A r.v. X is said to have a hypergeometric distribution if its p.m.f. is given by (5.4.6)

We have $\sum_{k=0} P(X=k) = 1$ by Theorem 2.3.3.

Example 5.4.9.

A lot of 20 radio tubes, 8 of which are defective. A sample of 5 items is selected from it. Find the probability that the sample will contain.

i. one defective radio tube,
ii. at least 4 defective radio tubes.

Solution

i. $$P(X = 1) = \frac{C(8,1).C(12,4)}{C(20,5)} = \frac{8(595)}{15504}$$

$$= 0{\cdot}2554 .$$

ii. $$P(X \geq 4) = \frac{C(8,4)C(12,1)}{C(20,5)} + \frac{C(8,5)C(12,0)}{C(20,5)}$$

$$= \frac{70(12) + 56(1)}{15504} = 0{\cdot}0577 .$$

5.4.6 Negative Binomial Distribution

In the geometric disribution we are interested in the first success. If we are interested in the k-th success instead of the first, in this case we apply the negative binomial distribution which can be explained as follows:

Suppose we have n independent Bernoulli trials, we can say that the probability of getting (k-1) success in (n-1) trials is

$$C(n-1, k-1)p^{k-1}q^{n-k}$$

and since the probability of getting another success on the n-th trial is p, it follows that the probability of getting the k-th success on the n-th trial is

$$P(X = k) = C(n-1, k-1)\, p^{k-1} q^{n-k} . p$$
$$= C(n-1, k-1)\, p^{k} q^{n-k}, \; n = k, k+1, \ldots \quad (5{\cdot}4{\cdot}7)$$

A r.v. X is said to have a negative binomial distribution with parameters p and k if its p.m.f. is given by (5.4.7).

Example 5.4.10.

Suppose that the probability for a child exposed to a certain disease to catch it is 0.2, what is the probability that the twelfth child exposed to the disease will be the third to catch it?

Solution

Substituting n=12, k=3, and p=0.2 in (5.4.7), we get

$$P(X=3) = C(11,2)\,(0{\cdot}2)^3\,(0{\cdot}8)^9$$
$$= 0.059$$

5.5 Continuous Random Variables

In Sections 5.3 and 5.4 we limited our discussion to discrete random variables. In Section 5.2 we defined the distribution function by defining its values F(b) as the probabilities that a r.v. X takes on a value less than or equal to b, for $-\infty < b < \infty$. At that time we restricted our discussion to discrete random variables, so that the graph of their distributions was step function as in Figures 5.1 and 5.2.

Here we consdier the case where the range of a r.v. X consists of the set of all real numbers R or any subset of R and the graph of F(x) is a continuous curve. as in Example 5.5.1 below. There are other types of distribution which are mixtures of discrete and continuous types. We shall study only the distribution functions which are continuous and whose derivative

$$F'(x) = \frac{dF(x)}{dx} = f(x) \qquad \dots (5 \cdot 5 \cdot 1)$$

exists for all but a finite set of values of x. The function f is called the *probability density function* (p.d.f.) of the r.v. X. The p.d.f. must satisfy

1. $f(x) \geq 0$ for all x

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

All probability statements about X can be answered in term of f. For example

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx = F(b) - F(a) \qquad \dots (5 \cdot 5 \cdot 2)$$

$$\text{Where } F(y) = \int_{-\infty}^{y} f(x)\,dx. \qquad \dots (5 \cdot 5 \cdot 3)$$

Also

$$P(X = a) = \lim_{h \to 0} P(a \leq X \leq a + h)$$

$$= \lim_{h \to 0} \int_a^{a+h} f(x)\,dx = 0 \qquad \dots (5 \cdot 5 \cdot 4)$$

That is, the probability that a continuous r.v. X will take on any specific value a is always zero.

From Equations (5.5.3) and (5.5.4), we get

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

Example 5.5.1.

