)\, dx = 0$ for $x \le 0$, $\int_1^\infty f(x)\, dx = 0$ for $x \ge 1$.

$\int_0^x f(x)\, dx = \int_0^x 6x(1-x)\, dx = x^2 (3 - 2x)$ for $0 < x < 1$.

The required distribution function is

$$F(x) = \begin{cases} \int_0^x 6x(1-x)\, dx = x^2(3-2x) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Finally,

$$P(X > \tfrac{1}{2}) = \int_{1/2}^{\infty} 6x(1-x)\, dx = 1 - \int_0^{1/2} 6x(1-x)\, dx$$

$$= 1 - [3x^2 - 2x^3]_0^{1/2} = 1 - \frac{1}{2} = \frac{1}{2}$$

**Example 4:** A random variable $X$ has the density function

$$f(x) = k\frac{1}{1+x^2} \quad \text{where } -\infty < x < \infty$$

Determine $k$ and the distribution function. Evaluate the probability $P(X \ge 0)$

**Solution:** Since $f(x) \ge 0$ for all $x$, it is clear that $k \ge 0$. Also,

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \Rightarrow \int_{-\infty}^{\infty} k\frac{1}{1+x^2}\, dx = 1 \Rightarrow k\,[\tan^{-1}x]_{-\infty}^{\infty} = 1$$



Or, $k\left[\frac{\pi}{2} - \frac{-\pi}{2}\right] = 1 \Rightarrow k[\pi] = 1 \Rightarrow k = \frac{1}{\pi}$

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

Next, we find the distribution function of $X$. We see that

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\, dx = \frac{1}{\pi}\int_{-\infty}^{x} \frac{1}{1+x^2}\, dx.$$

Integrating, we get,

$$F(x) = \frac{1}{\pi} = [\tan^{-1}(x)]_{-\infty}^{x} = \frac{1}{\pi}\left[x - \left(-\frac{\pi}{2}\right)\right] = \frac{1}{\pi}\left[x + \frac{\pi}{2}\right] \text{ for all } x$$

Next, we find that

$$P(X \ge 0) = 1 - P(X < 0) = 1 - F(0) = -\frac{1}{\pi} \times \frac{\pi}{2} = 1 - \frac{1}{2} = \frac{1}{2}$$

**Example 5:**

$$\text{If, } f(x) = \begin{cases} xe^{-x^2/2} & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases}$$

(i) Show that $f(x)$ is a *pdf* of a continuous random variable $X$.

(ii) Find its distribution function.

**Solution:**

(i) By definition, $f(x) \le 0$ for all $x \in R$. Thus, condition (i) is satisfied.

Also, note that $\int_{-\infty}^{\infty} f(x)\, dx = \int_0^\infty xe^{-x^2/2}\, dx = \int_0^\infty e^{-x} dx = [-e^{-x}]_0^\infty = 1$

Thus, condition (ii) also is satisfied. $f(x)$ is a *pdf* of a continuous random variable $X$

(ii) Let $F$ be the distribution function of $X$. By the definition of $f$, it follows that

$$F(x) = 0 \text{ for } x < 0.$$

Next, if $x \ge 0$, then

$$F(x) = \int_{-\infty}^{x} f(x)\, dx = \int_{-\infty}^{x} xe^{-x^2/2}\, dx = \int_0^{x^2/2} e^{-x} dx = [e^{-x}]_0^{x^2/2} = 1 - e^{-x^2/2}$$

Hence, the distribution function of $X$ is given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x^2/2} & \text{if } x \ge 0 \end{cases}$$

**Example 6:** If the probability function of a random variable is given by

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

[TU, BE, 2064 Poush]

(a) Verify that this function is probability density.

(b) Find the probability that a random variable having this probability density will take on a value (i) between 0.2 and 0.8; (ii) between 0.6 and 1.6

(c) Find the corresponding distribution function and use it to determine the probabilities that a random variable having this distribution function will take on a value (i) greater than 1.8; (ii) between 0.4 and 1.6

(d) Find $\mu$ and $\sigma^2$ for the probability density.

**Solution:**

(a) (i) $\int_{-\infty}^{\infty} f(x)\, dx = \int_{-\infty}^{0} f(x)\, dx + \int_0^1 f(x)\, dx + \int_1^2 f(x)\, dx + \int_2^\infty f(x)\, dx$

$$= 0 + \int_0^1 x\,dx + \int_1^2 (2-x)\, dx + 0 = 1$$

(ii) Since $f(x) = x > 0$ for $0 < x < 1$;

$f(x) = 2-x > 0$ for $1 \le x < 2$;

$f(x) = 0$ for $x < 0$ and $x \ge 2$.

So, $f(x) \geq 0$ for all $x$. Hence, $f(x)$ is probability density

(b) (i) $\int_{0.2}^{0.8} f(x)\,dx = \int_{0.2}^{0.8} x\,dx = \left[\frac{x^2}{2}\right]_{0.2}^{0.8} = \frac{1}{2}[(0.8)^2 - (0.2)^2] = 0.3$

(ii) $\int_{0.6}^{1.2} f(x)\,dx = \int_{0.6}^{1} f(x)\,dx + \int_{1}^{1.6} f(x)\,dx$

$= \int_{0.6}^{1} x\,dx + \int_{1}^{1.6} (2-x)\,dx = \left[\frac{x^2}{2}\right]_{0.8}^{1} + \left[2x - \frac{x^2}{2}\right]_{1}^{1.6} = 0.74$

(c) For any $x$, where $-\infty < x \leq 0$, $F(x) = \int_{-\infty}^{0} f(x)\,dx = \int_{-\infty}^{0} 0\,dx = 0$

For any $x$, where $0 < x < 1$, $F(x) = \int_{-\infty}^{0} 0\,dx + \int_{0}^{x} x\,dx = \frac{x^2}{2}$

For any $x$, where, $1 \leq x < 2$, $F(x) = \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} x\,dx + \int_{1}^{x} (2-x)\,dx$

$= 0 + \frac{1}{2} + 2x - \frac{x^2}{2} - \frac{3}{2} = 2x - \frac{x^2}{2} - 1$

For any $x$, where $2 \leq x \leq \infty$

$F(x) = \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} x\,dx + \int_{1}^{2} (2-x)\,dx + \int_{2}^{x} 0\,dx = 1$

Hence the distribution function $F(x)$ is given by

$$F(x) = \begin{cases} 0 & \text{for } -\infty < x \leq 0 \\ x^2/2 & \text{for } 0 < x < 1 \\ 2x - (x^2/2) - 1 & \text{for } 1 \leq x < 2 \\ 1 & \text{for } 2 \leq x < \infty \end{cases}$$

(i) $P(X > 1.8) = 1 - P(X \leq 1.8) = 1 - F(1.8) = 1 - [2 \times 1.8 - (1.8)^2/2 - 1] = 0.02$

(ii) $P(0.4 \leq X \leq 1.6) = F(1.6) - F(0.4) = 2 \times 1.6 - \frac{(1.6)^2}{2} - 1 - \frac{(0.4)^2}{2} = 0.84$

(d) $\mu = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{0}^{2} x f(x)\,dx = \int_{0}^{1} x\, x\,dx + \int_{1}^{2} x(2-x)\,dx = 1$

$\sigma^2 = V(X) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2 = \int_{0}^{1} x^2 x\,dx + \int_{1}^{2} x^2(2-x)\,x - 1 = \frac{1}{6}$

**Example 7:** Suppose the *pdf* of the magnitude $X$ of a dynamic load on a bridge (in Newton's) is given by

$$f(x) = \begin{cases} 1/8 + (3/8)\,x & \text{for } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

. Find the distribution function $F(x)$ and use it to determine $P(1 \leq X \leq 1.5)$ and $P(X > 1)$

**Solution:** $F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{x} f(x)\,dx$

$= 0 + \int_{0}^{x} \left(\frac{1}{8} + \frac{3}{8}x\right) dx = \frac{x}{8} + \frac{3x^2}{16}$

Thus, $$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \frac{x}{8} + \frac{3}{16}x^2 & \text{for } 0 < x \leq 2 \\ 1 & \text{for } 2 < x \end{cases}$$

Therefore, $P(1 \leq X \leq 1.5) = F(1.5) - F(1)$

$= \left[\frac{1}{8}(1.5) + \frac{3}{16}(1.5)^2\right] - \left[\frac{1}{8}(1) + \frac{3}{16}(1)^2\right] = \frac{19}{64} = 0.2969$

**Example 8:** The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous random variable $X$ with *pdf*



$$f(x) = \begin{cases} (3/2)(1-x)^2 & 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) *cdf* $F(x)$; (b) $E(x)$; (c) $V(x)$

**Solution:** (a) The *cdf* of sales for any $X$ between 0 and 1 is

$P(X \leq x) = F(x) = \int_{0}^{x} \frac{3}{2}(1 - x^2)\,dx = \frac{3}{2}\left(x - \frac{x^3}{3}\right)$

$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{0}^{1} x \frac{3}{2}(1-x^2)\,dx = \frac{3}{2}\left[\frac{x^2}{2} - \frac{x^4}{4}\right]_0^1 = \frac{3}{8}$

$\sigma^2 = V(X) = E(X^2) - [E(X)]^2$

$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2 = \int_{0}^{1} x^2 \frac{3}{2}(1-x^2)\,dx - (3/8)^2 = 0.0594$

$\sigma_x = \sqrt{V(X)} = 0.2437$

**Example 9:** Two species are competing in a region for control of a limited amount of a certain resource. Let $X$ = the proportion of the resource controlled by species 1 and suppose $X$ has *pdf*

$$f(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

which is uniform distribution on [0, 1]. Then the species that controls the majority of this resources control the amount

$$h(X) = \max(x, 1-x) = \begin{cases} 1-x & \text{if } 0 \leq x < 1/2 \\ x & \text{if } 1/2 \leq x \leq 2 \end{cases}, \text{ find } E[h(X)].$$

**Solution:** The expected amount controlled by the species having majority control is

then $E[h(X)] = \int_{-\infty}^{\infty} \max(x, 1-x) f(x)\,dx$

$= \int_{0}^{1} \max(x, 1-x)\,dx = \int_{0}^{1/2} (1-x)\,dx + \int_{1/2}^{1} x\,dx = \frac{3}{4}$

**Example 10:** Let $X$ be a continuous random variable with *pdf* given by

$$f(x) = \begin{cases} kx & \text{for } 0 \leq x < 1 \\ k & \text{for } 1 \leq x < 2 \\ -kx + 3k & \text{for } 2 \leq x < 3 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Determine $k$; (b) Determine $F(x)$, the *cdf*; (c) If $x_1, x_2, x_3$ are three independent observations from $X$, what is the probability that exactly one of these three numbers is larger than 1.5 ?

**Solution:**

(a) Since $f(x)$ is the *pdf* of $X$ we have

$\int_{0}^{3} f(x)\,dx = \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{3} f(x)\,dx = 1$

$\Rightarrow \int_{0}^{1} kx\,dx + \int_{1}^{2} k\,dx + \int_{2}^{3} (-kx + 3k) = 1 \qquad \Rightarrow k = \frac{1}{2}$

(b) For any $x$, where $-\infty < x < 0$, $F(x) = \int_{-\infty}^{x} f(x)\,dx = 0$

For any $x$, where, $0 \leq x < 1$,

$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{x} f(x)\,dx = 0 + \int_{0}^{x} \frac{1}{2}\,dx = \frac{x^2}{4}$

For any $x$, where $1 \leq x < 2$, $F(x) = \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} \frac{1}{2}\,dx + \int_{1}^{x} \frac{1}{2}\,dx = \frac{2x-1}{4}$

For any $x$, where $2 \leq x < 3$,

$F(x) = \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} \frac{1}{2}\,dx + \int_{1}^{2} \frac{1}{2}\,dx + \int_{2}^{x} \left(-\frac{x}{2} + \frac{3}{2}\right) dx$

Scanned by CamScanner

$$= 0 + \left(1 - \frac{1}{2}\right) + \left(-\frac{x^2}{4} + \frac{3x}{2} - 2\right) = -\frac{x^2}{4} + \frac{3x}{2} - \frac{5}{4}$$

For any $x$, where $3 \le x < \infty$,

$$F(x) = \int_{-\infty}^{0} 0\,dx + \int_0^1 \frac{1}{2}\,dx + \int_1^2 \frac{1}{2}\,dx + \int_2^3 \left(-\frac{x}{2} + \frac{3}{2}\right)dx + \int_3^x 0\,dx = 1$$

Hence the distribution function is given by

$$F(x) = \begin{cases} 0 & \text{for } -\infty \le x \le 0 \\ x^2/4 & \text{for } 0 < x < 1 \\ \left(\frac{2x-1}{4}\right) & \text{for } 1 \le x < 2 \\ -x^2/4 + 3x/2 - 5/4 & \text{for } 2 \le x < 3 \\ 1 & \text{for } 3 \le x < \infty \end{cases}$$

(c) The probability that $X$ is larger than 1.5 is given by

$$P(X > 1.5) = 1 - P(X \le 1.5) = 1 - F(1.5) = 1 - \frac{2 \times 1.5 - 1}{4} = \frac{1}{2}$$

The probability that $X$ is not larger than $1.5 = P(X < 1.5) = 1 - \frac{1}{2} = \frac{1}{2}$

Hence out of the three numbers $x_1$, $x_2$ and $x_3$, the probability that exactly one is larger than 1.5 is given by: $3 \times \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{3}{8}$.

**Example 11:** The mileage (in thousands of mile) that car owners get with a certain kind of tire is a random variable having probability density

$$f(x) = \begin{cases} \frac{1}{20} e^{-x/20} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

Find the probabilities that one of these lives will last

(a) at most 10,000 miles; (b) anywhere from 16,000 to 24,000 miles; (c) at least 30,000 miles. Also find $\mu$.

**Solution:** Let $X$ denote the miles (in 1000 miles) with a certain kind of tire. Then,

(a) $P(X \le 10) = \int_0^{10} f(x)\,dx = \int_0^{10} \frac{1}{20} e^{-x/20}\,dx = 1 - e^{-1/2} = 1 - 0.6065 = 0.3935$

(b) $P(16 \le X \le 24) = \int_{16}^{24} f(x)\,dx = \int_{16}^{24} \frac{1}{20} e^{-x/20}\,dx = e^{-0.8} - e^{-1.2} = 0.1481$

(c) $P(X \ge 30) = \int_{30}^{\infty} \frac{1}{20} e^{-x/20}\,dx = e^{-1.5} = 0.2231$

Now, $\mu = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-\infty}^{0} x f(x)\,dx + \int_0^{\infty} x f(x)\,dx$

$$= 0 + \int_0^{\infty} x \frac{1}{20} e^{-x/20}\,dx = 20{,}000 \text{ miles.}$$

**Example 12:** Suppose the probability density function of weekly gravel sales $X$ (in tons) is $f(x) = \begin{cases} 2(1-x) & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

(a) Obtain cdf of $X$. (b) What is $P(X \le 0.5)$; (c) Using (a) Find $P(0.25 \le X \le 0.5)$

[TU, BE, 2058 Bhadra] [2064 Magh/2065 Chaitra]

**Solution:** (a) cdf of $X$ is

$$F(x) = P(X \le x) = \begin{cases} \int_0^x 2(1-x)\,dx = 2x - x^2 & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$





(b) $P(X \le 0.5) = \int_0^{0.5} (2 - 2x)\,dx = 0.75$

(c) $P(0.25 \le X \le 0.5) = F(0.5) - F(0.25) = 1 - (0.5)^2 - 0.5 + (0.25)^2 = 0.3125$

**Example 13:** The life of a laser ray device used to inspect crack in aircraft wings is given by $X$, a continuous random variable assuming all values in the range $x \ge 0$. The probability density function of the lifetime, in year is as follows:

$$f(x) = \begin{cases} \frac{1}{2} e^{-x/2} & \text{for } x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

[TU, BE 2064 Shrawan]

(a) Find the probability that life of laser ray device is between 2 and 3 years.

(b) Find the probability that life of laser ray device is less than 2.5 years.

**Solution:**

(a) $\int_2^3 f(x)\,dx = -e^{-3/2} + e^{-1} = 0.1447$

(b) $P(X < 2.5) = \int_0^{2.5} f(x)\,dx = 1 - e^{-1.25} = 0.7135$

**Example 14:** In a city, the daily consumption of electric power (in millions of kilowatt hours) is a random variable having the probability density

$$f(x) = \begin{cases} \frac{1}{9} x e^{-x/3} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

If the city's power plant has a daily capacity of 12 million of kilowatt hours, what is the probability that this power supply will be inadequate on any given day?

**Solution:** Let $X$ denote that the daily consumption of electric power. The probability that the power supply will be inadequate is

$$P(X > 12) = 1 - P(X \le 12) = 1 - \int_0^{12} \frac{1}{9} x e^{-x/3}\,dx$$

$$= 1 - \frac{1}{9}\left[\left\{\frac{xe^{-x/3}}{\frac{-1}{3}}\right\}_0^{12} - \int_0^{12}\left(\frac{dx}{dx}\frac{e^{-x/3}}{\frac{-1}{3}}\right)dx\right]$$

$$= 1 - \frac{1}{9}\left[-36e^{-4} - 0 - 9e^{-4} + 9\right] = 1 + 4e^{-4} + e^{-4} - 1 = 5e^{-4} = 0.0916$$

**Example 15:** The probability distribution of a rv $X$ is:

$$f(x) = k \sin \frac{1}{5}\pi x,\ 0 \le x \le 5.$$

(a) Determine the constant $k$; (b) obtain the median and quartiles of the distribution.

**Solution:**

(a) $k\int_0^5 \sin \frac{1}{5}\pi x\,dx = 1 \Rightarrow k = \frac{\pi}{10}$.

(b) Median $M$ is given by

$$k\int_0^M \sin \frac{1}{5}\pi x\,dx = \frac{1}{2} \Rightarrow M = \frac{5}{2}$$

The first quartile $Q_1$ is given by

$$k\int_0^{Q_1} \sin \frac{1}{5}\pi x\,dx = \frac{1}{4} \Rightarrow \cos\left(\frac{\pi Q_1}{5}\right) = \frac{1}{2} \Rightarrow \cos\left(\frac{\pi}{3}\right) \Rightarrow Q_1 = \frac{\pi}{3}\cdot\frac{5}{\pi} = \frac{5}{3}$$

The third quartile $Q_3$ is given by

$$k\int_0^{Q_3} \sin \frac{1}{5}\pi x\,dx = \frac{3}{4} \Rightarrow \cos\left(\frac{\pi Q_3}{5}\right) = -\frac{1}{2} = \cos\left(\frac{2\pi}{3}\right) \Rightarrow Q_3 = \frac{2\pi}{3}\cdot\frac{5}{\pi} = \frac{10}{3}$$

**Example 16:** Suppose that the life in hours of a certain part of radio tube is a continuous random variable $X$ with pdf given by

$$f(x)=\begin{cases}\dfrac{100}{x^2} & \text{for } x\ge 100\\ 0 & \text{eleswhere}\end{cases}$$

(a) What is the probability that all of three such tubes in a given radio set will have to be replaced during the first 150 hours of operation?

(b) What is the probability that none of the original tubes will have to be replaced during that first 150 hours of operation?

(c) What is the probability that a tube will last less than 200 hours if it known that the tube is still functioning after 150 hours of service?

**Solution:** (a) $P(X\le 150)=\int_{100}^{150}\frac{100}{x^2}\,dx=\frac{1}{3}$

By compound probability theorem, the required probability $=\frac{1}{3}\times\frac{1}{3}\times\frac{1}{3}=\frac{1}{27}$

(b) $P(X>150)=1-P(X\le 150)=1-\frac{1}{3}=\frac{2}{3}$

By compound probability theorem, the required probability $=\frac{2}{3}\times\frac{2}{3}\times\frac{2}{3}=\frac{8}{27}$

(c) The required probability $=P(X<200\,|\,X>150)$

$$=\frac{P(150<X<200)}{P(X>150)}=\frac{\int_{150}^{200}f(x)\,dx}{1-P(X\le 150)}=\frac{1}{6}\times\frac{3}{2}=0.25$$

**Example 17:** Suppose the distance time $X$ between a point target and a shot aimed at the point in a coin operated target game is continuous random variable with probability density function **[TUBE 2062 Baishakh]**

$$f(x)=\begin{cases}\dfrac{3}{4}(1-x^2) & \text{for } -1\le x\le 1\\ 0 & \text{otherwise.}\end{cases}$$

Compute (a) $P(X>0)$; (b) $P(-0.5<X<0.5)$; (c) $P(X<-0.25 \text{ or } X>0.25)$.

