4

Probability Distribution

4.1. RANDOM VARIABLE

If the numerical values assumed by a variable are the result of some chance factors, so that a particular value cannot be exactly predicted in advance, the variable is then called a *random variable*. A random variable is also called '*chance variable*' or '*stochastic variable*'.

Random variables are denoted by capital letters, usually, from the last part of the alphabet, for instance, X, Y, Z etc.

Continuous and Discrete Random Variables

A *continuous random variable* is one which can assume any value within an interval, *i.e.*, all values of a continuous scale. For example (*i*) the weights (in kg) of a group of individuals, (*ii*) the heights of a group of individuals.

A *discrete random variable* is one which can assume only isolated values. For example,

(*i*) the number of heads in 4 tosses of a coin is a discrete random variable as it cannot assume values other than 0, 1, 2, 3, 4.

(*ii*) the number of aces in a draw of 2 cards from a well shuffled deck is a random variable as it can take the values 0, 1, 2 only.

4.2. DISCRETE PROBABILITY DISTRIBUTION

Let a random variable X assume values $x_1, x_2, x_3, \ldots\ldots, x_n$ with probabilities $p_1, p_2, p_3, \ldots\ldots, p_n$ respectively, where $P(X = x_i) = p_i \geq 0$ for each x_i and $p_1 + p_2 + p_3 + \ldots\ldots + p_n = \sum_{i=1}^{n} p_i = 1$.

Then

$$\begin{array}{lllll} X & : & x_1, & x_2, & x_3, \ldots\ldots, x_n \\ P(X) & : & p_1, & p_2, & p_3, \ldots\ldots, p_n \end{array}$$

is called the discrete probability distribution for X and it spells out how a total probability of 1 is distributed over several values of the random variable.

4.3. MEAN AND VARIANCE OF RANDOM VARIABLES

Let

$$\begin{array}{lll} X : x_1, & x_2, & x_3, \ldots\ldots, x_n \\ P(X) : p_1, & p_2, & p_3, \ldots\ldots, p_n \end{array}$$

be a discrete probability distribution.

We denote the *mean* by μ and define $\mu = \dfrac{\Sigma p_i x_i}{\Sigma p_i} = \Sigma p_i x_i \qquad (\because \; \Sigma p_i = 1)$

Other names for the mean are *average or expected value E(X).*

We denote the *variance* by σ^2 and define $\sigma^2 = \Sigma p_i(x_i - \mu)^2$

If μ is not a whole number, then $\sigma^2 = \Sigma p_i x_i^2 - \mu^2$

Standard deviation $\sigma = +\sqrt{\text{Variance}}$.

ILLUSTRATIVE EXAMPLES

Example 1. *Five defective bulbs are accidentally mixed with twenty good ones. It is not possible to just look at a bulb and tell whether or not it is defective. Find the probability distribution of the number of defective bulbs, if four bulbs are drawn at random from this lot.*

Sol. Let X denote the number of defective bulbs in 4. Clearly X can take the values 0, 1, 2, 3 or 4.

Number of defective bulbs = 5

Number of good bulbs = 20

Total number of bulbs = 25

$$P(X = 0) = P \text{ (no defective)} = P \text{ (all 4 good ones)}$$

$$= \frac{^{20}C_4}{^{25}C_4} = \frac{20 \times 19 \times 18 \times 17}{25 \times 24 \times 23 \times 22} = \frac{969}{2530}$$

$$P(X = 1) = P\text{(one defective and 3 good ones)} = \frac{^5C_1 \times {^{20}C_3}}{^{25}C_4} = \frac{1140}{2530}$$

$$P(X = 2) = P \text{ (2 defectives and 2 good ones)} = \frac{^5C_2 \times {^{20}C_2}}{^{25}C_4} = \frac{380}{2530}$$

$$P(X = 3) = P\text{(3 defectives and 1 good one)} = \frac{^5C_3 \times {^{20}C_1}}{^{25}C_4} = \frac{40}{2530}$$

$$P(X = 4) = P \text{ (all 4 defectives)} = \frac{^5C_4}{^{25}C_4} = \frac{1}{2530}$$

$\therefore$ The probability distribution of the random variable X is

X	:	0	1	2	3	4
P(X)	:	$\frac{969}{2530}$	$\frac{1140}{2530}$	$\frac{380}{2530}$	$\frac{40}{2530}$	$\frac{1}{2530}$

Example 2. *A die is tossed thrice. A success is 'getting 1 or 6' on a toss. Find the mean and the variance of the number of successes.*

Sol. Let X denote the number of success. Clearly X can take the values 0, 1, 2 or 3.