If the random variable X has the p.d.f.

$$f(x) = \begin{cases} k(x+1) & , -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

1. Find the constant k.
2. Find F(x) and draw its graph.

Solution

1. Since f is a p.d.f., then we must have

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

$$k \int_{-1}^{1} (x+1)\, dx = 1$$

$$k \left[\frac{x^2}{2} + x \right]_{-1}^{1} = 1$$

$$\therefore k = \frac{1}{2}$$

2. $F(x) = \int_{-x}^{x} f(u)\, du \quad , \quad -1 < x < 1$

$$= \frac{1}{2} \int_{-1}^{x} (u + 1)\, du = \frac{1}{2} \left[\frac{x^2}{2} + x + \frac{1}{2} \right]$$

Therefore

$$F(x) = \begin{cases} 0 & x < -1 \\ \dfrac{x^2}{4} + \dfrac{x}{2} + \dfrac{1}{4} & -1 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

The graph of F(x) is given below

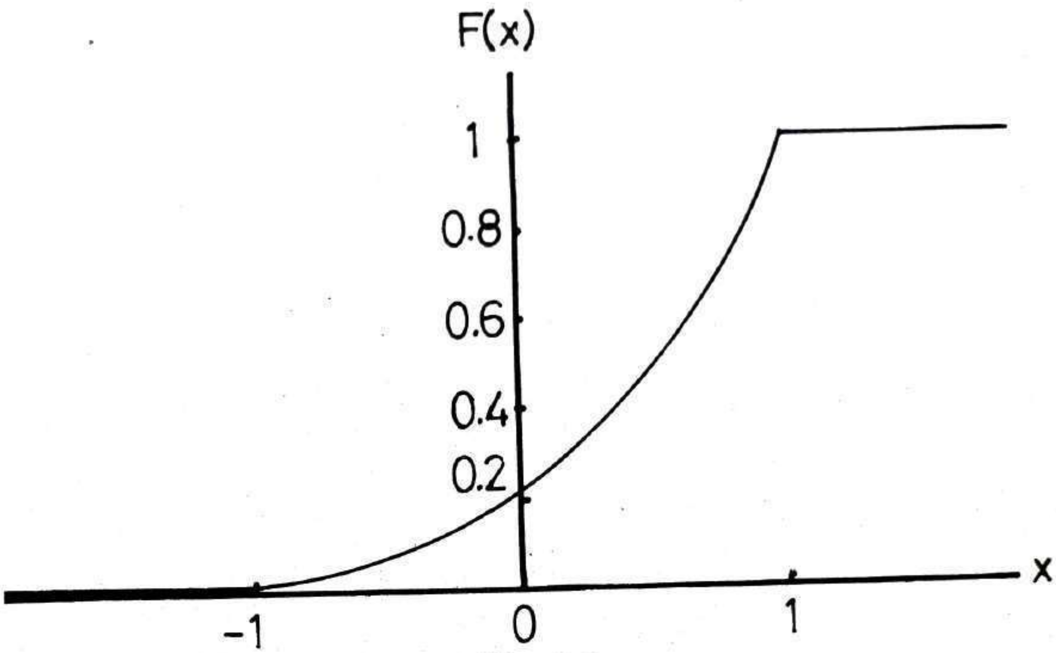

Fig. 5.5

Example 5.5.2.

Let X be a r.v. with distribution function F given by

$$F(x) = \begin{cases} 0 & x < 0 \\ x^3 & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

i. Find the p.d.f f(x).
ii. Find $P(x \le 0.7)$, $P(X > 0.5)$, and $P(X=1)$.

Solution :-

i. By (5.5.1), we have

$$f(x) = \frac{dF(x)}{dx} = \frac{d(x^3)}{dx}, 0 \le x < 1$$

Then

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

ii. $P(X \le 0.7) = F(0.7) = (0.7)^3 = 0.343,$
$P(X > 0.5) = 1 - P(X \le 0.5)$
$= 1 - F(0.5) = 1-(0.5)^3$
$= 0.875,$
$P(X=1) = 0$, since X is continuous r.v.

5.6. Special Univariate Continuous Distribution Functions

5.6.1 Uniform Distribution

A very simple distribution for a continuous r.v. is the uniform distribution (or rectangular distribution).

A r.v. X is said to have uniform distribution in the interval [a,b] if its p.d.f. is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases} \quad \text{.......... (5.6.1)}$$

where a and b are real constants with $a < b$. The graph of f(x) is shown in Fig. 5.6.