**Solution:** (a) $P(X>0)=\int_0^1\frac{3}{4}(1-x^2)\,dx=0.5$

(b) $P(-0.5<X<0.5)=\int_{-0.5}^{0.5}\frac{3}{4}(1-x^2)\,dx=0.6875$

(c) $P(X<-0.25 \text{ or } X>0.25)$

$=P(X<-0.25)+P(X>0.25)$ $\quad[\because X<-0.25\cap X>0.25=\phi]$

$=\frac{3}{4}\int_{-1}^{-0.25}(1-x^2)\,dx+\frac{3}{4}\int_{0.25}^{1}(1-x^2)\,dx$

$=\frac{3}{4}\left[x-\frac{x^3}{3}\right]_{-1}^{-0.25}+\frac{3}{4}\left[x-\frac{x^3}{3}\right]_{0.25}^{1}=0.84375$

**Example 18:** The amount of bread (in hundreds of pounds) that a certain bakery is able to sell in a day is found to be a numerical valued random phenomenon, with a probability function specified by the *pdf* $f(x)$, given by:

$$f(x)=\begin{cases}kx & \text{for } 0\le x<5\\ k(10-x) & \text{for } 5\le x<10\\ 0 & \text{otherwise}\end{cases}$$

(a) Find $k$;

(b) What is the probability that the number of pounds of bread that will be sold tomorrow is: (i) more than 500 pounds; (ii) less than 500 pounds?
(iii) between 250 and 750 pounds.



(c) If $A: 5<X\le 10$; $B: 0\le X\le 5$; $C: 2.5\le X\le 7.5$
Find (i) $P(A\,|\,B)$, $P(A\,|\,C)$; (ii) Are $A$ and $B$ independent events?
(iii) Are $A$ and $C$ independent events?

**Solution:** (a) $\int_{-\infty}^{\infty}f(x)\,dx=1\Rightarrow\int_0^5 kx\,dx+\int_5^{10}k(10-x)\,dx=1\Rightarrow k=\frac{1}{25}$.

(b) (i) $P(5<X\le 10)=\int_5^{10}k(10-x)\,dx=0.5$

(ii) $P(0\le X<5)=\int_0^5 k(10-x)\,dx=0.5$

or, $P(0\le X<5)=1-P(X\ge 5)=1-0.5=0.5$

(iii) $P(2.5\le X\le 7.5)=\int_{2.5}^{5}k(10-x)\,dx+\int_5^{7.5}k(10-x)\,dx=0.75$

(c) $P(A)=0.5$; $P(B)=0.5$; $P(c)=0.75$; Also, $A\cap B=\phi$, $A\cap C=5<X\le 7.5$

$P(A\cap B)=P(\phi)=0$, $P(A\cap C)=\int_5^{7.5}k(10-x)\,dx=\frac{3}{8}=0.375$

(i) $P(A\,|\,B)=\frac{P(A\cap B)}{P(B)}=0$, $P(A\,|\,C)=\frac{P(A\cap C)}{P(C)}=\frac{0.375}{0.75}=0.5$

(ii) $P(A).P(B)=\frac{1}{4}\ne P(A\cap B)\Rightarrow A$ and $B$ are not independent.

(iii) $P(A).P(C)=\frac{3}{8}=P(A\cap C)\Rightarrow A$ and $C$ are independent.

**Example 19:** A bombing plane carrying three bombs files directly above a railway track. If a bomb falls within 40 meters of track, the track will be sufficiently damaged to disrupt the traffic. With a certain bomb site the points of impact of bomb have *pdf*:

$$f(x)=\begin{cases}(100+x)/10000 & \text{for } -100\le x<0\\ (100-x)/10000 & \text{for } 0\le x<100\\ 0 & \text{elsewhere}\end{cases}$$

where $x$ represents the vertical deviation (in meters) from the aiming point, which is the track in this case. Find the distribution function. If all the three bombs are used, what is the probability that the track will be damaged?

**Solution:** If $x<-100$, $F(x)=0$.

If $-100\le x<0$, $F(x)=\int_{-100}^{x}f(x)\,dx=\int_{-100}^{x}\frac{100+x}{10000}\,dx=\frac{1}{10{,}000}\left(100x+\frac{x^2}{2}+\frac{10^4}{2}\right)$

If $0\le x\le 100$, $F(x)=\int_{-\infty}^{-100}0\,dx+\int_{-100}^{0}\frac{100+x}{10000}\,dx+\int_0^x\frac{100-x}{10000}\,dx$

$=\frac{1}{10{,}000}\left(100x-\frac{x^2}{2}+\frac{10^4}{2}\right)$

If $x>100$, $F(x)=\int_{-\infty}^{-100}0\,dx+\int_{-100}^{0}\frac{100+x}{10000}\,dx+\int_0^{100}\frac{100-x}{10000}+\int_{100}^{\infty}0\,dx=1$

Hence,

$$F(x)=\begin{cases}0 & \text{for } x<-100\\ \dfrac{1}{10{,}000}\left(100x+\dfrac{x^2}{x}+\dfrac{10^4}{2}\right) & \text{for } -100\le x<0\\ \dfrac{1}{10{,}000}\left(100x+\dfrac{x^2}{x}+\dfrac{10^4}{2}\right) & \text{for } 0\le x<100\\ 1 & \text{for } x>100\end{cases}$$

Let $X$ denote the random variable that a bomb will damage the track.

The Track will be damaged if a bomb falls within 40 meters of track. The probability that the track will be damaged by a bomb is

$$P(|X| < 40) = P(-40 < X < 40) = \int_{-40}^{0} \frac{100+x}{10000}\,dx + \int_{0}^{40} \frac{100-x}{10000}\,dx = \frac{16}{25} = 0.64$$

The probability that a bomb will not damage the track = 1 – 0.64 = 0.36

The probability that none of the three bombs will damage the track

$$= 0.36 \times 0.36 \times 0.36 = 0.046656$$

If three bombs are dropped the required probability $p$ that the track will be damaged is given by

$p = P$ [at least one of the three bombs damage the track]

$= 1 - P$ [none of the three bombs damage the track] $= 1 - 0.046656 = 0.953344$

## 4.5 Rectangular (or Uniform) Distribution

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

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

**Remarks:**

1. Parameters of uniform distribution are $a$ and $b$ with $a < b$. It is called *uniform distribution* on $(a, b)$ since it assumes a *constant* (*uniform*) *value* of all $x$ in $(a, b)$.
2. It is called *rectangular distribution* because, the curve $y = f(x)$ describes a rectangle over $x$–axis and between ordinates $x = a$ and $x = b$.
3. A uniform variate $X$ on $(a, b)$ is written as $X \sim U[a, b]$ or $X \sim R[a, b]$.

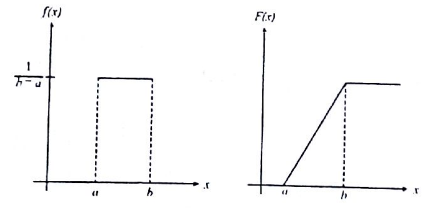



Figure: *pdf* and *cdf* of $X$

4. The *cdf* $F(x)$ is given by $F(x) = \begin{cases} 0 & \text{for } x \le a \\ \frac{x-a}{b-a} & \text{for } a < x < b \\ 1 & \text{for } b \le x \end{cases}$

Since $F(x)$ is not continuous at $a$ and $b$, it is not differentiable at these points. Then $\frac{d}{dx}F(x) = f(x) = \frac{1}{b-a} \neq 0$ exist everywhere except $a$ and $b$.

5. For a uniform variate $X$, *pdf* is : $f(x) = \begin{cases} \frac{1}{2a} & \text{for } -a < x < a \\ 0 & \text{otherwise} \end{cases}$
6. Mean and variance of a uniform rv $X$ are:





$$\mu = E(X) = \int_a^b x f(x)\,dx = \frac{1}{b-a} \times \frac{(b^2-a^2)}{2} = \frac{(b+a)}{2}$$

$$\sigma^2 = V(X) = E(X^2) - \mu^2 = \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

7. Total probability is always unity. That is, $\int_a^b f(x)\,dx = \frac{1}{b-a}\int_a^b dx = \frac{b-a}{b-a} = 1$

**Example 20:** Buses arrive at a specified stop at 15 minute intervals starting at 7 *A.M.* That is, they arrive at 7, 7.15, 7.30, 7.45 and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7.30, find the probability that he waits (a) less than 5 minutes for a bus;

(b) at least 12 minutes for a bus.

**Solution:** Let $X$ denote the time in minutes past 7 *AM* that the passenger arrives at the stop. Since $X$ is a uniform random variable over the interval (0, 30), it follows that the passenger will have to wait less than 5 minutes if he arrives between 7:10 and 7:15 or between 7:25 and 7:30. Hence, the desired probability is

(a) $P(10 < X < 15) + P(25 < X < 30) = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}$ $[\because \beta - \alpha = 30 - 0 = 30]$

Similarly, he would have to wait at least 12 minutes if he arrives between 7 and 7:03 or between 7:15 and 7:18 and so the probability is

(b) $P(0 < X < 3) + P(15 < X < 18) = \int_0^3 \frac{1}{30-0}\,dx + \int_{15}^{18} \frac{1}{30-0}\,dx = \frac{3}{30} + \frac{3}{30} = \frac{1}{5}$

**Example 21:** The current in a semiconductor diode is often measured by the Shockley equation

$$I = I_0\,(e^{aV} - 1)$$

where $V$ is the voltage across the diode; $I_0$ is the reverse current; $a$ is a constant, and I is the resulting diode current. Find $E(I)$ if $a = 5$, $I_0 = 10^{-6}$, and $V$ is uniformly distributed over (1,3).

**Solution:** $E(I) = E[I_0\,(e^{aV} - 1)] = I_0\,E[e^{aV} - 1] = 10^{-6}\,[E(e^{aV}) - 1]$

$$= 10^{-6}\int_1^3 e^{5x}\frac{1}{2}\,dx - 10^{-6} = 10^{-7}[e^{15} - e^{5}] - 10^{-6} = 0.3269.$$

**Example 22:** The buses 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 bus stop at a random time during this period will have to wait at least twenty minutes?

**Solution:** Let the rv $X$ denote the waiting time (in minutes) for the next bus. Under the assumption that a man arrives at the stop at random, $X$ is distributed uniformly on (0, 30) with *pdf* $f(x) = \begin{cases} \frac{1}{30} & \text{for } 0 < x < 30 \\ 0 & \text{otherwise} \end{cases}$

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

**Example 23:** For the rectangular distribution:

$$F(x) = \begin{cases} \frac{x}{100} & \text{for } 100 < x < 200 \\ 0 & \text{otherwise} \end{cases}$$

Compute (a) $\mu$ and $\sigma^2$; (b) $P(X \ge 150)$; (c) $P(125 \le X \le 160)$.

**Solution:** Since $F'(x)$ exists, $f(x) = \frac{dF(x)}{dx} = \begin{cases} 1/100 & \text{for } 100 < x < 200 \\ 0 & \text{otherwise} \end{cases}$

(a) $\mu = E(X) = \dfrac{100+200}{2} = 150$

$\sigma^2 = V(X) = \dfrac{(b-a)^2}{12} = \dfrac{(200-100)^2}{12} = \dfrac{2500}{3}$

(b) $P(X \geq 150) = \int_{150}^{200} f(x)\,dx = \dfrac{1}{2} = 0.5$

(c) $P(125 \leq X \leq 160) = \int_{150}^{200} f(x)\,dx = \dfrac{35}{100} = 0.35$

**Example 24:** In a certain experiment, the error made in determining the solubility of a substance is a random variable having the uniform density $\alpha = -0.025$ and $\beta = 0.025$. What are the probabilities that such an error will be (a) between 0.010 and 0.015; (b) between – 0.012 and 0.012?

**Solution:** Let the rv $X$ denote the error made in determining the solubility. Here *pdf* of $X$ is: $f(x) = \dfrac{1}{\beta - \alpha} = 20$ for $-0.025 < x < 0.025$. Therefore

(a) $P(0.010 < X < 0.015) = \int_{0.010}^{0.015} f(x)\,dx = 20\,[0.015 - 0.010] = 0.1$

(b) $P(-0.012 < X < 0.012) = \int_{-0.012}^{0.012} f(x)\,dx = 20[0.012 + 0.012] = 0.48.$

**Example 25:** In computing work, *I* must get on a bus near the house and then transfer to second bus. If the waiting time (in minutes) at bus stop has uniform distribution $A = 0$ and $B = 5$. Then it can be shown that my total waiting time $Y$ has probability density function: **[TU, BE, 2067 Mangsir]**

$$f(y) = \begin{cases} \left(\dfrac{1}{25}\right)y & \text{for } 0 < y \leq 5 \\ \left(\dfrac{2}{5}\right) - \left(\dfrac{1}{25}\right)y & \text{for } 5 < y < 10 \end{cases}$$

(a) What is the probability that waiting time is at most three minutes?

(b) What is the probability that total waiting time is at most 8 minutes?

**Solution:** Here $f(y)$ is *pdf* because

(i) $f(y) \geq 0$ for all $y \in (0, 10)$

(ii) $\int_0^{10} f(y)\,dy = \int_0^5 \left(\dfrac{1}{25}\right) y\,dy + \int_5^{10} \left[\left(\dfrac{2}{5}\right) - \left(\dfrac{1}{25}\right)y\right]dy = 1$

(a) Required probability is given by

$$P(Y \leq 3) = \int_0^3 \left(\frac{1}{25}\right) y\,dy = \frac{1}{25}\left[\frac{y^2}{2}\right]_0^3 = \frac{1}{25} \times \frac{9}{2} = 0.18.$$

(b) Required probability is given by

$$P(Y \leq 8) = \int_0^5 \left(\frac{1}{25}\right) y\,dy + \int_5^8 \left[\left(\frac{2}{5}\right) - \left(\frac{1}{25}\right)y\right] dy$$

$$= \frac{1}{25}\left[\frac{y^2}{2}\right]_0^5 + \left[\frac{2}{5}y - \frac{1}{25}\frac{y^2}{2}\right]_5^8 = 0.92.$$

## 4.6 The Normal Distribution (or Gaussian Distribution)

### 4.6.1 Introduction

The **normal probability density**, usually referred to as the **normal distribution**, is the most widely used model for the distribution of a random variable. This distribution was discovered by English Mathematician **De-Moivre** in 1733, who obtained this continuous distribution as a limiting case of binomial distribution and applied it to problems arising in the game of chance. Normal distribution is also known as ***Gaussian distribution*** (or *Gaussian law of errors*) because Gauss used this distribution to describe





the theory of accidental errors of measurements involved in the calculation of orbits of heavenly bodies.

Many numerical populations have distributions that are normally distributed either exactly or approximately. Physical measurements in areas such as temperature studies, metrological experiments, rain fall studies, product dimension in manufacturing industries, heights, weights, life–time of electric bulbs, measurement errors in scientific experiments, anthropometric measurements on fossils, reaction times in psychological experiments, measurements of intelligence and aptitude, scores on various test, and numerous economic measures and indicators are often conveniently explained with a normal distribution.

Random variables with different means and variances can be modeled by normal probability density functions with appropriate choices of the center and width of the curve. The value of $E(X) = \mu$ determines the center of the probability density function and the $V(X) = \sigma^2$ determines the width. The following dentition provides the formula for normal probability density functions.

**Definition: (normal distribution)**

A continuous random variable $X$ is said to have a ***normal distribution*** with parameters $\mu$ and $\sigma$ (or $\mu$ and $\sigma^2$) if the *pdf* of $X$ is given by

$$f(x) = f(x;\, \mu,\, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]; \quad \begin{matrix} -\infty < \pi < \infty, \\ -\infty < x < \infty, \quad \sigma > 0 \end{matrix}$$

where $e = 2.71828$, $\pi = 3.1416$, $\mu$ = mean, $\sigma$ = standard deviation.

Any continuous random variable $X$ with above *pdf* is called ***Normal variable*** and the distribution given by the same *pdf* is called the ***Normal distribution***. The graph of *pdf* $f(x;\, \mu;\, \sigma^2)$ is called ***Normal curve***.

**Note:**

1. Above *pdf* can be written as

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

2. The statement that $X$ is normally distributed with parameters $\mu$ and $\sigma^2$ is often abbreviated $X \sim N(\mu, \sigma^2)$.

3. Clearly $f(x\,;\, \mu,\, \sigma^2) \geq 0$ and

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

$$= \frac{1}{\sqrt{2\pi}}\sqrt{2\pi} = 1 \qquad \left[\text{put } \frac{x-\mu}{\sigma} = y\right]$$

Therefore, the above *pdf* is eligible.

### 4.6.2 Mean and variance of Normal Distribution

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

put $(x-\mu)/\sigma = y$, then $dx = \sigma\,dy$

$$\therefore \mu = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2}\,dy + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\,dy$$

$$= 0 + \frac{\mu}{\sqrt{2\pi}}\sqrt{2\pi} = \mu \qquad \left[\because \int_{-\infty}^{\infty} e^{-y^2/2}\,dy = \sqrt{2\pi}\,\right]$$

Similarly $V(X) = E(X^2) - [E(X)]^2 = \sigma^2.$

### 4.6.3 The Standard Normal Distribution

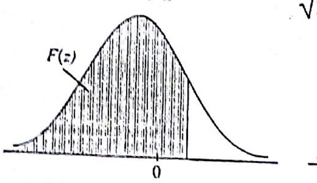
**Definitions:** The normal distribution of random variable $X$ with parameter values $\mu = 0$ and $\sigma^2 = 1$ is called the ***Standard normal distribution***. A random variable having a standard normal distribution is called a ***standard normal variate*** ($SNV$) and is denoted by

$$Z = \frac{X-\mu}{\sigma}$$

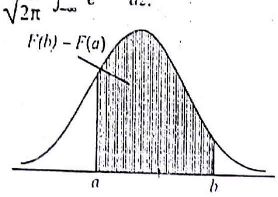
The $pdf$ of Z is $f(z) = f(z\,;0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, $-\infty < z < \infty$.