Probability of success $= \frac{2}{6} = \frac{1}{3}$; Probability of failure $= 1 - \frac{1}{3} = \frac{2}{3}$

$$P(X = 0) = P \text{ (no success)} = P \text{ (all 3 failures)} = \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{8}{27}$$

$$P(X = 1) = P \text{ (one success and 2 failures)} = {^3C_1} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{12}{27}$$

$$P(X = 2) = P \text{ (two successes and one failure)} = {^3C_2} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{6}{27}$$

$$P(X = 3) = P \text{ (all 3 successes)} = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{27}$$

$\therefore$ The probability distribution of the random variable X is

X	:	0	1	2	3
P(X)	:	$\frac{8}{27}$	$\frac{12}{27}$	$\frac{6}{27}$	$\frac{1}{27}$

To find the mean and variance

x_i	p_i	$p_i x_i$	$p_i x_i^2$
0	$\frac{8}{27}$	0	0
1	$\frac{12}{27}$	$\frac{12}{27}$	$\frac{12}{27}$
2	$\frac{6}{27}$	$\frac{12}{27}$	$\frac{24}{27}$
3	$\frac{1}{27}$	$\frac{3}{27}$	$\frac{9}{27}$
		1	$\frac{5}{3}$

Mean $\mu = \Sigma p_i x_i = 1$

Variance $\sigma^2 = \Sigma p_i x_i^2 - \mu^2 = \frac{5}{3} - 1 = \frac{2}{3}$.

Example 3. *A random variable X has the following probability function :*

Values of X,	x	:	0	1	2	3	4	5	6	7
	$p(x)$	:	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2 + k$

(i) Find k,

(ii) Evaluate $P(X < 6)$, $P(X \geq 6)$, $P(3 < X \leq 6)$

(iii) Find the minimum value of x so that $P(X \leq x) > \frac{1}{2}$.

Sol. (*i*) Since $\sum_{x=0}^{7} p(x) = 1$, we have

$$0 + k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$$

$\Rightarrow \quad 10k^2 + 9k - 1 = 0 \quad \Rightarrow \quad (10k - 1)(k + 1) = 0$

$\Rightarrow \quad k = \frac{1}{10}$ $\quad [\because \; p(x) \geq 0]$

(*ii*) $P(X < 6) = P(X = 0) + P(X = 1) + \ldots\ldots + P(X = 5)$

$$= 0 + k + 2k + 2k + 3k + k^2 = 8k + k^2 = \frac{8}{10} + \frac{1}{100} = \frac{81}{100}$$

$P(X \geq 6) = P(X = 6) + P(X = 7)$

$$= 2k^2 + 7k^2 + k = \frac{9}{100} + \frac{1}{10} = \frac{19}{100}$$

$P(3 < X \leq 6) = P(X = 4) + P(X = 5) + P(X = 6)$

$$= 3k + k^2 + 2k^2 = \frac{3}{10} + \frac{3}{100} = \frac{33}{100}$$

(*iii*) $P(X \le 1) = k = \frac{1}{10} < \frac{1}{2}$; $\quad P(X \le 2) = k + 2k = \frac{3}{10} < \frac{1}{2}$

$P(X \le 3) = k + 2k + 2k = \frac{5}{10} = \frac{1}{2}$; $\quad P(X \le 4) = k + 2k + 2k + 3k = \frac{8}{10} > \frac{1}{2}$

$\therefore$ The maximum value of x so that $P(X \le x) > \frac{1}{2}$ is 4.

TEST YOUR KNOWLEDGE

1. Find the probability distribution of the number of doublets in four throws of a pair of dice.
2. Two bad eggs are mixed accidently with 10 good ones. Find the probability distribution of the number of bad eggs in 3, drawn at random, without replacement, from this lot.
3. A die is tossed twice. Getting a number greater than 4 is considered a success. Find the variance of the probability distribution of the number of successes.
4. Two cards are drawn simultaneously from a well-shuffled deck of 52 cards. Compute the variance for the number of aces.
5. An urn contains 4 white and 3 red balls. Three balls are drawn, with replacement, from this urn. Find μ, σ^2 and σ for the number of red balls drawn.
6. A random variable X has the following probability distribution :

Values of X, x :	0	1	2	3	4	5	6	7	8
$p(x)$:	a	$3a$	$5a$	$7a$	$9a$	$11a$	$13a$	$15a$	$17a$

(*i*) Determine the value of a. (*ii*) Find $P(X < 3)$, $P(X \ge 3)$, $P(2 \le X < 5)$

(*iii*) What is the smallest value of x for which $P(X \le x) > 0.5$?