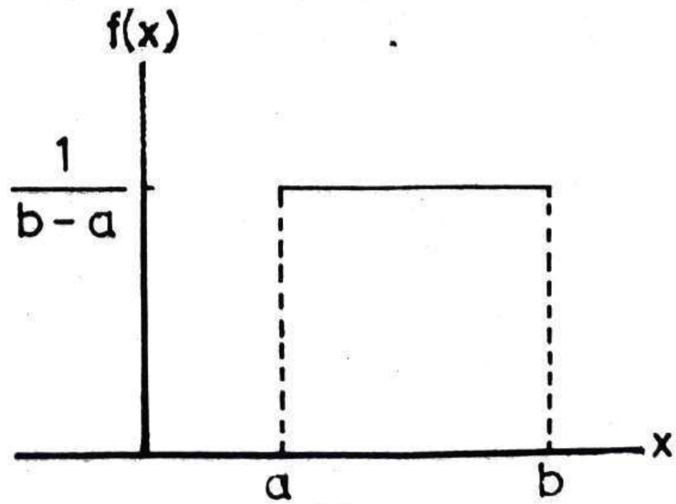

Fig. 5.6

The distribution function F is given by

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

The following p.d.f.s are some examples of uniform distribution.

1. $f(x) = 1 \quad 0 < x < 1$
   $\quad\quad\; 0 \quad$ otherwise

2. $f(x) = \dfrac{1}{5} \quad 2 < x < 7$
   $\quad\quad\; 0 \quad$ otherwise

3. $f(x) = \dfrac{1}{4} \quad -1 < x < 3$
   $\quad\quad\; 0 \quad$ otherwise

Example 5.6.1.

Buses arrive at a bus stop at 10-minute intervals starting at 6 A.M. That is they arrive at 6, 6:10, 6:20, 6:30, and so on. If a passenger arrives at the bus stop at a time that is uniformly distributed between 6 and 6:20, find the probability that he waits for

1. less than 7 minutes for a bus?
1. more than 4 minutes for a bus?

Solution :

Let X denote the number of minutes past 6 that the passanger arrives at the bus stop, then

$f(x) = \dfrac{1}{20} \quad 0 < x < 20$
$\quad\quad\; 0 \quad$ otherwise

1. The passenger will wait for less than 7 minutes if he arrives between 6:3 and 6:10 or between 6:13 and 6:20. Hence the required probability is

$$P(3 < X < 10) + P(13 < X < 20) = \int_3^{10} \frac{1}{20}\,dx + \int_{13}^{20} \frac{1}{20}\,dx$$

$$= \frac{7}{10}$$

2. The passenger will wait for more than 4 minutes if he arrives between 6. and 6:4 or between 6:10 and 6:14, so the required probability is

$$P(0 < X < 4) + P(10 < X < 14) = \int_0^{4} \frac{1}{20}\,dx + \int_{10}^{14} \frac{1}{20}\,dx$$

$$= \frac{2}{5}$$

5.6.2. Exponential Distribution

A r.v. X is said to have an exponential distribution if its p.d.f. is given by

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{.............} (5.6.2)$$

where $\theta > 0$.

We have $$\int_0^{\infty} \frac{1}{\theta} e^{-x/\theta}\,dx = -e^{-x/\theta}\Big]_0^{\infty} = 0 - (-1) = 1.$$

The distribution function F is given by

$$F(x) = \int_0^x f(u)\,du = \int_0^x \frac{1}{\theta} e^{-u/\theta}\,du = 1 - e^{-x/\theta}, x > 0,$$

while for $x \leq 0$, $F(x) = 0$.

Example 5.6 2.

The mileage (in thousand of miles) which car owners get with a certain kind of tyre is a r.v. having an exponential with $\theta = 20$, i.e.

$$f(x) = \begin{cases} \frac{1}{20} e^{-\frac{x}{20}} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the probability that one of these tyres will last

1. at most 12000 miles
2. anywhere from 18000 to 24000 miles.