The graph of $f(z;\, 0,1)$ is called the ***standard normal curve*** (or ***Z curve***). The corresponding ***distribution function of*** $Z$ is

$$F(z) = P(Z \le z) = \int_{-\infty}^{z} f(z\,;0,1)\,dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-z^2/2}\,dz.$$




Standard Normal Probabilities
$F(z) = P(Z \le z)$




Standard Normal Probabilities
$P(a < Z \le b) = F(b) - F(a)$

**Remark:**

1. Since $f(z;\,0,1) \ge 0$ and $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-z^2/2}\,dz = 1$, the $pdf$ $f(z;\,0,1)$ is eligible.
2. We write $Z \sim N(0,\,1)$ to denote that the standard normal variate Z follows normal distribution with mean 0 and variance 1.
3. When $X \sim N(\mu,\,\sigma^2)$ then Normal probabilities are given by
$$P(a < X \le b) = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right).$$
4. $F(-z) = 1 - F(z)$

### 4.6.4 Properties of Normal Distribution

Let $X \sim N(\mu\,,\sigma^2)$. Some important properties of normal distribution are:

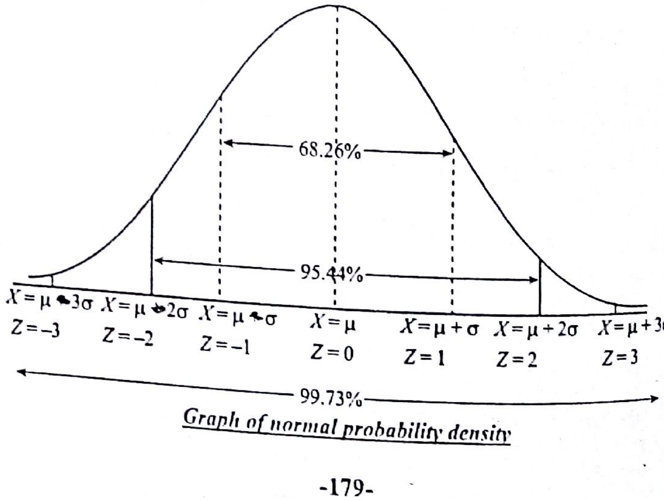
1. The normal curve is bell–shaped and symmetrical about the vertical line $X = \mu$ (i.e., $Z = 0$)
2. The mean, median and mode of normal distribution coincide. Thus Mean = Median = Mode = $\mu$
3. The total area under the normal curve within the limits $-\infty$ to $\infty$ is 1 or 100%. That is,
$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\,dx = 1$$
If $\mu = 0$, $\sigma^2 = 1$, then $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\,dx = 1$
4. The ordinate $X = \mu$ (i.e., $Z = 0$) divides the total area under the normal curve into two equal parts. So the area to the right of the ordinates as well as to the left of the ordinate at $X = \mu$ (i.e., $Z = 0$) is 0.5 i.e., 50%.
5. Since there is only one point of maximum frequency at the mean $X = \mu$, it is unimodal distribution with mode $X = \mu$. Thus distribution has the maximum probability at the mean value and the maximum value is given by





$$f_{max} = [f(x)]_{x=\mu} = \frac{1}{\sigma\sqrt{2\pi}}.$$

6. By virtue of symmetry, the quartiles are equidistant from median ($M_d$). That is, $Q_3 -$ Median = Median $- Q_1$.
7. The points of inflection (the points at which the curve changes its direction) are at $X = \mu \pm \sigma_x$. That is, they are equidistant from mean at a distance of $\sigma$.
8. The two tails of the normal curve extend indefinitely in both directions and never touch the horizontal axis. So, $x$–axis is an asymptote to the curve. It continues to approach but never touches the base line ($x$–axis).
9. Since the distribution is symmetrical, the coefficient of skewness is zero. A normal curve is mesokurtic.
10. In a normal distribution, mean deviation ($MD$) about mean is approximately equal to $\frac{4}{5}\sigma$. So, $MD \equiv \frac{4}{5}\sigma$. Also, $QD \equiv \frac{2}{3}\sigma$
Hence $Q.D = \frac{2}{3}\sigma = \frac{4}{6}\sigma = \frac{5}{6}\times\frac{4}{5}\sigma = \frac{5}{6}M.D.$
11. All central moments of odd order are zero.
12. As the value of $\sigma$ increases, the curve becomes more and more flat and vice versa.
13. A linear combination of independent normal variates is also normal variate. If $X_1, X_2, \ldots, X_n$ are independent variates with means $\mu_1, \mu_2, \ldots, \mu_n$ and standard deviations $\sigma_1, \sigma_2, \ldots, \sigma_n$ respectively then their linear combination $a_1X_1 + a_2X_2 + \ldots + a_nX_n$ where $a_1, a_2, \ldots, a_n$ are constants, is also a normal variate with
$$\text{Mean } (\mu) = a_1\mu_1 + a_2\mu_2 + \ldots + a_n\mu_n\text{; and}$$
$$\text{Variance } (\sigma^2) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \ldots + a_n^2\sigma_n^2.$$
Thus *sum (or difference) of independent normal variates is a normal variate.*
14. Normal distribution is a limiting case of binomial distribution when the number of trials is very large and neither $p$ nor $q$ is very small. It is also limiting case Poisson distribution when its mean ($\lambda$) is very large.
15. No portion of the curve lies below the $x$–axis, since the probability can never be negative.
16. **Area property:**
The most important property of normal distribution is the area property which is described as given below:




Graph of normal probability density

(i) The area under the normal probability curve between the ordinates $X = \mu - \sigma$ and $X = \mu + \sigma$ is 0.6826. In other words the range $\mu \pm \sigma$ covers 68.26% of the observations. So, $P(\mu - \sigma, \mu + \sigma) = 0.6826$

(ii) The range $X = \mu \pm 2\sigma$ covers 95.44%. So $P(\mu - 2\sigma, \mu + 2\sigma) = 0.9544$.

(iii) The range $X = \mu \pm 3\sigma$ covers 99.73% of the observations. So, $p(\mu - 3\sigma, \mu + 3\sigma) = 0.9973$. Hence, for practical purposes, the range $\mu \pm 3\sigma$ covers the entire area which is approximately 1.

The standard normal variate corresponding to $X$ is $Z = \dfrac{X-\mu}{\sigma}$.

When $X = \mu \pm \sigma$, $Z = \pm 1$

When $X = \mu \pm 2\sigma$, $Z = \pm 2$

When $X = \mu \pm 3\sigma$, $Z = \pm 3$

Hence the area under the standard normal probability curve

(a) between the ordinates $Z = \pm 1$ is 0.6826

(b) between the ordinates $Z = \pm 2$ is 0.9544

(c) between the ordinates $Z = \pm 3$ is 0.9973

## 4.6.5 Properties of Standard Normal Distribution

1. *pdf* of $Z$ is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \text{ for } -\infty < Z < \infty.$$

2. $Z$ has mean, median and mode all equal to zero.
3. Standard deviation of $Z$ is 1. Also the approximate value of mean deviation (*MD*) and quartile deviation (*Q.D.*) are 0.8 and 0.675 respectively.
4. The standard normal distribution is symmetrical about $Z = 0$.
5. The points of inflection of the probability curve of the standard normal distribution are $-1$ and 1.
6. The two tails of the standard normal curve never touch the horizontal axis.
7. If $\bar{X}$ denotes the arithmetic mean of a random sample of size $n$ drawn from population then,

$$Z = \frac{\sqrt{n}\,(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$$

## 4.6.6 Importance of Normal distribution

Normal distribution has a very importance role in statistics. Some of its applications are:

1. Most of the discrete distribution occurring in practice, e.g., *Binomial, Poisson, Hypergemotric distribution* etc. can be approximated by normal distribution. For large value of $n$, computation of probabilities for discrete distributions becomes quite tedious and time consuming. In such cases, *normal approximation* can be used with great ease and convenience.
2. For large value of $n$ (i.e., $n \to \infty$) almost all the sampling distributions, e.g., *Student's t–distribution, Z–distribution, F–distribution* and *Chi square distribution* conform to normal distribution.
3. The whole theory of *exact sample* (*small sample*) tests, namely, $t$, $F$, $\chi^2$ tests, etc. is based on the assumption that the parent population from which the samples have been drawn follows *Normal distribution*.





4. For large values of $n$ (i.e., $n \to \infty$), the central limit theorem follows normal distribution.
5. It is extensively used in large sampling theory to find estimates of parameters from statistic and confidence limits, etc.
6. Many of the distributions which are not normal can be made normal by simple transformation.
7. It is used in statistical Quality control in Industry for the setting of control limits.

Normal distribution has wide applications in engineering science also. There are two principle applications of normal distribution to engineering science. One application deals with the analysis of items which exhibit failure due to wear, such as mechanical devices. Frequently the wear–out failure distribution is sufficiently close to normal that the use of this distribution for predicting or assessing reliability is valid.

Another application is in the analysis of manufactured items and their ability to meet specifications. No two parts made to the same specification are exactly alike. The variability of those parts leads to a variability in systems composed of those parts. The design must take this part variability into account, otherwise the system may not meet the specification requirement due to combined effect of part variability. Another aspect of this application is in quality control procedures.

## 4.6.7 Computation of the areas under the normal probability curve (Finding Probabilities) when the value of standard normal variate (*Z* Scores) is known.

Using Table A-3 (in Appendix A and the Formula), we can find areas (or probabilities) for many different regions. Such areas can be found by using Table A-3, a calculator, or software such as SPSS, Minitab, or Excel. It is not necessary to know all five methods; you only need to know the method you will be using for class and tests.

Because the following examples and exercises are based on Table A-3, it is essential to understand these points.

1. Table A-3 is designed only for the standard normal distribution, which has a mean of 0 and a standard deviation of 1.
2. Table A-3 is on two pages, with one page for negative z scores and the other page for positive $Z$ scores.
3. Each value in the body of the table is a cumulative area from the left up to a vertical boundary above a specific z score.
4. When working with a graph, avoid confusion between z scores and areas.

   ***Z*-score:** Distance along the horizontal scale of the standard normal distribution; refer to the leftmost column and top row of Table A-3.

   **Area:** Region under the curve; refer to the values in the body of Table A-3.
5. The part of the Z-score denoting hundredths is found across the top row of Table A-3.

**Example 26:** **(Calculation of standard normal probabilities)**

Find the probabilities that a random variable having the standard normal distribution will take on a value: (a) between 0.87 and 1.28; **(b)** between $-$ 0.34 and 0.62; (c) greater than 0.85; (d) greater than $-$ 0.65; (e) less than $-$ 0.8; (f) between 0 and 2.34; (g) less than $-$ 0.84 or greater than 2.08.

What is the value of Z if only 31.87% of all possible $Z$ values are smaller?

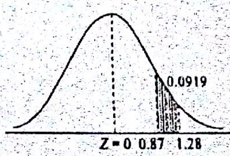
**Solution:**

Here, $Z$ is a Standard Normal variate

(a) $P(0.87 < Z < 1.28) = F(1.28) - F(0.87)$

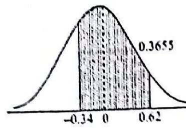
$= 0.8997 - 0.8087$ [Using Z–Table A-3]

$= 0.0919$

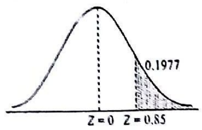


or, $P(0.87 < Z < 1.28)$
$= P(0 < Z < 1.28) - P(0 < Z < 0.87)$
$= 0.3997 - 0.3078 = 0.0919$

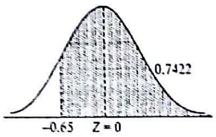
(b) $P(-0.34 < Z < 0.62) = F(0.62) - F(-0.34)$
$= 0.7324 - 0.3669 = 0.3655$
or, $P(-0.34 < Z < 0.62)$
$= P(-0.34 < Z < 0) + P(0 < Z < 0.62)$
$= P(0 < Z < 0.34) + P(0 < Z < 0.62)$ [By symmetry]
$= 0.1331 + 0.2324 = 0.3655$



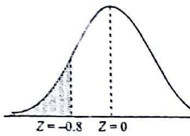

(c) $P(Z > 0.85) = 1 - P(Z \le 0.85)$
$= P(0 < Z < \infty) - P(0 < Z \le 0.85)$
$= 0.5 - 0.3023$ [Using Z–Table A-3]
$= 0.1977$



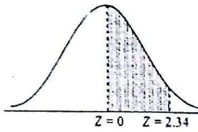

(d) $P(Z > -0.65) = P(-0.65 < Z < 0) + P(0 < Z < \infty)$
$= P(0 < Z < 0.65) + 0.5$
$= 0.2422 + 0.5 = 0.7422$
or, $P(Z > -0.65) = 1 - F(-0.65)$
$= 1 - 0.2578$
$= 0.7422$
or, $P(Z > -0.65) = 1 - F(-0.65)$
$= 1 - [1 - F(0.65)] = F(0.65) = 0.7422$



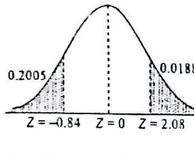

(e) $P(Z < -0.8) = F(-0.8)$
$= 0.2199$
or, $P(Z < -0.8)$
$= P(-\infty < Z < 0) - P(-0.8 < Z < 0)$
$= 0.50 - P(0 < Z < .8)$ [*By symmetry*]
$= 0.50 - 0.2881 = 0.2199$




(f) $P(0 < Z < 2.34) = 0.4904$




(g) $P(Z < -0.84 \text{ or } Z > 2.08)$
$= P(Z < -0.84) + P(Z > 2.08)$
$= F(-0.84) + 1 - P(Z < 2.08)$
$= 0.2193$




Now, Let the value of $Z$ be $-z_1$ such that only 31.87% of all values of $z$ are smaller than $-z_1$,
i.e., $P(Z < -z_1) = 0.3187$
or, $F(-z_1) = 0.3187$
The value of probability closest to 0.3187 in Z– table is 0.3192 at $z = 0.47$.
$\therefore\ z_1 = 0.47$
Hence, the required value of $Z$ is $-0.47$.

### 4.6.8 Finding Values from Known Areas

If we are given specific limit values and we must find an area (or probability or percentage). In many practical and real cases, the area (or probability or percentage)





is known and we must find the relevant value(s). When finding values from known areas, be careful to keep these cautions in mind:

1. *Don't confuse Z scores and areas*. Remember, Z scores are distances along the horizontal scale, but areas are regions under the normal curve. The normal table lists Z scores in the left columns and across the top row, but areas are found in the body of the table.
2. *Choose the correct (right/left) side of the graph*. A value separating the top 10% from the others will be located on the right side of the graph, but a value separating the bottom 10% will be located on the left side of the graph.
3. A *Z-score* must be *negative* whenever it is located in the *left* half of the normal distribution.
4. Areas (or probabilities) are *positive* or *zero* values, but they are never negative.

Graphs are extremely helpful in visualizing, understanding, and successfully working with normal probability distributions, so they should be used whenever possible.

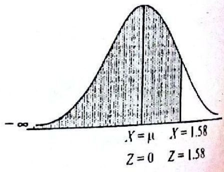
**Notation:** $P(a < Z < b)$ denotes the probability that the $Z$ score between $a$ and $b$.
$P(Z > a)$ denotes the probability that the $Z$ score is greater than $a$.
$P(Z < a)$ denotes the probability that the $Z$ score is less than $a$.

### 4.6.9 Computation of Area of Normal curve when the value of variable is known.

**Example - 27:** The Precision Scientific Instruction Company manufactures thermometers that are supposed to give reading of 0°c at the freezing point of water. Tests on a large sample of these instruments reveal that at the freezing point of water, some thermometers give reading below 0° (denoted by negative numbers) and some give readings above 0° (denoted by positive numbers). Assume that the mean reading is 0°c and the standard deviation of the readings is 1°c. Also assume that the readings are normally distributed. If one thermometer is randomly selected, find the probability of randomly selecting one thermometer that reads (at the freezing point of water) (a) The reading is less that 1.58°; (b) Above −1.23°
(c) Between −2.00° and 1.50°. [*TU, BE, 2067 Mangsir*]

**Solution:** Let $X$ be the normal variate denoting the readings of the thermometer with
Mean ($\mu$) = 0°C, $\sigma$ = 1.00°C

(a) For $X = 1.58$,
$$Z = \frac{X-\mu}{\sigma} = \frac{1.58-0}{1} = 1.58$$

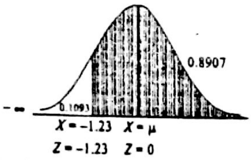



$\therefore\ P(X < 1.58) = P(Z < 1.58)$
$= F(1.58) = 0.9429$ [Using Z–table 3]

[That is, the probability distribution of readings is a standard normal distribution, because the readings are normally distributed with $\mu = 0$ and $\sigma = 1$. We need to find the area in the figure below $z = 1.58$. The area below $z = 1.58$ is equal to the probability of randomly selecting a thermometer with a reading less than 1.58°. From Standard Normal Table A-3, we find that this area is 0.9429.]

**Interpretation:** The probability of randomly selecting a thermometer with a reading less than 1.58° (at the freezing point of water) is equal to the area of 0.9429 shown as the shaded region in figure. Another way to interpret this result is to conclude that 94.29% of the thermometers will have readings below 1.58°.

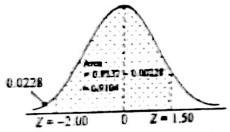
(b) $P(X > -1.23) = P\left(\frac{X-\mu}{\sigma} > \frac{-1.23-0}{1}\right)$
$= P(Z > -1.23) = 1 - P(Z \le -1.23)$
$= 1 - F(-1.23)$
$= 1 - 0.1093$ [*Using Z–table* A-3]
$= 0.8907$




[That is, we again find the desired probability by finding a corresponding area. We are looking for the area of the region that is shaded in figure but standard normal table is designed to apply only to cumulative areas from the left. Referring to standard normal table for the page with negative Z scores, we find that the cumulative area from the left up to $z = -1.23$ is 0.1093 as shown. Knowing that the total area under the curve is 1, we can find the shaded area by subtracting. 0.1093 from 1. The result is 0.8907. Even though standard normal table is designed only for cumulative areas from the left, we can use it to find cumulative areas from the right, as shown in figure].

**Interpretation:** Because of the correspondence between probability and area, we conclude that the probability of randomly selecting a thermometer with a reading above −1.23° at the freezing point of water is 0.8907 (which is the area above $z = -1.23$). In other words, 89.07% of the thermometers have readings above −1.23°.

(c) $P(-2.00° < X < 1.50°)$
$= P(-2.00° < Z < 1.50)$
$= F(1.50) - F(-2.00)$
$= 0.9332 - 0.0228 = 0.9104$




[That is, we are again dealing with normally distributed values having a mean of 0° and a standard deviation of 1°. The probability of selecting a thermometer that reads between −2.00° and 1.50° corresponds to the shaded area in figure. Normal table cannot be used to find that area directly, but we can use the table to find that $z = -2.00$ corresponds to the area of 0.0228, and $z = 1.50$ corresponds to the area of 0.9332, as shown in the figure. Refer to figure to see that the shaded area is the difference between 0.9332 and 0.0228. The shaded area is therefore 0.9332 − 0.0228 = 0.9104.]

**Interpretation:** Using the correspondence between probability and area, we conclude that there is a probability of 0.9104 of randomly selecting one of the thermometers with a reading between −2.00° and 1.50° at the freezing point of water. Another way to interpret this result is to state that if many thermometers are selected and tested at the freezing point of water, then 0.9104 (or 91.04%) of them will read between −2.00° and 1.50°.

**Example 28:** The diameter of a shaft in an optical storage drive is normally distributed with mean 0.2508 inch and standard deviation 0.0005 inch. The specifications on the shaft are 0.2500 ± 0.0015 inch. What proportion of shafts conforms to specifications?

**Solution:** Let X denote the shaft diameter in inches. The requested probability is

$$P(0.2485 < X < 0.2515) = P\left(\frac{0.2485-0.2508}{0.0005} < Z < \frac{0.2515-0.2508}{0.0005}\right)$$
$$= P(-4.6 < Z < 1.4) = P(Z < 1.4) - P(Z < -4.6)$$
$$= 0.91924 - 0.0000 = 0.91924$$



Most of the nonconforming shafts are too large, because the process mean is located very near to the upper specification limit. If the process is centered so that the process mean is equal to the target value of 0.2500,

$$P(0.2485 < X < 0.2515) = P\left(\frac{0.2485-0.2500}{0.0005} < Z < \frac{0.2515-0.2500}{0.0005}\right)$$
$$= P(-3 < Z < 3) = P(Z < 3) - P(Z < -3)$$
$$= 0.99865 - 0.00135$$
$$= 0.9973$$

By reentering the process, the yield is increased to approximately 99.73%.

**Example 29: (Calculation of probabilities using normal distribution)**

The average height of students of Thapathali Engineering Campus is 165 *cms* and standard deviation is 10 *cms*.

(a) Find the percentage of students whose height is (i) less than 172 *cms*; (ii) between 150 and 180 *cms*; (iii) more than 174 *cms*.

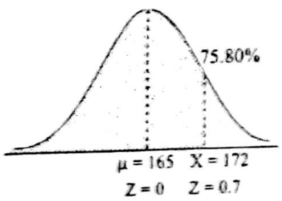
(b) For what value of the variable $X$, $P(X < x_1) = 0.10$ holds?

**Solution:** Let $X$ be the normal variate that follows normal distribution with mean $\mu$ = 165 *cms* and $\sigma$ = 10 *cms*

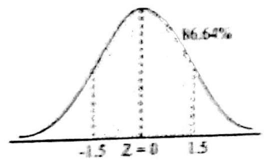
**(a)** (i) For $P(X < 172)$ :

when $X = 172$, $Z = \frac{X-\mu}{\sigma} = \frac{172-165}{10} = 0.7$

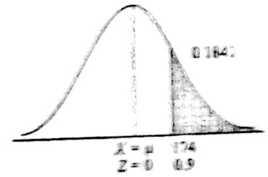
Therefore, $P(X < 172)$
$= P\left(\frac{X-\mu}{\sigma} < \frac{172-165}{10}\right)$
$= P(Z < 0.7) = F(0.7) = 0.7580$
$= 75.80\%$



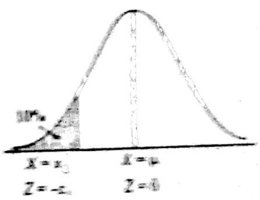

(ii) $P(150 < X < 180)$
$= P\left(\frac{150-165}{10} < \frac{X-\mu}{\sigma} < \frac{180-165}{10}\right)$
$= P(-1.5 < Z < 1.5)$
$= F(1.5) - F(-1.5) = 0.9332 - 0.0668$
$= 0.8664 = 86.44\%$




(iii) $P(X > 174) = P\left(\frac{X-\mu}{\sigma} > \frac{174-165}{10}\right)$
$= P(Z > 0.9) = 1 - P(Z \le 0.9)$
$= 1 - 0.8159 = 0.1841$




**(b)** When $X = x_1$
$Z = \frac{X-\mu}{\sigma} = \frac{x_1-165}{10} = -z_1$
[∵ $-x_1$ lies to the left of $\mu$]
∴ $P(Z < -z_1) = 0.1000$



The value of the probability closest to 0.1000 in Z-table is 0.1003 at $z_1 = -1.28$. Hence, $\frac{x_1 - 165}{10} = -1.28$

$\therefore$ $x_1 = 152.2$ cm. Hence the required value of $X$ is 152.2 cm.