7. Find the standard deviation for the following discrete distribution :

x :	8	12	16	20	24
$p(x)$:	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{12}$

Answers

1.

X :	0	1	2	3	4
P(X) :	$\frac{625}{1296}$	$\frac{500}{1296}$	$\frac{150}{1296}$	$\frac{20}{1296}$	$\frac{1}{1296}$

2.

X :	0	1	2
P(X) :	$\frac{12}{22}$	$\frac{9}{22}$	$\frac{1}{22}$

3. $\frac{4}{9}$ 4. $\frac{400}{2873}$ 5. $\frac{9}{7}, \frac{36}{49}, \frac{6}{7}$

6. (*i*) $a = \frac{1}{81}$ (*ii*) $\frac{1}{9}, \frac{8}{9}, \frac{7}{27}$ (*iii*) 5 7. $2\sqrt{5}$.

4.4. THEORETICAL DISTRIBUTIONS

Frequency distributions can be classified under two heads :

(*i*) Observed Frequency Distributions.

(*ii*) Theoretical or Expected Frequency Distributions.

Observed frequency distributions are based on actual observation and experimentation. If certain hypothesis is assumed, it is sometimes possible to derive mathematically what the frequency distribution of certain universe should be. Such distributions are called **Theoretical Distributions.**

There are many types of theoretical frequency distributions but we shall consider only three which are of great importance :

(*i*) Binomial Distribution (or Bernoulli's Distribution) ;

(*ii*) Poisson's Distribution ;

(*iii*) Normal Distribution.

BINOMIAL (OR BERNOULLI'S) DISTRIBUTION

4.5. BINOMIAL PROBABILITY DISTRIBUTION

Let there be n independent trials in an experiment. Let a random variable X denote the number of successes in these n trials. Let p be the probability of a success and q that of a failure in a single trial so that $p + q = 1$. Let the trials be independent and p be constant for every trial.

Let us find the probability of r successes in n trials.

r successes can be obtained in n trials in ${}^nC_r$ ways.

$$\therefore \quad P(X = r) = {}^nC_r P(\underbrace{S\,S\,S\ldots\ldots S}_{r \text{ times}} \quad \underbrace{F\,F\,F\ldots\ldots F}_{(n-r) \text{ times}})$$

$$= {}^nC_r \underbrace{P(S)\,P(S)\ldots\ldots P(S)}_{r \text{ factors}} \quad \underbrace{P(F)\,P(F)\ldots\ldots P(F)}_{(n-r) \text{ factors}}$$

$$= {}^nC_r \underbrace{p\,p\,p\ldots\ldots p}_{r \text{ factors}} \quad \underbrace{q\,q\,q\ldots\ldots q}_{(n-r) \text{ factors}}$$

$$= {}^nC_r p^r q^{n-r} \qquad \ldots(1)$$

Hence $\quad \mathbf{P(X = r) = {}^nC_r\, q^{n-r} p^r}$, **where p + q = 1 and r = 0, 1, 2, ..., n.**

The distribution (1) is called the *binomial probability distribution* and X is called the *binomial variate.*

Note 1. $P(X = r)$ is usually written as $P(r)$.

Note 2. The successive probabilities $P(r)$ in (1) for $r = 0, 1, 2, \ldots\ldots, n$ are

$${}^nC_0 q^n,\ {}^nC_1 q^{n-1} p,\ {}^nC_2 q^{n-2} p^2,\ \ldots\ldots,\ {}^nC_n p^n$$

which are the successive terms of the binomial expansion of $(q + p)^n$. That is why this distribution is called "binomial" distribution.

Note 3. n and p occurring in the binomial distribution are called the *parameters* of the distribution.

Note 4. In a binomial distribution :

(*i*) n, the number of trials is finite.

(*ii*) each trial has only two possible outcomes usually called success and failure.

(*iii*) all the trials are independent.

(*iv*) p (and hence q) is constant for all the trials.