Solution :

Let X denote the mileage of this tyre (in 1000 miles), then

$$P(X < 12) = \int_0^{12} \frac{1}{20} e^{-\frac{x}{20}} dx = -e^{-\frac{x}{20}}\Big]_0^{12}$$

$$= 1 - e^{-\frac{12}{20}} = 1 - 0{\cdot}548 = 0{\cdot}452.$$

$$P(18 < X < 24) = \int_{18}^{24} \frac{1}{20} e^{-\frac{x}{20}} dx$$

$$= e^{-\frac{18}{20}} - e^{-\frac{24}{20}} = 0{\cdot}406 - 0{\cdot}301 = 0{\cdot}105$$

5.6.3. Gamma Distribution

The gamma distribution is an important generalization of the exponential distribution

A r.v. X is said to have a gamma distribution with parameters α and β, if its p.d.f. is given by

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} e^{-\frac{x}{\beta}} x^{\alpha-1} \quad x > 0 \qquad \text{..........} (5.6.3)$$

Where α, β are positive constants and the gamma function $\Gamma(\alpha)$ is defined as

$$\Gamma(\alpha) = \int_{0}^{\infty} x^{\alpha-1} e^{-x}\, dx \text{ for } \alpha > 0 \qquad \text{............} (5.6.4)$$

If we integrate (5.6.4) by parts, we can show that the gamma function satisfies the recursive formula

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) \qquad \text{.............} (5.6.5)$$

If α is a positive integer, then

$\Gamma(\alpha) = (\alpha - 1)!$

Equation (5.6.3) is reduced to (5.6.2) when $\alpha = 1, \beta = \frac{1}{\theta}$

By using double integration one can show that

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

(The proof is out of the scope of this book.)

5.7. Normal Distribution*

The r.v. X is said to have a normal distribution with parameters μ and σ^2 if its p.d.f. is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad -\infty < x < \infty, \qquad \text{..........} (5.7.1)$$

---

* The normal distribution is sometimes referred to as the Gaussian distribution in honor of Karl Gauss (1777-1855)

where μ and σ are real constants and $\sigma > 0$. The values μ and σ^2 represent, the average value (the mean) and the possible variation of X (variance), respectively.

The graph of f(x) is a bell-shape curve, that is symmetric about μ (see Fig. 5.7)

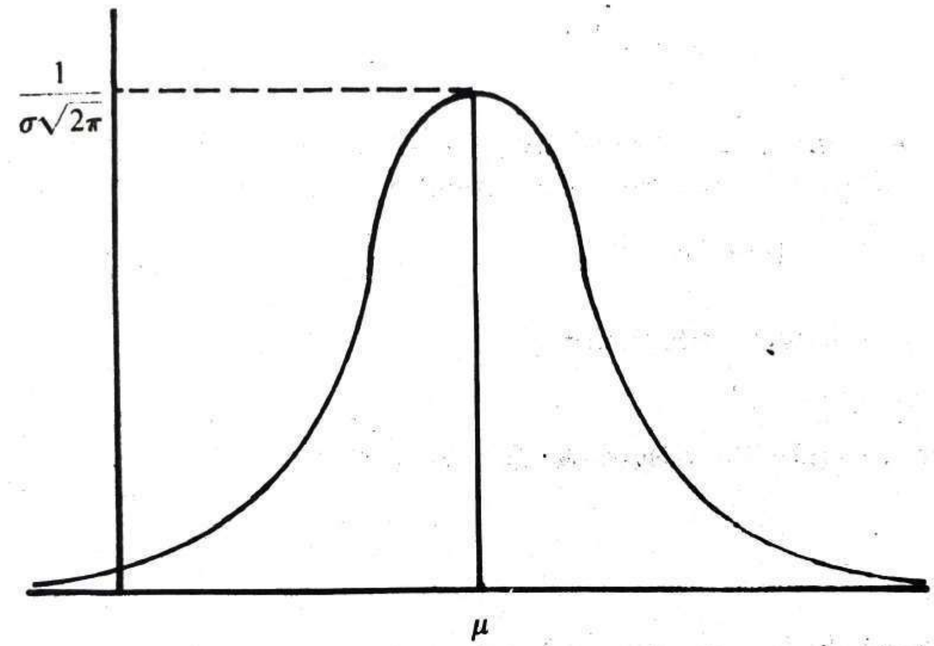

Fig. 5.7 Graph of Normal Distribution.

A r.v. X having a normal distribution with parameters μ and σ^2 is expressed by $X \sim N(\mu, \sigma^2)$.