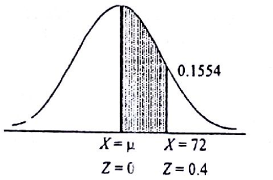
**Example 30:** The daily wages of 1,000 workmen are normally distributed around a mean of Rs. 70 and with a standard deviation of Rs.5. Estimate the number of workers whose daily wages will be: (a) between Rs. 70 and Rs. 72; (b) between Rs. 69 and Rs. 72; (c) more than Rs. 75; (d) less than Rs. 63; (e) Also estimate the lowest daily wages of the 100 highest paid workers. [TU 2061]

**Solution:** Let the random variable $X$ denote the daily wages in rupees. Then $X$ is a normal random variable with mean $\mu = 70$ and $\sigma = 5$. The standard normal variable corresponding to $X$ is

$$Z = \frac{X-\mu}{\sigma} = \frac{X-70}{5}$$

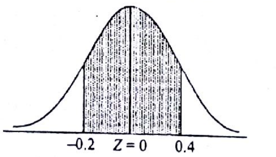
| $X$ | 63 | 69 | 70 | 72 | 75 | 80 |
|---|---|---|---|---|---|---|
| $Z = \frac{X-70}{5}$ | $\frac{63-70}{5} = -1.4$ | -0.2 | 0 | 0.4 | 1 | 2 |

(a) $P(70 < X < 72) = P(0 < Z < 0.4)$
$= F(0.4) - F(0) = 0.6554 - 0.5 = 0.1554$
[*Using Z–table* A-3]
Therefore the number of workers with daily wages between Rs.70 and Rs. 72 = 1000 × 0.1554 = 155.4 ≈ 155.

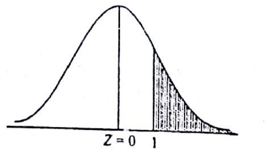



(b) $P(69 < X < 72) = P(-0.2 < Z < 0.4)$
$= F(0.4) - F(-0.2)$
$= 0.6554 - 0.4207 = 0.2347$

or, $P(69 < X < 72) = P(-0.2 < Z < 0.4)$
$= P(-0.2 < Z < 0) + P(0 < Z < 0.4)$
$= P(0 < Z < 0.2) + P(0 < Z < 0.4)$
[By symmetry]
$= 0.0793 + 0.1554 = 0.2347$
Hence the required number of workers = 1000 × 0.2347 = 243.7 ≈ 235.

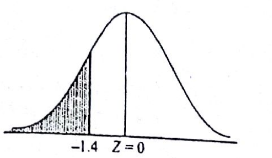



(c) $P(X > 75) = P(Z > 1)$
$= 1 - P(Z \le 1) = 1 - 0.8413 = 0.1587$
Required no. of workers = 1000 × 0.1587
= 158.7 ≈ 159




(d) $P(X < 63) = P(Z < -1.4)$
$= F(-1.4) = 0.0808$ [*Using Z table* A-3]
Required number of workers
= 1000 × 0.0808 = 80.8 ≈ 81

(e) Proportion of the 100 highest paid workers = 100/1000 = 0.1000
We want to determine $X = x_1$, say, such that $P(X > x_1) = 0.1000$








When $X = x_1$, $Z = \frac{X-70}{5} = \frac{x_1 - 70}{5} = z_1$ (say)

$\therefore$ $P(Z > z_1) = 0.10 \Rightarrow 1 - P(Z \le z_1) = 0.10 \Rightarrow P(Z \le z_1) = 0.90$

The value of the probability closure to 0.90 in Z–table A-3, (i.e., normal probability table) is 0.8997 at $z_1$ is 1.28

$\therefore$ $z_1 = \frac{x_1 - 70}{5}$ $\Rightarrow x_1 = 70 + 5 \times 1.28 = 76.40$

Hence, the lowest daily wages of the 100 highest paid workers are Rs. 70.40.

**Example 31:** A sample of 600 dry battery cells tested to find the length of life produced the following result $\bar{X} = 12$ hours, $\sigma = 2.5$ hours. Assuming the data to be normally distributed, find the number of battery cells that are expected to have life (a) more than 15 hours; (b) between 10 and 14 hours. (c) less than 6 hours.
[TU, BE 2061 Ashwin]

**Solution:** (a) 531; (b) 346; (c) 5. [T.U. BE 2061 Ashwin]

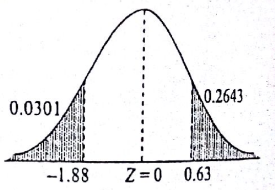
**Example 32:** The Average diameter of sample of 1000 pipes is 2.6 inches and standard deviation is 0.55. Assuming that the diameter of pipes is normally distributed. Find the number of pipes of diameter: (a) greater than 2 inches; (b) between 2 and 3.1 inches. [TU, BE, 2063 Kartik]

**Solution:** (a) 862; (b) 681.

**Example 33:** The mean elongation of steel bar under a particular tensile load has been established to be normally distribution with parameters $\mu = 0.06''$ and $\sigma = 0.008''$. Assuming the same distribution applies to new bar, find the probability that the mean elongation falls:
(a) above 0.08"; (b) somewhere between 0.05" and 0.07"; (c) either below 0.045" or above 0.065". [TU, BE 2063 Kartik]

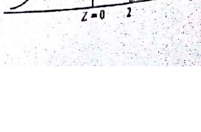
**Solution:** (a) $P(X > 0.08) = P(Z > 2.5) = 1 - F(2.5) = 0.0062$
(b) $P(0.05 < X < 0.07) = P(-1.25 < Z < 1.25)$
$= F(1.25) - F(-1.25) = 0.8944 - 0.1056 = 0.7888$ [*from Z table* A-3]

(c) $P(X < 0.045 \text{ or } X > 0.065)$
$= P(X < 0.045 \cup X > 0.065)$
$= P(Z < -1.88 \cup Z > 0.63)$
$= P(Z < -1.88) + P(Z > 0.63)$
$- P(Z < -1.88 \cap Z > 0.63)$
$= 0.0301 + 1 - P(Z \le 0.63) - 0$
$= 0.0301 + 1 - 0.7357 = 0.2944.$



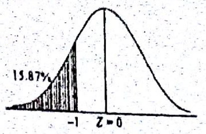

**Example 34:** Scores on a trade school entrance examination exhibit the characteristics of a normal distribution with mean and standard deviation of 50 and 5 respectively.
(a) What proportion of the scores on this examination would be greater than 60?
(b) What proportion of the scores on this examination would be less than 45
(c) What proportion of the scores on this examination would be between 35 and 65? [TU, BE 2064 Poush]

**Solution:** Let $X$ be the normal variate denoting the scores with mean $\mu = 50$ and $\sigma = 5$.
(a) $P(X > 60) = P(Z > 2) = 1 - F(2)$
$= 1 - 0.9772 = 0.0228 = 2.28\%$



(b) $P(X < 45) = P\left(\frac{X-50}{5} < \frac{45-50}{5}\right)$

$= P(Z < -1) = 0.1587 = 15.87\%$

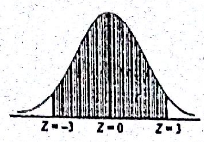



(c) $P(35 < X < 65)$

$= F\left(\frac{65-50}{5}\right) - F\left(\frac{35-50}{5}\right)$

$= F(3) - F(-3) = 0.9987 - 0.0013 = 0.9974$

$= 99.74\%$




**Example 35:** Incomes of a group of 10,000 persons were found to be normally distributed with mean Rs. 1,520 and *s.d.* Rs.160 Find:

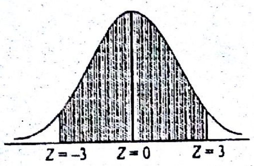
(a) highest income of poorest 2000 persons;

(b) lowest income of richest 1000 persons. [TU 2065]

**Solution:** Let $X$ be the normal variate denoting the income of persons with $\mu = 1520$, $\sigma = 160$. Then $Z = \frac{X-\mu}{\sigma} = \frac{X-1520}{160}$

(a) Let the highest income of poorest 2000 persons be $x_1$. Percentage of persons $= \frac{2000}{10000} \times 100\% = 20\%$

When, $X = x_1$, then $Z = \frac{x_1 - 1520}{160} = -z_1$(say)




[$\because z_1$ *lies to the left of mean*]

Now, $P(X < x_1) = P(Z < -z_1) = 20\%$

or, $F(-z_1) = 0.2000$

From Standard Normal Table A-3, the probability closest to 0.2000 is 0.2005 at $z_1$ = 0.84.

Therefore, $\frac{x_1 - 1520}{160} = 0.84$

$x_1 = 1520 - 134.4 = 1385.6$

The highest income of the poorest 2000 persons in Rs. 1385.60.

(b) Percentage of the richest 1000 persons $= \frac{100}{10000} \times 100\% = 10\%$

Let, the lowest income of the richest 1000 persons be $x_2$.

When $X = x_2$, then $z = \frac{x_2 - 1520}{160} = z_2$ (say)

[$\because z_2$ *lies to the right of mean*]

Now $P(X > x_2) = P(Z > z_2) = 10\%$

That is, $1 - P(Z < z_2) = 0.1000$

Therefore, $F(z_2) = 1 - 0.10 = 0.90$.

From Standard Normal Table A-3, the value of the probability closest to 0.9000 is 0.8997 at $z_2 = 1.28$.

Hence $\frac{x_2 - 1520}{160} = 1.28$



or, $x_2 = 1520 + 204.8 = 1724.8$.

The lowest income of the richest 1000 persons is Rs. 1724.80.

**Example 36:** In an aptitude test administered to 900 college students, the mean was found to be 50 and *s.d.* of 20. Assuming the marks in an aptitude test of students approximately follows normal distribution. Find

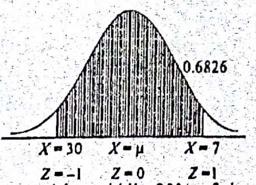
(a) The number of students securing between 30 and 70;

(b) The range of marks of middle 80% of the students;

(c) The value of the score exceeded by top 90 students. [TU 2050 MBA]

**Solution:** Let $X$ be the normal variate denoting the marks of the students in an aptitude test that follows normal distribution with $\mu = 50$ and $\sigma = 20$. The *SNV* corresponding to $X$ is

$Z = \frac{X-\mu}{\sigma} = \frac{X-50}{20}$.




(a) $P(30 < X < 70) = P(-1 < z < 1)$

$= F(1) - F(-1) = 0.8413 - 0.1587$

$= 0.6826 = 68.26\%$

(b) Let $x_1$ and $x_2$ be te lowest and highest scores secured by middle 80% of the students. Here

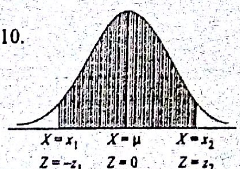
$P(X < x_1) = 10\% = 0.10$ and $P(X > x_2) = 10\% = 0.10$.

For $X = x_1$, $Z = \frac{X-\mu}{\sigma} = \frac{x_1 - 50}{20} = -z_1$(say)

[$\because x_1$ *lies to the left of mean*]

$\therefore$ $P(X < x_1) = 0.10$

or, $P(Z < -z_1) = 0.10$ or $F(-z_1) = 0.10$




From normal table, the value of probability closest to 0.10 is 0.1003 at $z = 1.28$. So, $z_1 = 1.28$.

$\frac{x_1 - 50}{20} = -1.28 \Rightarrow x_1 = 24.4$

Again, For $X = x_2$, $Z = \frac{x_2 - \mu}{\sigma} = z_2$ (say)

Therefore, $P(X > x_2) = 0.10 \Rightarrow P(Z > z_2) = 0.10 \Rightarrow 1 - P(Z < z_2) = 0.10$

Therefore, $P(Z < z_2) = 0.90$

From normal table, the value of probability closer to 0.90 is 0.8997 at $z = 1.28$. so, $z_2 = 1.28$

Hence $\frac{x_2 - 50}{20} = 1.28 \Rightarrow x_2 = 75.6$

The lowest and the highest scores secured by middle 80% of the students are 24.4 and 75.6 respectively.

$\therefore$ Range $= 75.6 - 24.4 = 51.2$

(c) Let $x$ be the value of the score exceeded by the top 90 students such that $P(X > x) = \frac{90}{900} = 0.10$.

For $X = x$,

$Z = \frac{X-\mu}{\sigma} = \frac{X-50}{20} = z_3$ (say)

$\therefore$ $P(X > x) = 0.10$

or, $P(Z > z_1) = 0.10$

or, $1 - P(Z \le z_1) = 0.10$

or, $F(z_1) = 0.90$

From the Standard Normal Table A-3, the value of the probability closest to 0.90 is 0.8997 at $z = 1.28$. So, $z_1 = 1.28$.

Hence $\frac{x - 50}{20} = 1.28 \Rightarrow x = 75.6$.

So, the top 90 students exceeded the score 75.6.

**Example 37:** The distribution of wages of a group of workers is known to be normal with mean Rs. 500 and s.d. Rs. 100. If the wages of 100 workers in the group are less than Rs. 430, what is the total number of workers in the group?

**Solution:** Let $X$ denote the wages. It is given that $X$ is normally distributed with mean $\mu$ = Rs 500 and s.d. as Rs. 100 and $P(X < 430) = 100/N$, $N$ being the total number workers in the group.

$$\Rightarrow P\left(\frac{X - 500}{100} < \frac{430 - 500}{100}\right) = \frac{100}{N}$$

$\Rightarrow P(Z < -0.70) = 100/N$.

or, $N\,F(-0.70) = 100$

Therefore, $N = \frac{100}{0.2420} = 413$

**Example 38:** The final examination scores in statistics of *BE* are normally distributed with an average score of 70 and variance of 25.

(a) If the lowest passing grade is 58, what percent of class is failing?

(b) If the professor gives the grade on a course and everybody getting 82 or above gets grade of *A*, then what percentage of students get *A* grade?

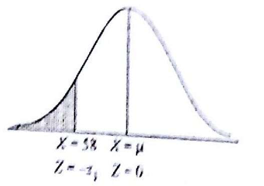
(c) If the highest 20% of the class are to pass the course, what is the lowest passing score?

(d) What should the score be so that only 15% of the students get a score higher than this?

**Solution:** Let $X$ be the normal variate with mean $\mu = 70$ and variance $\sigma^2 = 25$.

Therefore, $\sigma = 5$. Let $Z$ be *SNV* corresponding to $X$.

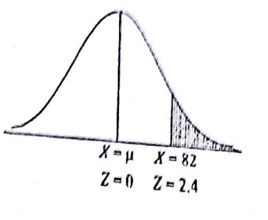
Then $Z = \frac{X - \mu}{\sigma} = \frac{X - 70}{5}$




(a) $P(X < 58) = P\left(\frac{X - 70}{5} < \frac{58 - 70}{5}\right)$

$= P(z < -2.4) = F(-2.4) = 0.0082$

$= 0.82\% \approx 1\%$ [Using normal table A-3]




(b) $P(X > 82) = \left(\frac{X - 70}{5} > \frac{82 - 70}{5}\right)$

$= P(Z > 2.4) = 1 - F(2.4)$

$= 1 - 0.9918 = 0.0082 = 0.82\% \approx 1\%$





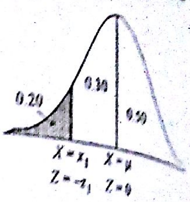
(c) Let $x_1$ be the lowest passing score and $z_1$ be its *SNV*.

Then $P(X > x_1) = 0.80$

or, $P(Z > -z_1) = 0.80$

or, $1 - P(Z \le -z_1) = 0.80$

or, $F(-z_1) = 0.20$




From the normal table the value of the probability closer to 0.20 is 0.2005 at $Z = 0.84$. So, $z_1 = -0.84$

$\frac{x_1 - 70}{5} = -z_1 \Rightarrow x_1 = 70 - 0.84 \times 5 = 65.8$

Hence, a students get at least 65.8 to pass the course.

(d) Let $x_2$ be the lowest score where only 15% of the students get a score higher than this. Let $z_2$ be its *SNV* value.

$\therefore\ Z = \frac{X - \mu}{\sigma} = \frac{x_2 - 70}{5} = z_2$

Now, $P(X > x_2) = 0.15 \Rightarrow P(Z > z_2) = 0.15 \Rightarrow 1 - F(z_2) = 0.15 \Rightarrow F(z_2) = 0.85$

From the normal table the value of probability closer to 0.85 is 0.8508 at $Z = 1.04$. So $z_2 = 1.04$.

Hence, $\frac{x_2 - 70}{5} = 1.04 \Rightarrow x_2 = 70 + 5.2 = 75.2$

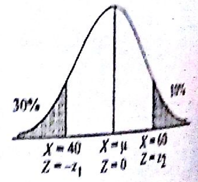
So, only 15% of the students get a score higher than 75.2.

### 4.6.10 Computation mean and *s.d.* when the value of variable and area under the normal curve are known

**Example 39:** In an examination it was observed that 10% of the candidates got first class and 30% failed. The pass mark is 40% and first division mark is 60%. Assuming that the marks follow normal distribution, find the mean and standard deviation of the distribution. [TU 2065 MBA]

**Solution:** Let $X$ be a random variable that follows normal distribution. Let $Z$ be SNV corresponding to $X$.

Then $Z = \frac{X - \mu}{\sigma}$




when $X = 40$, $Z = \frac{40 - \mu}{\sigma} = -z_1$ (say)

Now, $P(X < 40) = 0.30$

or, $P(Z < -z_1) = 0.30$

or, $F(-z_1) = 0.30$

From normal table A-3, the value of probability closest to 0.30 is 0.3015 at $Z = 0.52$. So $z_1 = 0.52$.

$\therefore\ \frac{40 - \mu}{\sigma} = -0.52 \Rightarrow 40 - \mu = -0.52\sigma$ ............... (1)

Again, when $X = 60$, $Z = \frac{X - \mu}{\sigma} = \frac{60 - \mu}{\sigma} = z_2$ (say)

Now, $P(X > 60) = 0.10 \Rightarrow P(Z > z_2) = 0.10 \Rightarrow 1 - P(Z \le z_2) = 0.10$

$\Rightarrow F(z_2) = 0.90$

From the normal table A-3 the value of the probability closer to 0.90 is 0.8997 at $Z = 1.28$. So, $z_2 = 1.28$.

Hence, $\frac{60-\mu}{\sigma} = 1.28 \Rightarrow 60-\mu = 1.28\sigma$ ------------ (2)

Solving equation (1) and (2) we get $\sigma = 0.09$ and $\mu = 40.0468$.

**Example 40:** The actual amount of instant coffee that a filling machine puts into "4 – ounce" jars may be looked upon as a random variable having a normal distribution with $\sigma = 0.04$ ounces. If only 2% of the jars are to contain less than 4 ounces, what should be the mean fill of these jars?

**Solution:** Let $X$ be rv denoting the actual amount of instant coffee with $\sigma = 0.04$ ounce. Let $Z$ be the corresponding *SNV* of $X$.

When $X = 4$, $Z = \frac{X-\mu}{\sigma} = \frac{4-\mu}{0.04} = -z_1$ (say)

Now, $P(X < 4) = 0.02$

or, $P(Z < -z_1) = 0.02$

or, $F(-z_1) = 0.02$

From the normal table the value of the probability closest to 0.02 is 0.0202 at $z = 2.05$. So, $z_1 = 2.05$

Now, $\frac{4-\mu}{0.04} = -2.05$, $\Rightarrow \mu = 4.082$ ounces.

**Example 41:** Of a large group of men 5% are under 60 inch in height and 40% are between 60 and 65 inches. Assuming a normal distribution, find the mean and *s.d.* of height.

**Solution:** Let $X$ be the normal variate denoting the height of the men with mean $\mu$ and *s.d.* $\sigma$. Let $Z$ be the corresponding *SNV* of $X$. When $X = 60$

$Z = \frac{X-\mu}{\sigma} = \frac{60-\mu}{\sigma} = -z_1$ (say)

Now, $P(X < 60) = 0.05$

or, $P(Z < -z_1) = 0.05$

or, $F(-z_1) = 0.05$

From the normal table A-3 the value of the probability closest to 0.05 is 0.0505 at $Z = 1.64$. So, $z_1 = 1.64$.