4.6. RECURRENCE OR RECURSION FORMULA FOR THE BINOMIAL DISTRIBUTION

In a binomial distribution,

$$P(r) = {}^nC_r\, q^{n-r} p^r = \frac{n!}{(n-r)!\, r!} q^{n-r} p^r$$

$$P(r+1) = {}^nC_{r+1}\, q^{n-r-1} p^{r+1} = \frac{n!}{(n-r-1)!\,(r+1)!} q^{n-r-1} p^{r+1}$$

$$\therefore \quad \frac{P(r+1)}{P(r)} = \frac{(n-r)!}{(n-r-1)!} \times \frac{r!}{(r+1)!} \times \frac{p}{q}$$

$$= \frac{(n-r)\times(n-r-1)!}{(n-r-1)!} \times \frac{r!}{(r+1)\times r!} \times \frac{p}{q} \times = \left(\frac{n-r}{r+1}\right) . \frac{p}{q}$$

$$\Rightarrow \quad P(r+1) = \frac{n-r}{r+1} . \frac{p}{q} P(r)$$

which is the required recurrence formula. Applying this formula successively, we can find P(1), P(2), P(3), ..., if P(0) is known.

4.7. MEAN AND VARIANCE OF THE BINOMIAL DISTRIBUTION

For the binomial distribution, $P(r) = {}^nC_r\, q^{n-r} p^r$ *(M.G.U. May 2010)*

$$\text{Mean } \mu = \sum_{r=0}^{n} rP(r) = \sum_{r=0}^{n} r \,.\, {}^nC_r q^{n-r} p^r$$

$$= 0 + 1 \,.\, {}^nC_1\, q^{n-1} p + 2 \,.\, {}^nC_2\, q^{n-2} p^2 + 3 \,.\, {}^nC_3\, q^{n-3} p^3 + \ldots\ldots + n \,.\, {}^nC_n\, p^n$$

$$= nq^{n-1} p + 2 \,.\, \frac{n(n-1)}{2.1} q^{n-2} p^2 + 3 \,.\, \frac{n(n-1)(n-2)}{3.2.1} q^{n-3} p^3 + \ldots\ldots + n.p^n$$

$$= nq^{n-1} p + n(n-1)\, q^{n-2} p^2 + \frac{n(n-1)(n-2)}{2.1} q^{n-3} p^3 + \ldots\ldots + np^n$$

$$= np\left[q^{n-1} + (n-1) q^{n-2} p + \frac{(n-1)(n-2)}{2.1} q^{n-3} p^2 + \ldots\ldots + p^{n-1}\right]$$

$$= np[{}^{n-1}C_0 q^{n-1} + {}^{n-1}C_1 q^{n-2} p + {}^{n-1}C_2 q^{n-3} p^2 + \ldots + {}^{n-1}C_{n-1} p^{n-1}]$$

$$= np(q+p)^{n-1} = np \qquad (\because \; p + q = 1)$$

Hence the mean of the binomial distribution is np.

$$\text{Variance } \sigma^2 = \sum_{r=0}^{n} r^2 P(r) - \mu^2 = \sum_{r=0}^{n} [r + r(r-1)]\, P(r) - \mu^2$$

$$= \sum_{r=0}^{n} rP(r) + \sum_{r=0}^{n} r(r-1)\, P(r) - \mu^2 = \mu + \sum_{r=2}^{n} r(r-1)\, {}^nC_r q^{n-r} p^r - \mu^2$$

(since the contribution due to $r = 0$ and $r = 1$ is zero)

$$= \mu + [2.1 \,.\, {}^nC_2 q^{n-2} p^2 + 3.2 \,.\, {}^nC_3 q^{n-3} p^3 + \ldots + n(n-1)\, {}^nC_n p^n] - \mu^2$$

$$= \mu + \left[2.1 . \frac{n(n-1)}{2.1} q^{n-2} p^2 + 3.2 . \frac{n(n-1)(n-2)}{3.2.1} q^{n-3} p^3 + \ldots + n(n-1) p^n\right] - \mu^2$$

Scanned by CamScanner

$$= \mu + [n(n-1)q^{n-2}p^2 + n(n-1)(n-2)q^{n-3}p^3 + \ldots + n(n-1)p^n] - \mu^2$$
$$= \mu + n(n-1)\,p^2[q^{n-2} + (n-2)q^{n-3}p + \ldots + p^{n-2}] - \mu^2$$
$$= \mu + n\,(n-1)p^2\,[{}^{n-2}C_0q^{n-2} + {}^{n-2}C_1q^{n-3}p + \ldots + {}^{n-2}C_{n-2}p^{n-2}] - \mu^2$$
$$= \mu + n(n-1)p^2(q+p)^{n-2} - \mu^2 = \mu + n(n-1)p^2 - \mu^2 \qquad [\because\ q + p = 1]$$
$$= np + n(n-1)p^2 - n^2p^2 \qquad [\because\ \mu = np]$$
$$= np[1 + (n-1)p - np] = np[1-p] = npq.$$

Hence the variance of the binomial distribution is npq.