To verify that $\int_{-\infty}^{\infty} f(x)\, dx = 1$, we make the substitution

$$y = \frac{x - \mu}{\sigma} \text{ and get}$$

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

$$= \frac{2}{\sqrt{2\pi}} \int_0^\infty e^{-y^2/2} \, dy \quad \text{(since } e^{-y^2/2} \text{ is an even function)},$$

Now let $z = \frac{y^2}{2}$, then

$$I = \frac{2}{\sqrt{2\pi}} \int_0^\infty e^{-z} . \frac{dz}{\sqrt{2z}} = \frac{1}{\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} e^{-z} dz$$

$$= \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{\sqrt{\pi}} = 1.$$

Notes :

1. If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X - \mu}{\sigma}$ ia a standard normal variate with mean O and Variance 1.
2. The p.d.f. of the standard normal variate Z is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty \qquad \text{......... (5.7.2)}$$

and the corresponding distribution function

$\phi(z) = P\{Z \le z\}$ is given by

$$\phi(z) = \int_{-\infty}^{z} f(u)\, du = \int_{-\infty}^{z} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du \qquad \text{... (5·7·3)}$$

The values of ϕ (z) for $z \geq 0$ are given in Table (2) in the Appendix. In Fig. 5.8, the shaded area gives the probability that the standard normal variate Z lies between Z=a and Z=b , i.e.

$$P(a < Z < b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

$\phi(b) - \phi(a)$ for $a < b$

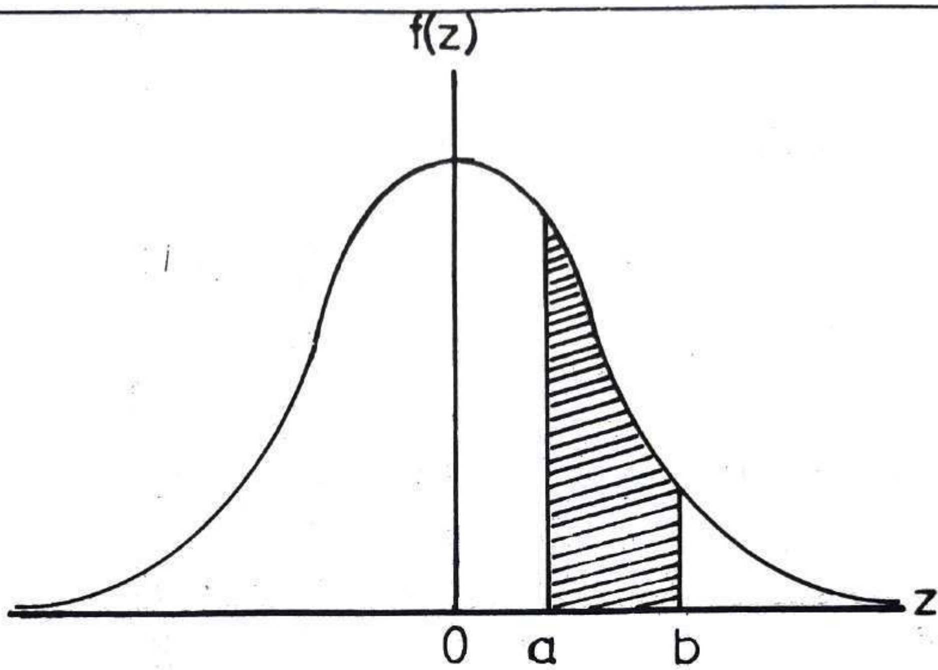

Fig 5·8

The values of ϕ (z) for negative values can be obtained from the relation

$$P(Z < -a) = P(Z > a)$$
$$= 1 - P(Z \leq a)$$
$$= 1 - \phi(a)$$
$$\therefore \phi(-a) = 1 - \phi(a).$$

For example,

1. $P(2 < Z < 3.1) = \phi(3.1) - \phi(2)$
$= 0.999 - 0.9772 = 0.0218$

2. $P(Z \leq -0.56) = 1 - \phi(0.56)$
$= 1 - 0.7123 = 0.2877$

If $X \sim N(\mu, \sigma^2)$, the distribution function of X can be expressed as

$$F(a) = P(X \leq a)$$