So, $\frac{60-\mu}{\sigma} = -1.64$

or, $60-\mu = -1.64\sigma$ ... (1)

Again, For, $X = 65$

$Z = \frac{X-\mu}{\sigma} = \frac{65-\mu}{\sigma} = -z_2$ (say)

$\therefore$ $P(60 < X < 65) = 0.40$

or, $P(-z_1 < Z < -z_2) = 0.40$

or, $F(-z_2) - F(-z_1) = 0.40$

or, $F(-z_2) = 0.40 + 0.05 = 0.45$

From normal table A-3, the value of the probability closest 0.45 is 0.4483 at $Z = 0.13$. So, $z_2 = 0.13$.

Therefore, $\frac{65-\mu}{\sigma} = -0.13$

or, $65-\mu = -0.13\sigma$ ... (2)

Solving equations (1) and (2) we get,

$\mu = 65.43$ inches and $\sigma = 3.31$ inches.





**Example 42:** In a normal distribution, 7% of the items are under 35 and 89% are under 63. What are the mean and standard deviation of the distribution?

[Pokhara Uni. BE 2007 Spring]

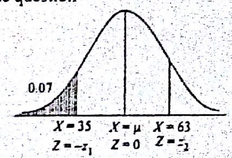
**Solution:** Let $X$ denote the normal variable and $Z$ be its corresponding *SNV*. Let $\mu$ and $\sigma$ be the mean and standard deviation of $X$. By the question

$P(X < 35) = 0.07$

$P(X < 63) = 0.89$

Since $P(X < 35) < 0.5$,

$X = 35$ lies to the left of the mean, so the corresponding value $Z$ is negative.




When $X = 35$, $Z = \frac{X-\mu}{\sigma} = \frac{35-\mu}{\sigma} = -z_1$ (say)

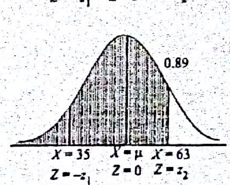
Now, $P(X < 35) = 0.07$

or, $P(Z < -z_1) = 0.07$

or, $F(-z_1) = 0.07$

From the normal table the value of probability closest to 0.07 is 0.0694 at

$Z = 1.48$. So, $z_1 = 1.48$.




Hence, $\frac{35-\mu}{\sigma} = -1.48$

or $35-\mu = -1.48\sigma$ ... (1)

Again, $P(X < 63) = 0.89 \Rightarrow P(Z < z_2) = 0.89 \Rightarrow F(z_2) = 0.89$

From the normal table the value of the probability closest to 0.89 is 0.8907 at $Z = 1.23$. So, $z_2 = 1.23$.

Hence, $\frac{63-\mu}{\sigma} = 1.23 \Rightarrow 63-\mu = 1.23\sigma$ ... (2)

Solving equations (1) and (2) we get, $\sigma = 10.33$ and $\mu = 50.3$.

**Example 43:** The distribution of marks obtained in examination is normal. 44% of the candidates get marks below 61 and 4% of the candidates got marks above 80. Find the percentage of candidates who got marks above 65.

**Solution:** Let $X$ denote the marks obtained in the examination and $Z$ be its *SNV*.

By question $P(X < 61) = 0.44$ and $P(X > 80) = 0.04$

Since $P(X < 61) < 0.5$, $X = 61$ lies to the left of the mean

So, when, $X = 61$, $Z = \frac{X-\mu}{\sigma} = \frac{61-\mu}{\sigma} = -z_1$ (say)

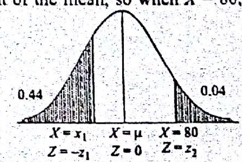
Since $P(X > 80) > 0.5$, $X = 80$ lies to the right of the mean, so when $X = 80$,

$Z = \frac{X-\mu}{\sigma} = \frac{80-\mu}{\sigma} = z_2$ (say)

Now, $P(X < 61) = 0.44$

or, $P(Z < -z_1) = 0.44$

or, $F(-z_1) = 0.44$




From the normal table the value of the probability closest to 0.44 is 0.4404 at

$Z = 0.15$. So, $z_1 = 0.15$

Hence, $\frac{61-\mu}{\sigma} = -0.15 \Rightarrow 61-\mu = -0.15\sigma$ ... (1)

Again, $P(X > 80) = 0.04 \Rightarrow P(Z > z_2) = 0.04 \Rightarrow 1 - P(Z \le z_2) = 0.04$

or, $F(z_2) = 0.96$

From the normal table the value of the probability closest 0.96 is 0.9599 at $Z = 1.75$. So, $z_2 = 1.75$

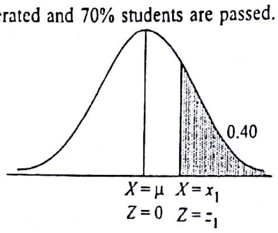
Hence $\frac{80-\mu}{\sigma} = z_2 \Rightarrow 80 - \mu = 1.75\sigma$ . . . (2)

Solving equation (1) and (2) we get $\sigma = 10$, $\mu = 62.5$

At last $P(X > 65) = P\left(Z > \frac{65-62.5}{10}\right)$

$= P(Z > 0.25) = 1 - P(Z \le 0.25) = 1 - 0.5987 = 0.4013$

**Example 44:** The marks of the students in a certain examination are normally distributed with mean marks as 40% and standard deviation marks as 20%. On this basis, 60% students failed. The result was moderated and 70% students are passed. Find the pass marks before after the moderation.




**Solution:** Let the percentage pass marks before and after the moderation be $x_1$ and $x_2$ respectively. If $X$ denote the percentage of marks obtained in the examination and $Z$ its SNV, then

$$Z = \frac{X-\mu}{\sigma} = \frac{X-40}{20}$$

**Before Moderation:**

Pass marks $= x_1\%$

60% failed $\Rightarrow$ 40% students passed.

So, $P(X \ge x_1) = 0.40$.

$\therefore$ Since $P(X \ge x_1) < 0.5$, so, $x_1$ lies to the right of mean and the corresponding value of $Z$ is positive.

When $X = x_1$, $Z = \frac{X-\mu}{\sigma} = \frac{x_1 - 40}{20} = z_1$ (say)

$\therefore$ $P(Z \ge z_1) = 0.40$ $\Rightarrow 1 - P(Z < z_1) = 0.40$

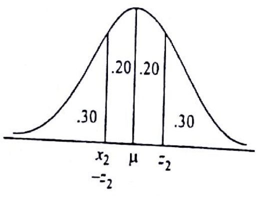
or, $F(z_1) = 0.60$

From the normal table A-3, the value of the probability closest to 0.60 is 0.5987 at $Z = 0.25$. So, $z_1 = 0.25$.

Hence, $\frac{x_1 - 40}{20} = z_1 \Rightarrow x_1 = 40 + 20 \times 0.25 = 45$

The pass mark before moderation are 45%.

**After Moderation:** Pass marks $= x_2\%$




Since 70% students passed,

$P(X \ge x_2) = 0.70$

Here $P(X \ge x_2) > 0.5$, $x_2$ lies to the left of mean. So, the corresponding value of $Z$ is negative. When $X = x_2$, $Z = \frac{X-\mu}{\sigma} = \frac{x_2 - 40}{20}$

$= -z_2$ (say)

$\therefore$ $P(Z \ge -z_2) = 0.70 \Rightarrow 1 - P(Z < -z_2) = 0.70 \Rightarrow$ $F(-z_2) = 0.30$.

From the normal table A-3, the value of the probability closest to 0.30 is 0.3015 at $Z = 0.52$. So, $z_2 = 0.52$

$\therefore$ $\frac{x_2 - 40}{20} = -0.52$ $\Rightarrow x_2 = 40 - 10.4 = 29.6$

Hence the pass marks after moderation are 29.6%.

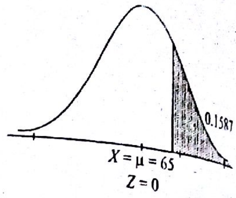




**Example 45:** In certain examination test 3000 students appeared in statistics. The average mark obtained was 65% and *s.d.* was 5%. How many students do you expect to obtain more than 70% mark? What are the minimum marks of top 200 students? Assume that the marks are normally distributed. **[TU, BE, 2062 Baisakh]**

**Solution:** Let $X$ denote the marks with $\mu = 65$ and $\sigma = 5$, and $Z$ be its SNV.

Then $Z = \frac{X-\mu}{\sigma} = \frac{X-65}{5}$




Now, Number of students obtaining more than 70%

$= N\,P(X > 70)$

$= 3000 \times P\left(\frac{X-\mu}{\sigma} > \frac{70-65}{5}\right)$

$= 3000 \times P(Z > 1)$

$= 3000 \times [1 - P(Z \le 1)] = 3000 \times [1 - 0.8413] = 476.1 \approx 476$

Again, Proportion of 200 students $= \frac{200}{3000} \times 100\% = 6.67\%$

Let $x_1$ be the minimum marks of 200 students.

When $X = x_1$, $Z = \frac{X-\mu}{\sigma} = \frac{x_1 - 65}{5} = z_1$ (say) [$\because z_1$ *lies to the right of mean*]

$P(X \ge x_1) = 6.67\% = 0.0667$

or, $P(Z \ge z_1) = 0.0667$

or, $1 - P(Z < z_1) = 0.0667$

or, $P(Z < z_1) = 1 - 0.0667 = 0.9333$

From normal table A-3, the value of the probability closer to 0.9333 is 0.9332 at $Z = 1.5$. So, $z_1 = 1.5$

So, $\frac{x_1 - 65}{5} = z_1$ $\Rightarrow$ $x_1 = 65 + 5 \times 1.5 = 72.5$

## 4.7 The Normal Distribution and Discrete Populations

The normal distribution is often used as an approximation to the distribution of values in a discrete population. In such situations, extra care should be taken to ensure that probabilities are computed in an accurate manner. Here we use continuity correction.

When we use the normal distribution (which is a *continuous* probability distribution) as an approximation to the ***binomial*** or ***Poisson distribution*** (which are discrete), a continuity correction is made to a discrete whole number $x$ in discrete distribution by representing the single value $x$ by *interval* from $x - 0.5$ to $x + 0.5$ (i.e. adding and subtracting 0.5).

The correction for discreteness of the underlying distribution is often called a *continuity correction*.

### 4.7.1 The Normal Approximation to the Binomial Distribution

The use of binomial distribution becomes very tedious when $n$ is very large. In such case, the normal distribution is used to approximate the binomial distribution. That is, binomial problems can be solved by using normal distribution. Note that for a binomial problem, the exact probability is obtained by using binomial formula. If we apply normal distribution to solve the binomial problem, the probability, we obtain is an approximation to exact probability. The approximation obtained by using the normal distribution is very close to the exact probability when $n$ is very large and $p$ is very close to 0.50. However, it does not mean that we should not use normal approximation when $p$ is not close to 0.50.

**Thus, the normal distribution is a limiting case of the binomial distribution under the following conditions:**

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

(ii) Neither $p$ nor $q$ is very small.

**Theorem 8.1:** If $X$ is a random variable having binomial distribution with parameters $n$ and $p$, the limiting form of the distribution function of the standard random variable

$$Z = \frac{X - np}{\sqrt{npq}},\ q = 1 - p$$

as $n \to \infty$, is given by the standard normal distribution

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

**Note:**

1. If $X$ is a binomial random variable, then $Z = \frac{X - np}{\sqrt{npq}}$ is approximately a standard random variable.
2. *Normal approximation* to the binomial distribution are best only when $np > 15$ *and* $nq > 15$. The *approximation are good for* $np \geq 5$ *and* $nq \geq 5$.

## 4.7.2 The Normal Approximation to the Poisson Distribution

The normal distribution is a limiting case of the Poisson distribution if the average number of occurrences ($\lambda$) is indefinitely large; i.e., $\lambda \to \infty$. So under this case the distribution of standard Poisson variate $Z = \frac{X - \lambda}{\sqrt{\lambda}}$, tends to the distribution of standard normal variate.

If $X$ is a **Poisson random variable** with $E(X) = \lambda$ and $V(X) = \lambda$, then

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$ is approximately a standard random variable.

The approximation is good for $\lambda > 5$.

Since the binomial and Poisson distribution are discrete probability distributions and normal distribution is continuous probability distribution, so while applying the normal approximation as an approximation to binomial and Poisson distribution we always convert discrete distribution to continuous distribution by using continuity correction factors.

## 4.7.3 Continuity correction to improve the Approximation

The addition of 0.5 and/or subtraction of 0.5 from the value(s) of $X$ when the continuous distribution is used as an approximation to the discrete distribution is called ***continuity correction*** and factor 0.5 is called ***continuity correction factor.***

The table below shows how the discrete probability is approximated by normal distribution by making continuity correction

| *Binomial or Poisson Probability* | *Approximated by Normal Distribution* |
|---|---|
| $P(X = 10)$ | $P(10 - 0.5 < X < 10 + 0.5) = P(9.5 < X < 10.5)$ |
| $P(X < 10)$ | $P(X \leq 10 - 0.5) = P(X \leq 9.5)$ |
| $P(X \leq 10)$ | $P(X < 10 + 0.5) = P(X < 10.5)$ |
| $P(X > 10)$ | $P(X \geq 10 + 0.5) = P(X \geq 10.5)$ |
| $P(X \geq 10)$ | $P(X > 10 - 0.5) = P(X > 9.5)$ |
| $P(10 \leq X \leq 20)$ | $P(10 - 0.5 < X < 20 + 0.5) = P(9.5 < X < 20.5)$ |
| $P(10 < X < 20)$ | $P(10 + 0.5 \leq X \leq 20 - 0.5) = P(10.5 \leq X \leq 19.5)$ |

So, (i) $P(X \leq x) = B(x; n, p) = F\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right)$





## Procedure for Using a Normal Distribution to Approximate a Binomial Distribution

1. Establish that the normal distribution is a suitable approximation to the binomial distribution by verifying that $np \geq 5$ and $nq \geq 5$. (If these conditions are not both satisfied, then you must use software, or a calculator, or calculations with the binomial probability formula.)
2. Find the value of the parameters $\mu$ and $\sigma$ by calculating $\mu = np$ and $\sigma = \sqrt{npq}$.
3. Identify the discrete value $X$ (the number of successes). Change the *discrete* value $X$ by replacing it with the *interval* from $X - 0.5$ to $X + 0.5$. (For further clarification, see the discussion in above table). Draw a normal curve and enter the values of $\mu$, $\sigma$, and either $X - 0.5$ or $X + 0.5$, as appropriate.
4. Change $X$ by replacing it with $X - 0.5$ or $X + 0.5$, as appropriate.
5. Using $X - 0.5$ or $X + 0.5$ (as appropriate) in place of $X$, find the area corresponding to the desired probability by first finding the $Z$ score: $Z = (X - \mu)/\sigma$. Now use that $Z$ score to find the area to the left of either $X - 0.5$ or $X + 0.5$, as appropriate. That area can now be used to identify the area corresponding to the desired probability.

We will illustrate this normal approximation procedure with the following example.

**Example 46:** (Loading Airliners: ***normal approximation to binomial distribution***): When an airliner is loaded with passengers, baggage, cargo, and fuel, the pilot must verify that the gross weight is below the maximum allowable limit, and the weight must be properly distributed so that the balance of the aircraft is within safe acceptable limits. Nepal Airlines has established a procedure whereby extra cargo must be reduced whenever a plane filled with 200 passengers includes at least 120 men. Find the probability that among 200 randomly selected passengers, there are at least 120 men. Assume that the population of potential passengers consists of an equal number of men and women.

**Solution:** Now we use *normal distribution* to approximate the *binomial distribution*. We must first verify that it is reasonable to approximate the binomial distribution by the normal distribution because $np \geq 5$ and $nq \geq 5$.

With $n = 200$, $p = 0.5$, and $q = 1 - p = 0.5$,
we verify the required conditions as follows:
$np = 200\ (0.5) = 100$ (Therefore $np \geq 5$).
$nq = 200\ (0.5) = 100$ (Therefore $nq \geq 5$).

$\mu = np = 200\ (0.5) = 100$; $\sigma = \sqrt{npq} = \sqrt{200\ (0.5)\ (0.5)} = 7.0710678$

Let $X$ denote the number of men.

Now, $P(X \geq 120) = P(X > 120 - 0.5) = P(X > 119.5)$ [*Using continuity correction*]

For $X = 119.5$, $Z = \frac{X - np}{\sqrt{npq}} = \frac{119.5 - 100}{7.0710678} = 2.76$

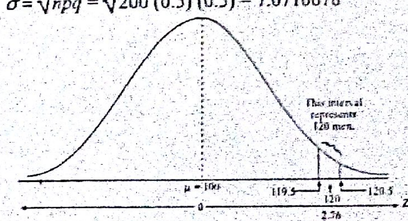



Figure : Finding the probability of "At least" 120 Men Among 200 Passengers

$\therefore$ $P(X \geq 120) = P(X > 119.5) = P(Z > 2.76) = 1 - P(Z \leq 2.76) = 1 - 0.9971 = 0.0029$

**Interpretation:** There is a 0.0029 probability of getting at least 120 men among 200 passengers. Because that probability is so small, we can conclude that a roster of 200 passengers will rarely include at least 120 men, so the reduction of extra cargo is not something to be very concerned about.

**Example 47:** *(Normal approximation to the Binomial distribution)*: If 20% of the memory chips made in a certain plant are defective, what are the probabilities that in a lot of 100 randomly chosen for inspection (a) at most 15 will be defective; (b) exactly 15 will be defective; (c) between 10 and 15 inclusive; (d) less than 32; (e) at least 10 will be defective. [TU, BE, 2063 Ashadh]

**Solution:** Here, $n = 100$, $P$ = Probability of success (i.e., defective) = 0.20

So, $q$ = Probability of failure (non defective) = 0.80

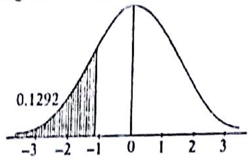
Let $X$ denote the number of successes (getting defective chips), then using normal approximation to the binomial distribution

$\mu = np = 100 \times 0.20 = 20$, $\sigma = \sqrt{10 \times 0.20 \times 0.80} = 4$

(a) $P(X \le 15) = P(X < 15 + 0.5) = P(X < 15.5)$ [Using continuity correction]

For $X = 15.5$, $Z = \dfrac{X - np}{\sqrt{npq}}$

$= \dfrac{15.5 - 20}{4} = -1.25 \approx -1.13$




$\therefore$ $P(X \le 15) = P(X < 15.5)$

$= P(Z < -1.13) = F(-1.13) = 0.1292$

[*From normal table A-3*]

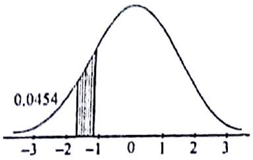
(b) $P(x = 15) = P(14.5 < X < 15.5)$ [*Using continuity correction*]

$= F\left(\dfrac{15.5 - 20}{4}\right) - F\left(\dfrac{14.5 - 20}{4}\right)$

$\left[\because P(a < X < b) = F\left(\dfrac{b - \mu}{\sigma}\right) - F\left(\dfrac{a - \mu}{\sigma}\right)\right]$

$= F(-1.13) - F(-1.38)$

$= 0.1292 - 0.0838 = 0.0454$




(c) Between 10 and 15 inclusive

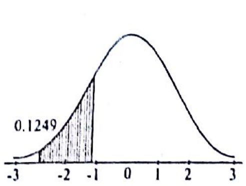
$P(10 \le X \le 15) = P(9.5 < X < 15.5)$

[*Using continuity correction*]

$= P\left(\dfrac{9.5 - 20}{4} < \dfrac{X - 20}{4} < \dfrac{15.5 - 20}{4}\right)$

$= P(-2.63 < Z < -1.13) = F(-1.13) - F(-2.63) =$

$0.1292 - 0.0043 = 0.1249$



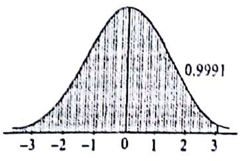

(d) $P(X < 32) = P(X \le 31.5)$

[*Using continuity correction*]

$= P\left(\dfrac{X - 20}{4} \le \dfrac{32.5 - 20}{4}\right)$

$= P(Z \le 3.13) = F(3.13)$

$= 0.9991$



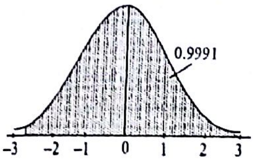

(e) $P(X \ge 10) = P(X > 9.5)$

[*Using continuity correction*]

$= P\left(\dfrac{X - 20}{4} \le \dfrac{9.5 - 20}{4}\right)$

$= P(Z > -2.63) = 1 - P(Z \le -2.63)$

$= 1 - F(-2.63) = 1 - 0.0043 = 0.9957$







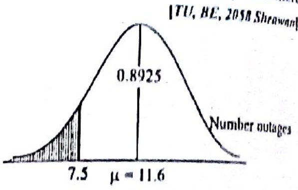

**Example 48:** *(Normal approximation to a count variable)*.

In a certain city, the number of power outages per month is a random variable, having a distribution with $\mu = 11.6$ and $\sigma = 3.3$. If this distribution can be approximated closely with a normal distribution, what is the probability that there will be at least 8 outages in any one month? [TU, BE, 2058 Shrawan]

**Solution:** Let $X$ be *rv* denoting the number of outages in a month with $\mu = 11.6$ and $\sigma = 3.3$ and $Z$ be its *SNV*.