Standard deviation of the binomial distribution is $\sqrt{npq}$.

Similarly, we can prove that

$$\beta_1 = \frac{{\mu_3}^2}{{\mu_2}^3} = \frac{(q-p)^2}{npq} = \frac{(1-2p)^2}{npq}\ ; \qquad \beta_2 = \frac{\mu_4}{{\mu_2}^2} = 3 + \frac{1-6pq}{npq}$$

Hence $\quad \gamma_1 = \sqrt{\beta_1} = \dfrac{q-p}{\sqrt{npq}} = \dfrac{1-2p}{\sqrt{npq}}\ ; \qquad \gamma_2 = \beta_2 - 3 = \dfrac{1-6pq}{\sqrt{npq}}.$

Note. $\gamma_1 = \dfrac{q-p}{\sqrt{npq}} = \dfrac{1-2p}{\sqrt{npq}}$ gives a **measure of skewness** of the binomial distribution. If $p < \frac{1}{2}$, skewness is positive, if $p > \frac{1}{2}$, skewness is negative and if $p = \frac{1}{2}$, it is zero.

$\beta_2 = 3 + \dfrac{1-6pq}{npq}$ gives a **measure of the kurtosis** of the binomial distribution.

ILLUSTRATIVE EXAMPLES

Example 1. *During war, 1 ship out of 9 was sunk on an average in making a certain voyage. What was the probability that exactly 3 out of a convoy of 6 ships would arrive safely ?*

Sol. p, the probability of a ship arriving safely $= 1 - \dfrac{1}{9} = \dfrac{8}{9}\ ; \quad q = \dfrac{1}{9}, n = 6$

Binomial distribution is $\left(\dfrac{1}{9} + \dfrac{8}{9}\right)^6$

The probability that exactly 3 ships arrive safely $= {}^6C_3\left(\dfrac{1}{9}\right)^3\left(\dfrac{8}{9}\right)^3 = \dfrac{10240}{9^6}.$

Example 2. *Assume that on the average one telephone number out of fifteen called between 2 P.M. and 3 P.M. on week-days is busy. What is the probability that if 6 randomly selected telephone numbers are called (i) not more than three, (ii) at least three of them will be busy ?*

Sol. p, the probability of a telephone number being busy between 2 P.M. and 3 P.M. on week-days $= \dfrac{1}{15}$

$$q = 1 - \frac{1}{15} = \frac{14}{15}, \quad n = 6\ ; \text{Binomial distribution is } \left(\frac{14}{15} + \frac{1}{15}\right)^6$$

The probability that not more than three will be busy

$$= p(0) + p(1) + p(2) + p(3)$$

$$= {}^6C_0\left(\frac{14}{15}\right)^6 + {}^6C_1\left(\frac{14}{15}\right)^5\left(\frac{1}{15}\right) + {}^6C_2\left(\frac{14}{15}\right)^4\left(\frac{1}{15}\right)^2 + {}^6C_3\left(\frac{14}{15}\right)^3\left(\frac{1}{15}\right)^3$$

$$= \frac{(14)^3}{(15)^6}\,[2744 + 1176 + 210 + 20] = \frac{2744 \times 4150}{(15)^6} = 0.9997$$

The probability that at least three of them will be busy

$$= p(3) + p(4) + p(5) + p(6)$$

$$= {}^6C_3\left(\frac{14}{15}\right)^3\left(\frac{1}{15}\right)^3 + {}^6C_4\left(\frac{14}{15}\right)^2\left(\frac{1}{15}\right)^4 + {}^6C_5\left(\frac{14}{15}\right)\left(\frac{1}{15}\right)^5 + {}^6C_6\left(\frac{1}{15}\right)^6 = 0.005.$$

Example 3. *Six dice are thrown 729 times. How many times do you expect at least three dice to show a five or six ?*

Sol. p = the chance of getting 5 or 6 with one die $= \frac{2}{6} = \frac{1}{3}$

$$q = 1 - \frac{1}{3} = \frac{2}{3},\ n = 6,\ N = 729$$

since dice are in sets of 6 and there are 729 sets.