The desire probability is approximated by

$P(X \ge 8) = P(X > 7.5)$ [*Using continuity correction*]




$= P\left(\dfrac{X - 11.6}{3.3} > \dfrac{7.5 - 11.6}{3.3}\right)$

$= P(Z > -1.24) = 1 - P(Z \le -1.24)$

$= 1 - F(-1.24) = F(1.24) = 0.8925$

**Example 49:** *(Normal approximation to the Binomial distribution)*: Suppose that 25% of all licensed drivers in Kathmandu do not have insurance. Let $X$ be the number of uninsured drivers in a random sample of size 50 (somewhat perversely, a success is an uninsured driver), so that $p = 0.25$.

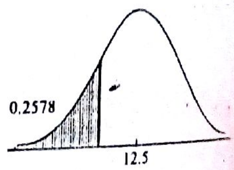
Find (a) $P(X \le 10)$; (b) $P(5 \le X \le 15)$

**Solution:** Here, $n = 50$, $p = 0.5$

So, $\mu = np = 12.5$, $\sigma = 3.06$, $nq = 37.5$

We know $P(X \le x)$

$= B(x; n, p) \approx F\left(\dfrac{x + 0.5 - np}{\sqrt{npq}}\right)$




So, $P(X \le 10)$

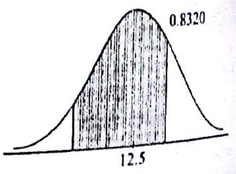
$= B(10; 50, 0.25) \approx F\left(\dfrac{11.5 - 12.5}{3.06}\right)$

$= F(-0.65) = 0.2578$

Similarly, $P(5 \le X \le 15)$

$= B(15; 50, 0.25) - B(4; 50, 0.25)$

$= F\left(\dfrac{15.5 - 12.5}{3.06}\right) - F\left(\dfrac{4.5 - 12.5}{3.06}\right)$




$= F(0.98) - F(-2.61) = 0.8365 - 0.0045 = 0.8320$

Here the exact probability are 0.2622 and 0.8348 respectively, so the approximations are quite good.

**Example 50:** *(Normal approximation to the binomial distribution)*: A safety engineer feels that 30% of all industrial accidents in her plant are caused by failure of employees to follow instruction. If this figure is correct, find, approximately, the probability that among 84 industrialized accidents in this plant anywhere from 20 to 30 (inclusive) will be due to failure of employees to follows instruction. [TU, BIE, 2065 Chaitra/ 2066 Magh]

**Solution:** Here, $n = 84$, $p$ = Prob. of accidents = 30% = 0.30

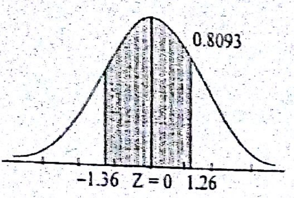
So, $q = 1 - p = 0.70$

Let $X$ denote the number of industrial accidents, then using normal approximation to the binomial distribution

$\mu = np = 84 \times 0.30 = 25.2$

$\sigma = \sqrt{npq} = \sqrt{84 \times 0.30 \times 0.70} = 4.2,$

$nq = 84 \times 0.70 = 58.8$




Now, $P(20 \le X \le 30) = P(19.5 < X < 30.5)$

*[Using continuity correction]*

$$= P\left(\frac{19.5 - 25.2}{4.2} < \frac{X - 25.2}{4.2} < \frac{30.5 - 25.2}{4.2}\right)$$

$$= P(-1.36 < Z < 1.26) = F(1.26) - F(-1.36) = 0.8962 - 0.0869 = 0.8093$$

**Example 51:** (*Normal approximation to the binomial distribution*): A manufacturer knows that, an average, 2% of the electronic toasters that he makes will require repairs within 90 days after they are sold. Use the normal approximation to the binomial distribution to determine the probability that among 1,200 of these toasters at least 30 will required repairs within the first 90 days after they are sold.

**Solution:** Here, $n = 1{,}200$; $p = 2\% = 0.02$, $q = 0.98$

Let $X$ denote the number of toasters that require repairs within 90 days. Then using normal approximation to the binomial distribution

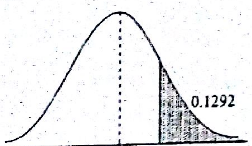
$\mu = np = 1{,}200 \times 0.02 = 24,\ nq = 1176$

$\sigma = \sqrt{npq} = 4.85$

Now, $P(X \ge 30) = P(X > 29.5)$

*[Using continuity correction]*

For $X = 29.5$, $Z = \dfrac{X - np}{\sqrt{npq}} = \dfrac{29.5 - 24}{4.85} = 1.13$




Therefore, $P(X \ge 30) = P(X > 29.5)$

$= P(Z > 1.13) = 1 - F(1.13) = 1 - 0.8708 = 0.1292$

So, required probability that the at least 30 toasters will require repair within the first 90 days = 0.1292.

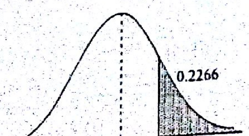
**Example 52:** (*Normal approximation to the Poisson distribution*): The number of accidents occurring on average each year in a factory is 36. They occur completely at random. Use the normal approximation to the Poisson distribution to find the probability that in 2000 there were more than 40 accidents.

**Solution:** Here, $\lambda$ = average no. of accidents each year = 36

Let $X$ denote the number of accidents in each year, then, using normal approximation to the Poisson distribution

$\mu = \lambda = 36, \quad \sigma = \sqrt{\lambda} = 6,$

so $\quad Z = \dfrac{X - \lambda}{\sqrt{\lambda}} = \dfrac{X - 36}{6}$




Therefore, $P(X > 40) = P(X \ge 40.5)$

*[Using continuity correction]*

$$= P\left(\frac{X - \lambda}{\sqrt{\lambda}} \ge \frac{40.5 - 36}{6}\right) = P(Z \ge 0.75)$$

$$= 1 - P(Z < 0.75) = 1 - 0.7734 = 0.2266.$$



**Example 53:** Assume that the number of asbestos particles in a squared meter of dust on a surface follows a Poisson distribution with a mean of 1000. If a squared meter of dust is analyzed, what is the probability that less than 950 particles are found?

**Solution:** This probability can be expressed exactly as

$$P(X \le 950) = \sum_{x=0} (e^{-1000}\, x^{1000})/x!$$

The computational difficulty is clear. The probability can be approximated as

$$P(X \le x) = P\left(Z \le \frac{950 - 1000}{\sqrt{1000}}\right) = P(Z \ge 0.75) = P(Z \le -1.58) = 0.057.$$

## Some more solved examples:

**Example 54:** (*Designing Cars*): The sitting height (from seat to top of head) of drivers must be considered in the design of a new car model. Men have sitting heights that are normally distributed with a mean of 36.0 in. and a standard deviation of 1.4 in. Engineers have provided plans that can accommodate men with sitting heights up to 38.8 in., but taller men cannot fit. If a man is randomly selected, find the probability that he has a sitting height less than 38.8 in., but taller men cannot fit. If a man is randomly selected, find the probability that he has a sitting height less than 38.8 in. Based on that result, is the current engineering design feasible?

**Solution:** Let $X$ denote the height of the drivers.

The height of 38.8 in. is converted to a $Z$ score as follows:

$$Z = \frac{X - \mu}{\sigma} = \frac{38.8 - 36.0}{1.4} = 2.00$$

This result shows that the sitting height of 38.8 in. is above the mean of 36.0 in by 2.00 standard deviations.

Referring to normal table we find that $Z = 2.00$ corresponds to an area of 0.9772.

So, $P(X < 38.8 \text{ in.}) = P(Z < 2.00) = 0.9772$

**Interpretation:** There is a probability of 0.9772 of randomly selecting a man with a sitting height less than 38.8 in. Another way to interpret this result is to conclude that 97.72% of men have sitting heights less than 38.8 in. An important consequence of that result is that 2.28% of men will not fit in the car. The manufacturer must now decide whether it can afford to lose 2.28% of all male car drivers.

**Example 55:** (*Jet Ejection Seats*): The Air Force had been using the ACES-II ejection seats designed for men weighing between 140 *lb* and 211 *lb*. Given that women's weights are normally distributed with a mean of 143 *lb* and a standard deviation of 29 *lb* (based on data from the National Health Survey), what percentage of women have weights that are within those limits?

**Solution:** Let $X$ denote the weight of women.

When $X = 140$, $Z = \dfrac{X - \mu}{\sigma} = \dfrac{140 - 143}{29}$

When $X = 211$, $Z = \dfrac{X - \mu}{\sigma} = \dfrac{211 - 143}{29} = 2.34$

Now $P(140 < X < 211) = P(-0.10 < Z < 2.34) = 0.5302$

**Interpretation:** We found that 53.02% of women have weights between the ejection seat limits of 140 *lb* and 211 *lb*. This means that 46.98% of women do not have weight between the current limits, so far too many women pilots would risk serious injure if ejection became necessary.

**Example 56:** The electric light tubes of manufacturer $A$ have mean lifetime of 1400 hours with *s.d.* of 200 hours, while those of manufactures $B$ have mean lifetime of 1200 hours with *s.d.* of 100. If random sample of 125 tubes of each brand are tested, what is the probability that the brand $A$ tubes will have a mean lifetime which is at least (a) 160 hrs more than the brand $B$ tubes and (b) 250 hrs more than the brand $B$ tubes.

**Solution:** $x_1 = 1200 + 160 = 1360$, $x_2 = 1200 + 250 = 1450$

$\mu = 1400$, $\sigma = 200$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{1360 - 1400}{200} = -0.2.$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{1450 - 1400}{200} = 0.25.$$

(a) $P(Z \geq -0.2) = 1 - P(Z < -0.2) = 1 - F(-0.2) = 1 - 0.4207 = 0.5793.$

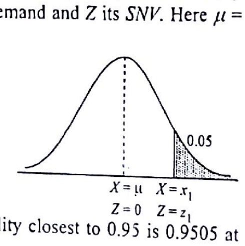
(b) $P(Z \geq 0.25) = 1 - P(Z < 0.25)$
$= 1 - F(0.25) = 1 - 0.5987 = 0.4013.$

**Example 57:** Lead time demand $X$ for an item is approximated by a normal distribution with mean of 25 and *s.d.* of 3. Find the value of lead - time that will exceed only 5% of the time. [TU, BE, 2064 2064 Poush]

**Solution:** Let $X$ denote the value of the lead- time demand and $Z$ its *SNV*. Here $\mu = 25$, $\sigma = 5$. Now $P(X > x_1) = 0.05$

Since, $P(X > x_1) < 0.5$, $x_1$ lies to the right of mean and the corresponding value of $Z$ is positive.

When $X = x_1$, $Z = \frac{X - \mu}{\sigma} = \frac{X - 25}{3} = z_1$ (say)




Therefore, $P(X > x_1) = P(Z > z_1) = 0.05$

or, $1 - P(Z \leq z_1) = 0.05$

or, $F(z_1) = 0.95$

From the normal table A-3, the value of the probability closest to 0.95 is 0.9505 at $Z = 1.65$. So, $z_1 = 1.65$.

Hence, $\frac{x_1 - \mu}{\sigma} = z_1 \Rightarrow x_1 = \mu + \sigma z_1 = 25 + 3 \times 1.65 = 29.95 \approx 30.$

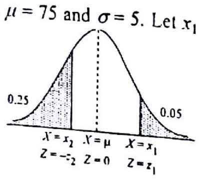
So, the required value of the time lead = 30.

**Example 58:** A set of examination marks is approximately normally distributed with $\mu = 75$ and $\sigma = 5$. If the top 5% of students get grade $A$ and the bottom 25% get grade $F$, what mark is the lowest $A$ and what mark is the highest $F$.

**Solution:** Let $X$ denote the marks in the examination with $\mu = 75$ and $\sigma = 5$. Let $x_1$ be the lowest marks for grade $A$ and $x_2$ be the highest marks for grade $F$. Then by question.

$P(X > x_1) = 0.05$ and $P(X < x_2) = 0.25$

Here, $P(X > x_1) < 0.05$, so $x_1$ lies to the right of mean and the corresponding value of $Z$ is positive.




Therefore, $Z = \frac{X - \mu}{\sigma} = \frac{x_1 - 75}{5} = z_1$(say) . . . (i)

$P(X > x_1) = 0.05 \Rightarrow P(Z > z_1) = 0.05$

or, $1 - P(Z \leq z_1) = 0.05$. So, $F(z_1) = 0.95$

From the normal table, the value of the probability closest to 0.95 is 0.9505 at $z = 1.65$. So $z_1 = 1.65$





From (i) $\frac{x_1 - 75}{5} = z_1 \Rightarrow x_1 - 75 = 8.25$

$\therefore$ $x_1 = 75 + 8.25 = 83.25 \approx 83.$

Again $P(X < x_2) = 0.25$

Here, $P(X < x_2) < 0.5$, so $x_2$ lies to the left of mean, so the corresponding value of $Z$ is negative.

Therefore, $Z = \frac{X - \mu}{\sigma} = \frac{x_2 - 75}{5} = -z_2$ (say) . . . (ii)

So, $P(X < x_2) = 0.25 \Rightarrow P(Z < -z_2) = 0.25 \Rightarrow F(-z_2) = 0.25$

From the normal table A-3, the value of the probability closest to 0.25 is 0.2514 at $Z = 0.67$. So, $z_2 = 0.67$.

From (ii) $\frac{x_2 - 75}{5} = -z_2 \Rightarrow x_2 = 75 - 5 \times 0.67 = 71.65 \approx 72$

Hence the lowest mark for grade $A$ is 83 and the highest mark for grade $F$ is 72.

**Example 59:** An industrial engineer has found that the standard household light bulbs produced by a certain manufacturer have a useful life that is normally distributed with a mean of 250 hours and a variance of 2500. What is the probability that a randomly selected bulb from this production process will have a useful life (a) in excess of 300 hours; (b) between 190 and 270 hours; (c) not exceeding 260 hours? [TU, BE 2064 Poush]

**Solution:** (a) $P(X > 300) = P(Z > 1) = 1 - F(1) = 1 - 0.8413 = 0.1587$

(b) $P(190 < X < 270) = P(-1.2 < Z < 0.4) = F(0.4) - F(-1.2) = 0.5403.$

(c) $P(X \leq 260) = P(Z \leq 0.2) = F(0.2) = 0.5973.$

**Example 60:** An auditor has reviewed the financial records of a hardware store and has found that its billing errors follow a normal distribution with mean \$1.5 and standard deviation \$1. [TU, BE, 2066 Magh]

(a) What proportion of the store's billings are in error by more than \$1?

(b) What is the probability that a billing represents an overcharge of at least \$1.50?

(c) What is the probability that a customer has been undercharged from \$0.50 to \$1.00?

**Solution:**

(a) $P(X > 1) = P(Z > 0.5) = 1 - F(-0.5) = 1 - 0.3085 = 0.6915 = 69.15\%$

(b) $P(X \geq 1.50) = P(Z \geq 0) = 0.5$

(c) $P(0.50 < X < 1) = P(-1 < Z < -0.5)$
$F(-0.5) - F(-1) = 0.3085 - 0.1587 = 0.1498.$

**Example 61:** The life-time in hours of certain electrical equipment has the normal distribution with mean 80 and standard deviation 16.

(a) What is the probability that the equipment lasts at least 100 hours?

(b) If the equipment has already lasted 88 hours, what is the conditional probability that it will last at least another 12 hours?

**Solution:** Let $X$ denote the life of electric equipment in hours. Then $X \sim N(\mu, \sigma^2)$ where $\mu = 80$, $\sigma = 16$.

(a) $P(X \geq 100) = P(Z \geq 1.25) = 1 - P(Z < 1.25) = 1 - 0.8944 = 0.1056$ . . . (1)

(b) $p = P(X \geq (88 + 12) \mid X \geq 88\} = P(X \geq 100 \mid X \geq 88)$

$$= \frac{P(X \geq 100 \cap X \geq 88)}{P(X \geq 88)} = \frac{P(X \geq 100)}{P(X \geq 88)} \quad \ldots (2)$$

Also, $P(X \geq 88) = P(Z \geq 0.5) = 1 - P(Z < 0.5) = 1 - 0.6915 = 0.3085$ . . . (3)

Hence, $P = \dfrac{P(X \geq 100)}{P(X \geq 88)} = \dfrac{0.1056}{0.3085} = 0.3423$

**Example 62:** Let $X$ be a normal variable with mean $\mu$ and standard deviation $\sigma$. If $Z$ is the standard normal variable such that $Z = -0.8$ when $X = 26$ and $Z = 2$ when $X = 40$, then find $\mu$ and $\sigma$. Also, find $P(X > 45)$ and $P(|X - 30| > 5)$.

**Solution:** We know that $Z = \dfrac{X - \mu}{\sigma}$. It is given that $Z = \dfrac{26 - \mu}{\sigma} = -0.8$

or, $-0.8\sigma + \mu = 26$ and $Z = \dfrac{40 - \mu}{\sigma} = 2$ or $2\sigma + \mu = 40$

Solving the above two equations, we get $\mu = 30$ and $\sigma = 5$.

Now, consider

$$P(X > 45) = P\left(\frac{X - 30}{5} > \frac{45 - 30}{5}\right) = P(Z > 3) = 0.00135$$

Similarly, we get

$$P(|X - 30| > 5) = 1 - P(|X - 30| \leq 5) = 1 - P(-5 \leq X - 30 \leq 5)$$

$$= 1 - P(25 \leq X \leq 35) = 1 - P\left(\frac{25 - 30}{5} \leq \frac{X - 30}{5} \leq \frac{35 - 30}{5}\right)$$

$$= 1 - P(-1 \leq Z \leq 1) = 1 - [P(Z \leq 1) - P(Z \leq -1)]$$

$$= 1 - [1 - F(-1)] = 1 - (0.8413 - 0.158) = 0.3168$$

**Example 63:** (*Hip Breadths and Airplane seats*): In designing seats to be installed in commercial aircraft, engineers want to make the seats wide enough to fit 98% of all males. (Accommodating 100% of males would require very wide seats that would be much too expensive) Men have hip breadths that are normally distributed with a mean of 14.4 in. and a standard deviation of 10 in. Find $P_{98}$. That is, find the hip breadth of men that separates the bottom (no pun intended) 98% from the top 2%.

**Solution:** Let $X$ denote the breadth of hip of men.

The area closest to 0.98 as 9.9798, and it corresponds to a $Z$ score of 2.05.

With $z = 2.05$, $\mu = 14.4$, and $\sigma = 1.0$, we find

$$x = \mu + (z \cdot \sigma) = 14.4 + (2.05)(1.0) = 16.45.$$

If we let $x = 16.45$ we see that this solution is reasonable because the 98th percentile should be greater than the mean of 14.4.

**Interpretation:** The hip breadth of 16.5 in. (rounded to one decimal place, as in $\mu$ and $\sigma$) separates the lowest 98% from the highest 2%. That is, seats designed for a hip breadth up to 16.5 in. will fit 98% of men. This type of analysis was used to design the seats currently used in commercial aircraft.

## 4.8 The Exponential and Gamma Distributions

There are many practical situations in which the variable of interest might have a skewed distribution. One family of distributions that has property is the gamma family.