The binomial distribution is $N(q+p)^n = 729\left(\frac{2}{3} + \frac{1}{3}\right)^6$

The expected number of times *at least three* dice showing five or six

$$= 729\left[{}^6C_3\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^3 + {}^6C_4\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^4 + {}^6C_5\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^5 + {}^6C_6\left(\frac{1}{3}\right)^6\right]$$

$$= \frac{729}{3^6}\,[160 + 60 + 12 + 1] = 233.$$

Example 4. *Out of 800 families with 4 children each, how many families would be expected to have (i) 2 boys and 2 girls (ii) at least one boy (iii) no girl (iv) at most two girls ? Assume equal probabilities for boys and girls.*

Sol. Since probabilities for boys and girls are equal

$$p = \text{probability of having a boy} = \frac{1}{2};\quad q = \text{probability of having a girl} = \frac{1}{2}$$

$$n = 4,\qquad N = 800\qquad \therefore\ \text{The binomial distribution is } 800\left(\frac{1}{2} + \frac{1}{2}\right)^4.$$

(*i*) The expected number of families having 2 boys and 2 girls

$$= 800\ {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = 800 \times 6 \times \frac{1}{16} = 300.$$

(*ii*) The expected number of families having *at least one boy*

$$= 800\left[{}^4C_1\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) + {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + {}^4C_3\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 + {}^4C_4\left(\frac{1}{2}\right)^4\right]$$

$$= 800 \times \frac{1}{16} [4 + 6 + 4 + 1] = 750.$$

(*iii*) The expected number of families having no girl, *i.e.*, having 4 boys

$$= 800 \cdot {}^4C_4 \left(\frac{1}{2}\right)^4 = 50.$$

(*iv*) The expected number of families having *at most two girls i.e.*, having *at least 2 boys*

$$= 800 \left[{}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 + {}^4C_3 \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 + {}^4C_4 \left(\frac{1}{2}\right)^4 \right] = 800 \times \frac{1}{16} [6 + 4 + 1] = 550.$$

TEST YOUR KNOWLEDGE

1. Ten coins are tossed simultaneously. Find the probability of getting at least seven heads.
2. The probability of any ship of a company being destroyed on a certain voyage is 0.02. The company owns 6 ships for the voyage. What is the probability of :
   (*i*) losing one ship (*ii*) losing at most two ships (*iii*) losing none.
3. The probability that a man aged 60 will live to be 70 is 0.65. What is the probability that out of ten men now 60, at least 7 would live to be 70 ?
4. The incidence of occupational disease in an industry is such that the workers have a 20% chance of suffering from it. What is the probability that out of six workers chosen at random, four or more will suffer from the disease ?
5. The probability that a pen manufactured by a company will be defective is $\frac{1}{10}$. If 12 such pens are manufactured, find the probability that
   (*i*) exactly two will be defective (*ii*) at least two will be defective
   (*iii*) none will be defective.
6. If the chance that one of the ten telephone lines is busy at an instant is 0.2.
   (*i*) What is the chance that 5 of the lines are busy ?
   (*ii*) What is the probability that all the lines are busy ?
7. If on an average 1 vessel in every 10 is wrecked, find the probability that out of 5 vessels expected to arrive, at least 4 will arrive safely.
8. A product is 0.5% defective and is packed in cartons of 100. What percentage contains not more than 3 defectives ?
9. A bag contains 5 white, 7 red and 8 black balls. If four balls are drawn, one by one, with replacement, what is the probability that
   (*i*) none is white (*ii*) all are white
   (*iii*) at least one is white (*iv*) only 2 are white ?
10. In a hurdle race, a player has to cross 10 hurdles. The probability that he will clear each hurdle is $\frac{5}{6}$. What is the probability that he will knock down fewer than 2 hurdles ?
11. Fit a binomial distribution for the following data and compare the theoretical frequencies with the actual ones :

x :	0	1	2	3	4	5
f :	2	14	20	34	22	8

12. If the sum of mean and variance of a binomial distribution is 4.8 for five trials, find the distribution.

13. If the mean of a binomial distribution is 3 and the variance is $\frac{3}{2}$, find the probability of obtaining at least 4 success.