### 4.8.1 The Exponential Distribution

**Definition:** A continuous random variable $X$ is said to have an *exponential distribution* with parameter $\lambda > 0$, if the *pdf* of $X$ is

$$f(x) = f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

**Note:**

1. Some sources write $f(x) = \dfrac{1}{\beta} e^{-x/\beta}$ so that $\lambda = \dfrac{1}{\beta}$.



2. Above *pdf* is eligible because
   (i) $f(x) \geq 0$ for all $x$
   (ii) $\int_{-\infty}^{\infty} f(x)\,dx = \lambda \int_{-\infty}^{\infty} e^{-\lambda x}\,dx = -1\,[e^{-\lambda x}]_0^{\infty} = 1.$

3. *Mean* of $X = E(X) = \int_0^{\infty} x\,\lambda\,e^{-\lambda x}\,dx = \lambda \int_0^{\infty} x^{2-1} e^{-\lambda x} = \dfrac{\lambda\,\Gamma(2)}{\lambda^2} = 1$

   [Using $\int_0^{\infty} x^{\alpha-1} e^{-ax} = \dfrac{\Gamma(\alpha)}{a^{\alpha}} = \dfrac{1}{a}$

   *Variance* of $X = V(X) = E(X^2) - [E(X)]^2$

   $= \int_0^{\infty} x^2 \lambda e^{-\lambda x}\,dx - \dfrac{1}{\lambda^2} = \lambda \int_0^{\infty} x^{3-1} e^{-\lambda x}\,dx - \dfrac{1}{\lambda^2} = \dfrac{\lambda\,\Gamma(3)}{\lambda^3} - \dfrac{1}{\lambda^2} = \dfrac{2}{\lambda^2} - \dfrac{1}{\lambda^2} = \dfrac{1}{\lambda^2}$

   *Cumulative distribution function $F(x)$ of $X$ is*

   $F(x) = P(X \leq x) = \int_0^{\infty} \lambda e^{-\lambda x}\,dx = 1 - e^{-\lambda x}, x \geq 0$

   So, $F(x) = F(x; \lambda) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-\lambda x} & \text{for } x \geq 0 \end{cases}$

4. Graphs of several exponential *pdf*s are:

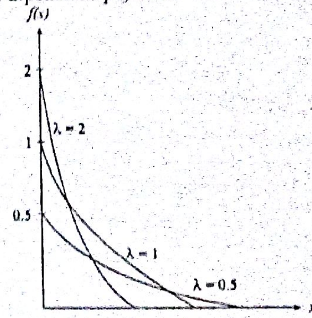



Figure : Exponential density curves

5. This distribution is used as a model for the distribution of times between the occurrence of successive events, such as the waiting time between successive arrivals (successes); the amount of time (starting from now) until an earthquake occurs, or until a new war breaks out, or until a telephone call you receive turns out to be a wrong number. It is closely related to Poisson process. Another application of this distribution is to provide appropriate model for calculating probabilities concerning the "*lifetime*" of a given piece of equipment. The exponential distribution *'lacks memory'* (or *it is memory less*).

**Example 64:** Suppose the response time $X$ at a certain on-line computer terminal (the elapsed time between the end of a user's inquiry and the beginning of the system's response to that inquiry has an exponential distribution with expected response time equal to 5 seconds. Find the probability that the response time is (a) at most 10 seconds; (b) between 5 and 10 seconds.

**Solution:** $X$ = the response time. $E(X) = \dfrac{1}{\lambda} = 5 \Rightarrow \lambda = 0.2$

(a) $P(X \leq 10) = F(10; 2) = 1 - e^{-(0.2)(10)} = 1 - e^{-2} = 1 - 0.1353 = 0.8647$

(b) $P(5 \leq X \leq 10) = F(10; 2) - F(5; 2) = (1 - e^{-2}) - (1 - e^{-1}) = 0.2325$

**Example 65:** Suppose that during rainy season on a tropical island the length of shower has an exponential distribution, with parameter $\lambda = 2$ minutes. What is the probability that a shower will last – more than three minutes?

**Solution:** Let $X$ denote the length of shower. Here $\lambda = 2$.

$$P(X \geq 3) = 1 - P(X < 3) = 1 - \int_0^3 2e^{-2x}\,dx = 1 + [e^{-2x}]_0^3 = e^{-6} = 0.00248$$

or, $P(X \geq 3) = 1 - P(X < 3) = 1 - F(3; 2) = 1 - [1 - e^{-3\times 2}] = e^{-6} = 0.00248$

**Example 66:** Suppose that a number of miles that a car can run before its battery wears out are exponentially distributed with an average value of 10,000 miles. If a person desires to take a 5000 mile trip, what is the probability that she will able to complete her trip without having to replace her car battery?

**Solution:** Let $X$ denote the remaining lifetime (in thousands of miles) of the battery.

Then $X$ is exponential with parameter $\lambda = \frac{1}{10}$. Hence the desired probability is

$$P(X > 5) = 1 - F(5; 0.1) = 1 - 1 + e^{-5(0.1)} \left[\because \mu = \frac{1}{\lambda} = 10 \Rightarrow \lambda = \frac{1}{10}\right]$$

$$= e^{-0.5} = 0.6065$$

**Example 67:** Suppose that the service life of a semiconductor is exponentially distributed with an average value of 50 hours. Find the probability that such a semiconductor will still be in operating condition after 100 hours?

**Solution:** Let $X$ denote the service life of a semiconductor with average life $\mu = 50$ hours. Now, $\mu = \frac{1}{\lambda} = 50 \Rightarrow \lambda = \frac{1}{50}$

*pdf* of $X$ is $f(x) = \lambda e^{-\lambda x}$, $\lambda > 0, x \geq 0$.

The required probability is $\quad P(X \geq 100) = 1 - P(X < 100)$

$$= 1 - \left[1 - F\left(100; \frac{1}{50}\right)\right] = 1 - 1 + e^{-2} = 0.1353$$

### 4.8.2 Gamma Distribution

**Definitions:** A continuous random variable $X$ is said to have a *gamma distribution* with parameters $\alpha$ and $\beta$ if it assumes only non-negative values and the *pdf* of $X$ is

$$f(x) = f(x; \alpha, \beta) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & \text{for } x > 0,\ \alpha > 0,\ \beta > 0 \\ 0 & \text{otherwise} \end{cases}$$

Any continuous random variable with above *pdf* is called ***gamma variable*** and probability distribution of this variable is called ***gamma distribution***. The parameter $\alpha$ is called ***shape parameter*** and $\beta$ is called ***scale parameter***.

**Remarks:**

1. (i) $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\,dx$. $\alpha > 0$ is gamma function.

   (ii) $\dfrac{\Gamma(\alpha)}{a^\alpha} = \int_0^\infty e^{-ax} x^{\alpha-1}\,dx$

   (iii) $\Gamma(\alpha) = (\alpha - 1)\,\Gamma(\alpha - 1) = (\alpha - 1)!$ and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

2. Above *pdf* is eligible because

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

   (ii) $\int_0^\infty f(x)\,dx = \int_0^\infty \dfrac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}\,dx$

   $= \dfrac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-x/\beta}\,dx = \dfrac{1}{\beta^\alpha \Gamma(\alpha)} \Gamma(\alpha)\,\beta^\alpha = 1$





3. If $\beta = 1$, the gamma distribution of $X$ with only parameter $\alpha$ is

$$f(x) = \begin{cases} \dfrac{e^{-x} x^{\alpha-1}}{\Gamma(\alpha)} & \text{for } x > 0 \quad \alpha > 0 \\ 0 & \text{otherwise} \end{cases}$$

4. If $\alpha = 1$ and $\beta = \frac{1}{\lambda}$, then an exponential distribution is obtained.
5. The application of this distribution is to provide appropriate model for calculating probabilities concerning the length of life of industrial equipments, electrical supply in certain area, distribution of petrol etc. gamma distribution is also applied to waiting time models in life testing, waiting time until death etc.
6. *Mean and variance of gamma distribution are*

   $E(X) = \mu = \dfrac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x\, x^{\alpha-1} e^{-x/\beta}\,dx = \alpha\beta$

   $V(X) = \sigma^2 = \alpha\beta^2$.

7. Graph of some gamma probability density functions is given below:

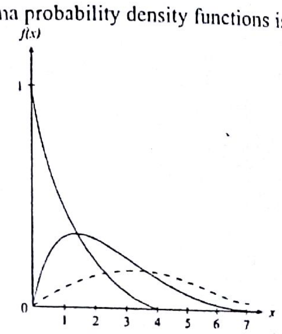


Figure: Gamma distribution

8. *Gamma distribution* tends to *Normal distribution* as $\lambda = \frac{1}{\beta} \to \infty$.
9. It is highly skewed distribution.
10. If $X$ is a gamma variable then $X \pm a$, where $a$ is constant, is also gamma variable.

**Example 68:** In Kathmandu city, the daily consumption of electric power (in million of kilowatt-hours) can be treated as a random variable having a gamma distribution with $\alpha = 3$ and $\beta = 2$. If the power plant of this city has daily capacity of 12 million kilowatt hours, what is the probability that this power supply will be inadequate on any given day?

[TU, BIE 2064 Shrawan]

**Solution:** Let $X$ denote the daily consumption of power (in million of kilowatt-hours). Then $X$ follows gamma distribution with parameters $\alpha = 3$ and $\beta = 2$. Then *pdf* of $X$ is:

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} = \frac{1}{2^3\,\Gamma(3)} x^2 e^{-x/2} = \frac{x^2 e^{-x/2}}{16},\ x > 0$$

The probability that the power supply will be inadequate on any given day is

$$P(X > 12) = 1 - P(X \leq 12) = 1 - \int_0^{12} \frac{1}{16} x^2 e^{-x/2}\,dx$$

$$= 1 - \frac{1}{16}\left[-2x^2 e^{-x/2} - 8x e^{-x/2} - 16e^{-x/2}\right]_0^{12}$$

$= 1 - \frac{e^{-3}}{[illegible]}(-288 - 96 - 16) - 1 = 0.0619$

**Example [illegible]:** In a certain city the daily consumption of water (in millions of gallons) follows approximately a gamma distribution with $\alpha = 2$ and $\beta = 3$. If the daily capacity of this city is 9 million gallons of water, show that the probability that on any given day the water supply is inadequate is $4e^{-3}$. [Pokhara Uni BE 2007]

**Solution:** Let X denote the daily consumption of water (in millions of gallons). Then X follows gamma distribution with parameters $\alpha = 2$ and $\beta = 3$.

So, pdf of X is

$$f(x) = \frac{1}{3^2\,\Gamma(2)} x^{2-1} e^{-x/3} = \frac{1}{9} x e^{-x/3},\ x > 0$$

The probability that on any given day the water supply is inadequate is

$$P(X > 9) = 1 - P(X \le 9) = 1 - \int_0^9 \frac{1}{9} x e^{-x/3}\,dx$$

$$= 1 - \frac{1}{9}\left[-27e^{-3} - 0 - 9e^{-3} + 9\right]$$

$$= 1 + 3e^{-3} + e^{-3} - 1 = 4e^{-3} = 4 \times 0.0498 = 0.1992$$

**Example [illegible]:** Suppose the survival time X in weeks of a randomly selected male mouse exposed to 240 rads of gamma radiation has a gamma distribution with $\alpha = 2$, $\beta = 15$. (a) Find the expected time and variance.

(b) Find the probability that a mouse survives between 10 and 30 weeks.

**Solution:** Here X denote the survival time in weeks.

(a) X is gamma distribution with $\alpha = 2$, $\beta = 15$

pdf of X is $f(x) = \frac{1}{(15)^2\,\Gamma(2)} x^{2-1} e^{-x/15} = \frac{1}{(15)^2\,\Gamma(2)} x e^{-x/15}$

(a) Expected time $E(X) = \alpha\beta = 2 \times 15 = 30$ weeks

variance $V(X) = \alpha\beta^2 = (2)(15)^2 = 450$ weeks

$\sigma = \sqrt{450} = 21.21$ weeks.

(b) Probability that a mouse survives between 10 and 30 weeks is

$$P(10 \le X \le 30) = \int_{10}^{30} f(x)\,dx = \frac{1}{(15)^2\,\Gamma(2)} \int_{10}^{30} x e^{-x/15}\,dx$$

$$= \frac{1}{450}\left[-15 x e^{-x/15} - 15 \times 15 e^{-x/15}\right]_{10}^{30}$$

$$= \frac{-15}{450}\left[30e^{-2} - 10e^{-10/15} + 15e^{-2} - 15e^{-10/15}\right]$$

$$= \frac{-1}{30}\left[45e^{-2} - 25e^{-10/15}\right] = [illegible]$$

**Example [illegible]:** The daily consumption of milk in Kathmandu valley in excess 2,00,000 liters is approximately distributed as gamma variate with parameter $\alpha = 2$ and $\beta = 10^5$. The valley has a daily stock of milk 3,00,000 liters. What is the probability that the stock is insufficient on a particular day?

**Solution:** Let X denote the daily consumption of milk in Kathmandu valley. Then the rv $Y = X - 2{,}00{,}000$ has also a gamma distribution with the pdf

$$g(y) = \frac{1}{\beta^\alpha\,\Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta} = \frac{1}{(10^5)^2\,\Gamma(2)} y^{2-1} e^{-y/10^5} = \frac{1}{(10^5)^2} y e^{-y/10^5}$$

The probability that the daily milk supply is insufficient is

$P(X > 3{,}00{,}000) = P(Y + 2{,}00{,}000 > 3{,}00{,}000) = P(Y > 1{,}00{,}000)$

$$= \int_{100000}^{\infty} g(y)\,dy = \int_{100000}^{\infty} \frac{1}{(10^5)^2} y e^{-y/10^5}\,dy$$



$$= \frac{1}{(10^5)^2}\left[-10^5 y e^{-y/10^5}\right]_{100000}^{\infty} + \frac{1}{(10^5)^2}\left[-(10^5)^2 e^{-y/10^5}\right]_{100000}^{\infty}$$

$= e^{-1} + e^{-1} = 0.7358$

## 4.9 Beta Distribution

**Definition:** A continuous random variable X is said to have a *beta distribution* with parameters $\alpha$ and $\beta$ if pdf of X is

$$f(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } 0 < x < 1,\ \alpha > 0,\ \beta > 0 \\ 0 & \text{otherwise} \end{cases}$$

**Remarks:**

1. Random Variable X takes on values on the interval from 0 to 1.
2. $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$.
3. Since $f(x) \ge 0$ for all $x \in (0, 1)$

   and $\int_0^1 f(x)\,dx = \int_0^1 \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx$

   $= \frac{B(\alpha,\beta)}{B(\alpha,\beta)} = 1$, so above pdf is eligible.
4. Mean and variance of this distribution are

   $\mu = \frac{\alpha}{\alpha+\beta}$, $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

**Example [illegible]:** In a certain country, the proportion of highway sections requiring repairs in any given year is a random variable having the beta distribution with $\alpha = 3$ and $\beta = 2$.

(a) On the average, what percentage of the highway sections require repairs in any given year?

(b) Find the probability that at most half of the highway sections will require repairs in any given year.

**Solution:** Let a random variable X denote the highway sections in any year with

$\alpha = 3,\ \beta = 2.$

Then (a) $\mu = \frac{\alpha}{\alpha+\beta} = \frac{3}{3+2} = \frac{3}{5} = 0.60 = 60\%$

So on the average 60% of the highway section require repairs in any given year.

(b) $\Gamma(\alpha+\beta) = \Gamma(5) = 4!;\ \Gamma(\alpha) = \Gamma(3) = 2!,\ \Gamma(\beta) = \Gamma(2) = 1!$

The beta distribution is

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

So, the required probability is

$$P\left(X \le \frac{1}{2}\right) = \int_0^{1/2} \frac{4!}{2!\,1!} x^{3-1}(1-x)^{2-1}\,dx$$

$$= \int_0^{1/2} 12x^2(1-x)\,dx = \frac{5}{16} = 0.3125$$

## 4.10. Chi-squared ($\chi^2$) Distribution

**Definition:** Let ν be a positive integer. Then a continuous random variable X is said to have a *chi-squared distribution* with parameter ν (nu) if the pdf of X is the gamma density with $\alpha = \frac{\nu}{2}$ and $\beta = 2$. The pdf of a chi-squared rv X is thus,

$$f(x\,;\,\nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2} & \text{for } x \ge 0 \\ 0 & \text{for } x < 0 \end{cases}$$

The parameter $\upsilon$ is called the number of ***degrees of freedom*** (*d.f.*) of *X*. The symbol $\chi^2$ is often used in place of "chi-squared". $\chi$ (chi) is Greek alphabet.

### 4.10.1 Degrees of freedom

**Definition:** ***Degrees of freedom*** is the number of independent observations in a set. It is denoted by the Greek alphabet ν(*nu*).

**For example:** If *k* is the number of independent constraints in a set of data of *n* observations then the degree of freedom is $\upsilon = n - k$.

So, degree of freedom is defined as the number of observations (or values) in a sample that we can choose freely. Suppose, we know the mean of four values is 30. Then sum of these four values is 120.

Say, $x_1 + x_2 + x_3 + x_4 = 120$

Out of these four, we can choose freely only three values. If we choose $x_1 = 50$, $x_2 = 40$, $x_3 = 10$, the fourth value $x_4$ is automatically determined i.e.,

$x_4 = 120 - (50 + 40 + 10) = 20$. So

$d.f. = \nu = 4 - 1 = 3$. Hence

For a sample of size 5, $d.f. = 5 - 1 = 4$

For a sample of size 7, $d.f. = 7 - 1 = 6$

For a sample of size *n*, $d.f. = \nu = n - 1$.

## Exercise 4

### Theoretical Questions

1. What do you understand by probability mass function and probability density function? Explain. [T.U. BE 2068 Bhadra]
2. Define probability Density function and its probability distribution. Give three engineering examples of discrete case. [TU, BE, 2067 Mangsir;]
3. Distinguish between probability mass function and probability density function with one example each. [TU, BE, 2067 Mangsir]
4. Define the random variable. Define the continuous random variable and its probability function and probability distribution function. [TU, BE, 2062 Baishakh]
5. Define continuous random variable and standard normal variate. [TU, BE, 2056 Bhadra]
6. Describe the conditions for probability mass function and probability density function. [TU, BE, 2064 Poush]
7. Find the distribution function of a random variable which has the uniform distribution.
8. Write down the four important properties of normal distribution. [T.U. BE 2068 Bhadra]
9. Discuss the area property of Normal distribution. [TU, BE, 2061 Ashwin/ 2064 Shrawan/ 2067 Mangsir:]
10. Define the normal distribution. Given condition for normal approximation of binomial distribution and Poisson distribution. [TU, BE 2062 Jestha/ 2063 Ashadh/ 2064 Poush/BE 2068 Bhadra]
11. Discuss the importance of normal distribution and write condition for normal approximation to the binomial distribution. [TU, BE, 2062 Bhadra]
12. Write important properties of normal distribution : [TU, BE, 2063 Kartik/ 2064 poush/2068 Bhadra/ 2068 Magh/Pokhara uni. BE 2004]
13. Define the normal distribution and standard normal distribution. [TU, BE, 2063 Kartik]
14. Define normal distribution for a continuous random variable. Discuss the area property and importance of the normal curve. [TU, BE, 2065 Kartik/2065 chaitra/2066 Magh]
15. Define the normal distribution and standard normal distribution and its application in engineering field. [TU, BE, 2067 Mangsir/ Purbanchal. Uni, BE 2004]
16. Define the terms : normal distribution, standard normal distribution, gamma distribution and chi-square distribution. [Purbanchal Uni, BE, 2003]
17. Define the gamma distribution, its parameter and write its applications. [TU, BE, 2062 Baishakh/2068 Bhadra]



### Numerical Problems

1. Let the phase error in a tracking device have probability density

$$f(x) = \begin{cases} \cos x & \text{for } 0 < x < \pi/2 \\ 0 & \text{otherwise} \end{cases}$$

Find the probability that the phase error is

(a) between 0 and $\pi/4$; (b) greater than $\pi/3$. [Ans : (a) 0.707; (b) 0.1339]

2. If the probability density of a random variable is given by

$$f(x) = \begin{cases} kx^3 & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) the value of *k*; (b) $P\left(\frac{1}{4} \le X \le \frac{3}{4}\right)$; (c) $P\left(X > \frac{2}{3}\right)$ (d) Mean and variance

[Ans : (a) $k = 4$; (b) 0.3125; (c) 0.8025; (d) 0.8, $\sigma^2 = 0.0267$]

3. If the *pdf* of *X* is

$$f(x) = \begin{cases} Ax(x - 2) & \text{for } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Find *A* and mean. [Ans : $A = \frac{3}{4}$, $\mu = 1$] [TU, BE, 2067 Mangsir]

4. Show that the following function is the probability density function

$$f(x) = \begin{cases} 0 & \text{for } 0 < x < 2 \\ \frac{1}{22}(5 + 2x) & \text{for } 2 \le x \le 4 \\ 0 & \text{for } x > 4 \end{cases}$$

5. Show that the following is a distribution function

$$F(x) = \begin{cases} 0 \text{ for } x < -a \\ \frac{1}{2a}(x + 1) \text{ for } -a \le x \le a \\ 0 \text{ for } x > a \end{cases}$$

[Hint: (a) $F'(x) = f(x) = \begin{cases} 0 & \text{for } x < -a \\ \frac{1}{2a}(x + 1) & \text{for } -a \le x \le a \\ 0 & \text{for } x > a \end{cases}$ exists.