14. In 800 families with 5 children each, how many families would be expected to have (*i*) 3 boys and 2 girls, (*ii*) 2 boys and 3 girls, (*iii*) no girl (*iv*) at the most two girls. (Assume probabilities for boys and girls to be equal.)

15. In 100 sets of ten tosses of an unbiased coin, in how many cases do you expect to get
(*i*) 7 heads and 3 tails (*ii*) at least 7 heads ?

16. The following data are the number of seeds germinating out of 10 on damp filter for 80 sets of seeds. Fit a binomial distribution to this data :

x :	0	1	2	3	4	5	6	7	8	9	10	Total
f :	6	20	28	12	8	6	0	0	0	0	0	80

[**Hint.** Here $n = 10$, N = 80, Mean = $\frac{\Sigma fx}{\Sigma f} = 2.175 \;\therefore\; np = 2.175$ etc.]

17. A bag contains 10 balls each marked with one of the digits 0 to 9. If four falls are drawn successively with replacement from the bag, what is the probability that none is marked with the digit 0 ?

18. A box contains 100 tickets each bearing one of the numbers from 1 to 100. If 5 tickets are drawn successively with replacement from the box, find the probability that all the tickets bear numbers divisible by 10.

19. The probability that a bomb dropped from a plane will strike the target is $\frac{1}{5}$. If six bombs are dropped, find the probability that (*i*) exactly two will strike the target, (*ii*) at least two will strike the target. (*M.G.U. Jan. 2007*)

20. In sampling a large number of parts manufactured by a machine, the mean number of defectives in a sample of 20 is 2. Out of 1000 such samples, how many would be expected to contain at least 3 defective parts.

21. Fit a binomial distribution to the following distribution.

x	0	1	2	3	4	5
f	27	14	6	3	0	0

22. If the probability of a defective bolt is 0.1, find (*a*) the mean (*b*) the standard deviation, for the number of defective bolt in a total of 400 bolts.

Answers

1. $\frac{11}{64}$ **2.** (*i*) 0.1085, (*ii*) 0.9997, (*iii*) 0.8858 **3.** 0.514

4. $\frac{53}{3125}$ **5.** (*i*) 0.2301 (*ii*) 0.3412 (*iii*) 0.2833

6. (*i*) 0.02579 (*ii*) 1.024×10^{-7} **7.** 0.91854 **8.** 99.83

9. (*i*) $\frac{81}{256}$ (*ii*) $\frac{1}{256}$ (*iii*) $\frac{175}{256}$ (*iv*) $\frac{27}{128}$ **10.** $\frac{5}{2}\left(\frac{5}{6}\right)^9$

11. $100\,(0.432 + 0.568)^5$ **12.** $\left(\frac{1}{5} + \frac{4}{5}\right)^5$ **13.** $\frac{11}{32}$

14. (*i*) 250, (*ii*) 250, (*iii*) 25, (*iv*) 400
15. (*i*) 12 nearly (*ii*) 17 nearly
16. 80 (0.7825 + 0.2175)^{10}
17. $\left(\frac{9}{10}\right)^4$
18. 0.00001
19. (*i*) 0.246 (*ii*) 0.345
20. 323
21. 50 (.86 + 0.14)^5
22. (*a*) mean = 40 (*b*) variance = 36, standard deviation = 6.

POISSON DISTRIBUTION

4.8. POISSON DISTRIBUTION AS A LIMITING CASE OF BINOMIAL DISTRIBUTION

If the parameters n and p of a binomial distribution are known, we can find the distribution. But in situations where n is very large and p is very small, application of binomial distribution is very labourious. However, if we assume that as $n \to \infty$ and $p \to 0$ such that np always remains finite, say λ, we get the Poisson approximation to the binomial distribution.

Now, for a binomial distribution

$$P(X = r) = {}^nC_r q^{n-r} p^r$$

$$= \frac{n(n-1)(n-2)\ldots(n-r+1)}{r!} \times (1-p)^{n-r} \times p^r$$

$$= \frac{n(n-1)(n-2)\ldots(n-r+1)}{r!} \times \left(1-\frac{\lambda}{n}\right)^{n-r} \times \left(\frac{\lambda}{n}\right)^r \qquad \text{since } np = \lambda \quad \therefore \quad p = \frac{\lambda}{n}$$

$$= \frac{\lambda^r}{r!} \times \frac{n(n-1)(n-2)\ldots(n-r+1)}{n^r} \times \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^r}$$

$$= \frac{\lambda^r}{r!}\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)\ldots\ldots\left(\frac{n-r+1}{n}\right) \times \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^r}$$

$$= \frac{\lambda^r}{r!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\ldots\left(1-\frac{r-1}{n}\right) \times \frac{\left[\left(1-\frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}\right]^{-\lambda}}{\left(1-\frac{\lambda}{n}\right)^r}$$