(b) (i) $f(x)$ lies between 0 and 1 in given range $[-a, a]$

(ii) $\int_{\infty}^{\infty} f(x)\, dx = 1$. so, $f(x)$ is *pdf*.

Since $F'(x) = f(x)$ exists and $f(x)$ is *pdf*, so $F(x)$ is *cdf*]

6. The length of time (in minutes) that a certain lady speaks on the telephone is found to be random phenomenon, with a probability function specified by the probability density function $f(x)$ as

$$f(x) = \begin{cases} Ae^{-x/5} & \text{for } x \ge 0 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find the value of A. (b) What is the probability that the number of minutes that she will take over the phone is

(i) more than 10 minutes; (ii) less than 5 minutes and (iii) between 5 and 10 minutes.

[Ans: (a) A = 1/5; (b) 0.1354; 0.6321; 0.2325]

7. If *pdf* of *X* is

$$f(x) = \begin{cases} ax & \text{for } 0 \le x < 1 \\ a & \text{for } 1 \le x < 2 \\ -ax + 3a & \text{for } 2 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$ find *a* and $P(X \le 1.5)$.

[Ans: $a = 0.5$, $P(X \le 1.5) = 0.5$]

8. If *pdf* of *X* is: $f(x) = e^{-x}$, $0 < x < \infty$; Find mean and variance.

[Hint: $\mu = \int_0^{\infty} x^{2-1} e^{-x}\, dx = \Gamma(2) = 1,\ \ \sigma^2 = \int_0^{\infty} x^{3-1} e^{-x}\, dx - \mu^2 = \Gamma(3) - 1 = 2$]

9. If the probability density of a random variable $X$ is given by

$$f(x) = \begin{cases} cx^2 & 0 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$ *[TU, BE, 2057 Bhadra/2067 Shrawan]*

Find constant $c$ and $P(1 < X < 2)$ [Ans: 1/9, 0.2593]

10. A college professor never finishes his lecture before the bell rings to end of the period, and always finishes his lecture within one minute after the bell rings. Let $X$ = the time that elapses between the bell and the end of the lecture, and suppose the probability density function of $X$ is:

$$f(x) = \begin{cases} kx^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of $k$;
(b) What is the probability that the lecture ends within 20 seconds of bell rings?
(c) What is the probability that the lecture continue beyond the bell for between 20 to 40 seconds? *[TU, BE, 2062 Jestha]*

[Hint: (a) $\int_0^1 kx^2 = 1 \Rightarrow k = 3$, (b) $\int_0^{1/3} 3x^2\, dx = 0.0370$; (c) $\int_{1/3}^{2/3} 3x^2\, dx = 0.2593$]

11. A continuous random variable $X$ has density function as follows:

$$f(x) = \begin{cases} cx^2 & \text{for } 1 \le x < 2 \\ cx & \text{for } 2 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$ *[TU, BE, 2063 Kartik]*

Find (*a*) the constant c; (*b*) $P(X>2)$; (*c*) $P\left(\frac{1}{2} < X < \frac{3}{2}\right)$.

[Hint: (*a*) $\frac{6}{29}$; (*b*) $\int_2^3 \frac{6}{29} x\, dx = 0.5172$; (*c*) $\int_{1/2}^{3/2} f(x)\, dx = \int_{1/2}^{1} 0\, dx + \int_1^{3/2} cx\, dx = 0.1293$]

12. The diameter, say $X$, of an electric cable, is assumed to be a continuous random variable with *pdf*: $f(x) = 6x(1 - x)$, $0 \le x \le 1$.
(a) Check that the above is a *pdf*. (b) obtain an expression for the *cdf* of $X$; (c) compute $P\left(X \le \frac{1}{2} \mid \frac{1}{3} \le X \le \frac{2}{3}\right)$, and (d) Determine the number $k$ such that $P(X < k) = P(X > k)$

[Hint: (*b*) $F(x) = \begin{cases} 0 & \text{for } x \le 0 \\ 3x^2 & \text{for } 0 < x \le 1 \\ 1 & \text{for } x > 1 \end{cases}$

(c) $P\left(X \le \frac{1}{2} \mid \frac{1}{3} \le X \le \frac{2}{3}\right) = \dfrac{P\left(X \le \frac{1}{2} \cap \frac{1}{3} \le X \le \frac{2}{3}\right)}{P\left(\frac{1}{3} \le X \le \frac{2}{3}\right)} = \dfrac{P\left(\frac{1}{3} \le X \le \frac{1}{2}\right)}{P\left(\frac{1}{3} \le X \le \frac{2}{3}\right)}$

$= \dfrac{\int_{1/3}^{1/2} 6x(1-x)\, dx}{\int_{1/3}^{2/3} 6x(1-x)\, dx} = \frac{11}{26} = 0.4231$

(d) $P(X < k) = P(X > k) \Rightarrow \int_0^k 6x(1-x)\, dx = \int_k^1 6x(1-x)\, dx$

$\Rightarrow 4k^3 - 6k^2 + 1 = 0 \Rightarrow k = \frac{1}{2}, \frac{1 \pm \sqrt{3}}{2}$]

13. A petrol pump is supplied with petrol once a day. If its daily volume of sales ($X$) in thousand of liters is distributed by:
$f(x) = 5(1 - x)^4$, $0 \le x \le 1$; what must be the capacity of its tank in order that the probability that its supply will be exhausted in a given day shall be 0.01?





[Hint: Let capacity of tank (in '000 of liters) be $V$ such that $P(X \ge V) = 0.01$
$\Rightarrow \int_V^1 5(1 - x)^4 = 0.01 \Rightarrow V = 0.6019$. So, capacity = 0.6019 × 1000 = 601.9 liters]

14. Show that the following function is a probability density function

$$f(x) = \begin{cases} 0 & \text{for } x < 2 \\ \frac{1}{22}(5 + 2x) & \text{for } 2 \le x \le 4 \\ 0 & \text{for } x > 4 \end{cases}$$ *[TU, BE, 2065 Chaitra]*

15. If $X$ is uniformly distributed over the interval [0, 10], compute the probability that (a) $2 < X < 9$; (b) $1 < X < 4$; (c) $X < 5$; (d) $X > 6$
[Ans: (a) 7/10; (b) 3/10; (c) 5/10; (d) 4/10]

16. Show that following distribution is rectangular distribution

$$f(x) = \frac{1}{2\alpha},\ -\alpha < x < \alpha$$

17. A rv $X$ has uniform distribution over (–3, 3). Compute (a) $P(X = 2)$; (b) $P(|X| \ge 2)$; (c) $P(|X - 2| < 2)$. [Ans: (a) 1/6; (b) 1/3; (c) 1/2]

18. If $X$ is uniformly distribution over ($a$, $b$) with mean 1 and variance $\frac{4}{3}$, find (a) $a$ and $b$; (b) $f(x)$; (c) $P(X < 0)$
[Ans : (a) $a = -1$ $b = 3$; (b) $f(x) = 1/4$; $-1 < x < 3$; (c) $P(X < 0) = 1/4$]

## Normal Distribution and Standard Normal Distribution

19. If a random var iable has the standard normal distribution, find the probability that it will take on a value (a) less than 1.50; (b) less than –1.20; (c) greater than 2.16; (d) greater than –1.75; (e) between –1.35 and –0.35.
[Ans : (a) 0.9332; (b) 0.1151; (c) 0.0514; (d) 0.9599; (e) 0.2747]

20. Find $Z$ if the probability that a random variable having the standard normal distribution will take on a value(a) less than $Z$ is 0.9911; (b) greater than $Z$ is 0.1093; (c) greater than $Z$ is 0.6443; (d) less than $Z$ is 0.0217; (e) between $-Z$ and $Z$ is 0.9298.
[Ans : (a) 2.73; (b) 1.23; (c) –0.37; (d) –2.02; (e) 1.81]

21. A corporation installs 10,000 electric lamps in the street of a city. If these lamps have an average life of 1850 hours with a standard deviation of 200 hours, what number of lamps may be expected to burn for (a) more than 2000 hours; (b) less than 1600 hours; (c) between 1540 and 1300 hours?
[Ans : (a) 2266; (b) 1056; (c) 3407] *[TU, BE 2065 Chaitra]*

22. The scores of candidates in certain test are normally distributed, with mean 500 and variance 10,000. Find the probability that the candidates to receive the score (a) between 350 and 550; (b) above 400; (c) below 500.
[Ans : (a) 0.6247; (b) 0.8413; (c) 0.5000] *[TU, BE 2065 Chaitra]*

23. In a certain examination 2000 students appeared in statistics. The average marks obtained were 50% and the s.d was 5%. How many students do you expect to obtain more than 60% marks? Assume that the marks are normally distributed.
[Ans: 46] *[TU, BE 2065 Chaitra (Ret)]*

24. A banker claims that the average life of a regular saving account opened in his bank is 18 months with a variance of 41.6025. What is the probability that (a) there will still be money in a saving account between 20 and 22 months by a depositor ; (b) The bank account will be closed (no money in the deposit) after two years.
[Ans: (a) 0.1107; (b) $P(X > 24) = 0.1762$] *[TU, 2066]*

25. The distribution of monthly incomes of 5,000 employees of a certain industrial unit was found to be normally distributed with mean of Rs. 2000 and a standard deviation of Rs. 200. Estimate (a) the range of incomes of the middle 60% employees; (b) the lowest income of richest 10% employees; (c) the biggest income of poorest 10% employees.
[Ans: (a) Rs. [illegible]; (b) Rs. 2256; (c) Rs. 1744] *[TU, [illegible] 2066]*

26. The mean and variance of height of 500 students were found to be 165 cm and 25cm² respectively. Find the range of the middle 80% of the students.
[Ans: Range = 12.8 cm] *[TU, [illegible] 2067]*

**27.** In a certain examination test 3000 student appeared in statistics. The average mark obtained was 65% and s.d. was 5%. How many students do you expect to obtain more than 70% mark. What is the minimum marks of 200 students. Assuming that the marks are normally distributed. [Ans: 0.1587] [*TU BE 2060*]

**28.** The mean weight of products is 68.22 grams with variance of 10.8 grams. How many products in a batch of 1000 would you expect (a) to be over 72 grams; (b) between 70 and 72 grams. [Ans: (a) 0.3749; (b) 0.2054] [*TU 2051 MBA*]

**29.** Assume that the marks in MBA examination are normally distributed with mean $\mu$ = 400 and $\sigma$ = 100. Of 600 students taking examination, it is desired to pass 500 of them, What should be the lowest marks permitted for passing? [Ans : 303] [*TU 2052 MBA*]

**30.** The mean I.Q. intelligence quotient of a large number of children of age 14 was 100 and *s.d.* 16. Assuming that the distribution was normal, find between what limits the IQ's of the middle 40% of the children lie? [Ans : 91.52 and 108.48] [*TU 2055 MBA*]

**31.** In an examination, average marks secured by 400 students is 45 with s.d. of 10. Assuming the distribution to be normal, find (i) the number of students securing marks between 50 and 60; (b) the range of marks within while middle 50% of students would lie ? [Ans : (i) 97; (b) 38.3; 51.7; Range = 13.4] [*TU 2056 MBA*]

**32.** 1000 bulbs with a mean life 120 days are installed in a new factory; their length of life is normally distributed with *s.d.* 20 days; (a) How many bulbs will expire in less than 90 days ? (c) If it is decided to replace all the bulbs together, what interval should be allowed between replacements if not more than 10% should expire before replacement? [**Hint:** $NP(X < 9) = 1000 \times 0.0668 = 66.88 \approx 67.$
Again $P(X < x_1) = 0.10 \Rightarrow z_1 = 1.28$. So, $x_1 = 120 - 20\,z_1 = 94.4 \approx 94$ days]

**33.** The breakdown voltage $X$ of randomly chosen diode of a particular type is known to be normally distributed with mean 40 and s.d. 1.5 volts. What is the probability that the breakdown voltage will be (a) between 39 and 42 volts; (b) at most 43 volts; (c) at least 39 volts. [Ans : (a) 0.1596; (b) 0.9772; (c) 0.2515] [*Purbanchal Uni. BE, 2006*]

**34.** The mean and standard deviation of the strengths of 40 steel bars from a particular samples are given as 60.14 and 5.02 KN. What is the probability that
(a) test steel bars will fail when subjected to a strength of 45 KN or less ?
(b) strength of bar is between 55 KN to 65 KN ?
[Ans: (a) 0.0013; (b) 0.6801] [*TU BE 2067 Mangsir*]

**35.** In an examination, 10% students got less than 20 marks and 95% got less than 75 marks. Assuming the distribution to be normal, find the mean and standard deviation. [Ans: 18.836; 44.11] [*TU 2063*]

**36.** The marks obtained by 1000 students in an examination are known to be normally distributed. If 15% of the students got less than 30 and 10% of the students got over 90, find the mean and variance of the distribution.[Ans: 56.896; 668.847] [*TU 2060*]

**37.** If 15% of the candidate got first class (60 marks or above) while 40% failed (securing below 40). Assuming the marks to normally distributed, estimate the mean and standard deviation. [Ans: 43.875; 15.5] [*TU 2042 MBA*]

**38.** The mean of the normal distribution is 60 and 6% of the values are greater than 80. Find the *s.d.* of the distribution. [Ans: 12.82] [*TU. 2045*]

**39.** In an examination, 10% of the students got less than 20 marks and 5% of the students got over 75 marks. Assuming the distribution to be normal, find the mean and *s.d.* of the distribution. [Ans: 44.068; 18.803] [*TU MBS 2063*]

**40.** In a normal distribution. 31% of the items are under 45 and 8% are over 64. Find the mean and *s.d.* of the distribution. [Ans: 1.4; 0.5] [*TU 2045 MBA/Pokhara Uni. BE 2002*]

**41.** The time for a super glue can be treated as a random variable having a normal distribution with mean 30 seconds. Find its standard deviation if the probability is 0.20 that it will take on a value greater than 39.2 seconds. [Ans: 10.93]

**42.** If a random variable has the binomial distribution with $n$ = 40 and $p$ = 0.60, use the normal approximation to determine the probabilities that it will take on (a) the value 14; (b) a value less than 12. [Ans: (a) 0.049; (b) 0.008]

**43.** The probability that an electronic component will fail in less than 1,000 hours of continuous use is 0.25. Use the normal approximation to find the probability that among 200 such components fewer than 45 will fail in less than 1,000 hours of continuous use. [Ans: 0.1841]





**44.** If 62% of all clouds seeded with silver iodide show spectacular growth, what is the probability that among 40 clouds seeded with silver iodide at most 20 will show spectaculated growth? [Ans: 0.0[illegible]]

**45.** On an average there are 64 robberies in Kathmandu city per month. Find the probability that there are: (a) exactly 80 robberies; (b) more than 80 robberies; (c) less than 80 robberies; by using normal approximation to the Poisson distribution. [Ans: (a) 0.0065; (b) 0.0197 (c) 0.9738]

**46.** The number of accidents occurring on each year in a busy road is 30. They occur completely at random. Use normal approximation to the Poisson distribution to find the probability that in 2000 there will be more than 45 accidents. [Ans: 0.0032]

## Additional Questions:

1. A random variable $X$ is distributed at random between the values 0 and 1 so that its probability density function is : $f(x) = kx^2\,(1 - x^3)$, where $k$ is constant. Find the value of $k$. Using this value of $k$. find its mean and variance. [Ans : $k = 6$, $\mu = 9/14$, $\sigma^2 = 9/245$]

2. A variable $X$ is distributed at random between the values 0 and 4 and its probability density function is given by : $f(x) = kx^3\,(4 - x)^2$. Find (a) $k$; (b) $\mu$ and $\sigma$.
[Ans: (a) $k = 15/1024$ (b) $\mu = 16/7$, $\sigma = 2\sqrt{2/7}$]

3. A normal population of 700 wages earners has a mean income of Rs. 220 per month and variance is 332. Find the no. of persons who earn (a) between Rs. 150 and Rs 200 per month (b) more than Rs. 300 per month. (c) less than Rs. 100 per month.
[Ans: (a) 0.1356] [*Purbanchal Uni. BE, 2005*]

4. Income of a group of 10,000 persons were found to be normally distributed with mean Rs. 520 and *s.d.* Rs. 60. Find
(a) the number of persons having income between Rs 400 and 550.
(b) the number of persons having income greater than Rs. 500.
[Ans: (a) 6687 (b) 6293] [*Purbanchal Uni. BE, 2004*]

5. The average daily sales of 500 branch offices were Rs. 150 thousands and the s.d. Rs. 15 thousands. Assuming distribution to be normal indicate how many branches have sales between [*Purbanchal Uni. BE, 2003*]
(a) Rs. 120 thousand and Rs. 145 thousand, and
(b) Rs 140 thousand and Rs. 165 thousand ? [Ans: (a) 174; (b) 295]

6. The mean yield for one-acre plot is 662 kilos with *s.d.* 32 kilos. Assuming normal distribution, how many one-acre plots in a batch of 1000 plots would you expect to have yield (a) over 700 kilos. (b) below 650 kilos ; (c) what is the lowest yield of the best 100 plots ? [*Pokhara Uni. BE, 2009 Fall*]
[**Hint** : (a) $P(X > 700) = P(Z > 1.19) = 0.117$. (b) $P(X < 650) = P(Z < -0.38) = 0.352$
(c) proportion of best 100 plots $= \frac{100}{1000} \times 100\% = 10\% = 0.10$
Let $x_1$ be the minimum production of the top 100 plots.
Then $P(X > x_1) = 0.10$. $z = \frac{x_1 - 662}{32} = z_1$ (say) or, $P(Z > z_1) = 0.10$
or, $1 - P(Z \le z_1) = 0.10 \Rightarrow P(Z \le z_1) = 0.90 \Rightarrow F(z_1) = 0.90$
From the normal table the value of the probability closest to 0.90 is 0.8997 at
$Z = 1.28$. At $Z = 1.28$; $\frac{x_1 - 662}{32} = z_1 \Rightarrow x_1 = 662 + 32 \times 1.28 = 702.96$]

7. The number of days sick leave requested annually by employees of a large insurance company is normally distributed with a mean of 9 days and standard deviation of 2.5 days.
(a) What proportion of the employees requested at least 10 days sick leave?
(b) If the company employs 600 people, how many use fewer than 2 days sick leave annually? [*TU, BE 2062 Bhadra*]
[**Hint:** $X$ = number of days sick leave requested.
(a) $P(X \ge 10) = P(Z \ge 0.4) = 1 - P(z < 0.4) = 1 - F(0.4) = 1 - 0.6554 = 0.3446$
(b) $N\,P(X < 2) = 600\,P(Z < -2.8) = 600 \times F(-2.8) = 600 \times 0.0026 = 1.56 \approx 2$ ].

8. The time requirement to assemble a piece of machinery is a random variable having approximately a normal distribution with $\mu = 12.9$ minutes and $\sigma = 2.0$ minutes. What are the probabilities that the assembly of piece of machinery of this kind will take
(a) at least 11.5 minutes;
(b) anywhere from 11 to 14.8 minutes. *[TU, BE 2064 shrawan/2068Magh]*

[Hint: $X$ = the time requirement. $Z = \frac{X - 12.9}{2}$

(a) $P(X \geq 11.5) = P\left(\frac{X - 12.9}{2} \geq \frac{11.5 - 12.9}{2}\right) = P(Z \geq -0.7)$
$= 1 - P(Z < -0.7) = 1 - 0.2420 = 0.7580$

(b) $P(11 \leq X \leq 14.8) = P\left(\frac{11 - 12.9}{2} \leq \frac{X - 12.9}{2} \leq \frac{14.8 - 12.9}{2}\right)$
$= P(-0.95 \leq Z \leq 0.95) = F(0.95) - F(-0.95) = 0.8289 - 0.1711 = 0.6578$]

9. The burning time for an experimental rocket is a random variable having the normal distribution with $\mu = 4.76$ seconds and $\sigma = 0.04$ seconds. What is the probability that this kind of rocket will burn *[TU, BE 2068 Bhadra]*
(a) less than 4.66 seconds ;(b) more than 4.80 seconds ;
(c) anywhere from 4.70 to 4.82 seconds. [Ans: (a) 0.0062; (b) 0.1587; (c) 0.8664]

10. If 70% of all consultations handled by students consultants at a computer centre involve programs with syntax errors, and let $X$ denote the number with such errors in a random sample of 100 consultations what is the approximate probability that $X$ is : (a) between 60 and 80 inclusive ; (b) at most 75 ? (c) less than 75 ?
[Ans: (a) 0.9708; (b) 0.862; (c) 0.8078]