As $n \to \infty$, each of the $(r-1)$ factors

$$\left(1-\frac{1}{n}\right), \left(1-\frac{2}{n}\right), \ldots\ldots, \left(1-\frac{r-1}{n}\right) \quad \text{tends to 1. Also } \left(1-\frac{\lambda}{n}\right)^r \text{ tends to 1.}$$

Since $\underset{x \to \pm\infty}{\text{Lt}} \left(1+\frac{1}{x}\right)^x = e$, the Naperian base. $\therefore \quad \left[\left(1-\frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}\right]^{-\lambda} \longrightarrow e^{-\lambda}$ as $n \to \infty$

Hence in the limiting case when $n \to \infty$, we have

$$P(X = r) = \frac{\lambda^r e^{-\lambda}}{r!} \quad (r = 0, 1, 2, 3, \ldots) \qquad \ldots(A)$$

where λ is a finite number $= np$.

(A) Represents probability distribution which is called the Poisson probability distribution.

Note 1. λ is called the parameter of the distribution.

Note 2. $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots\ldots + \frac{x^n}{n!} + \ldots\ldots$ to ∞.

Note 3. The sum of the probabilities $P(r)$ for $r = 0, 1, 2, 3, \ldots\ldots$ is 1, since

$$P(0) + P(1) + P(2) + P(3) + \ldots\ldots = e^{-\lambda} + \frac{\lambda e^{-\lambda}}{1!} + \frac{\lambda^2 e^{-\lambda}}{2!} + \frac{\lambda^3 e^{-\lambda}}{3!} + \ldots\ldots$$

$$= e^{-\lambda}\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots\ldots\right) = e^{-\lambda} \cdot e^{\lambda} = 1.$$

4.9. RECURRENCE FORMULA FOR THE POISSON DISTRIBUTION

For Poisson distribution, $P(r) = \frac{\lambda^r e^{-\lambda}}{r!}$ and $P(r + 1) = \frac{\lambda^{r+1} e^{-\lambda}}{(r+1)!}$

$$\therefore \quad \frac{P(r+1)}{P(r)} = \frac{\lambda r!}{(r+1)!} = \frac{\lambda}{r+1} \quad \text{or} \quad P(r + 1) = \frac{\lambda}{r+1} P(r),\ r = 0, 1, 2, 3, \ldots\ldots$$

This is called the recurrence formula for the Poisson distribution.

4.10. MEAN AND VARIANCE OF THE POISSON DISTRIBUTION

For the Poisson distribution, $P(r) = \frac{\lambda^r e^{-\lambda}}{r!}$

$$\text{Mean } \mu = \sum_{r=0}^{\infty} rP(r) = \sum_{r=0}^{\infty} r \cdot \frac{\lambda^r e^{-\lambda}}{r!}$$

$$= e^{-\lambda} \sum_{r=1}^{\infty} \frac{\lambda^r}{(r-1)!} = e^{-\lambda}\left(\lambda + \frac{\lambda^2}{1!} + \frac{\lambda^3}{2!} + \ldots\ldots\right)$$

$$= \lambda e^{-\lambda}\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \ldots\ldots\right) = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda$$

Thus, the mean of the Poisson distribution is equal to the parameter λ.

$$\text{Variance } \sigma^2 = \sum_{r=0}^{\infty} r^2 P(r) - \mu^2 = \sum_{r=0}^{\infty} r^2 \cdot \frac{\lambda^r e^{-\lambda}}{r!} - \lambda^2 = e^{-\lambda} \sum_{r=1}^{\infty} \frac{r^2 \lambda^r}{r!} - \lambda^2$$

$$= e^{-\lambda}\left[\frac{1^2 \cdot \lambda}{1!} + \frac{2^2 \cdot \lambda^2}{2!} + \frac{3^2 \lambda^3}{3!} + \frac{4^2 \lambda^4}{4!} + \ldots\ldots\right] - \lambda^2$$

$$= \lambda e^{-\lambda}\left[1+\frac{2\lambda}{1!}+\frac{3\lambda^2}{2!}+\frac{4\lambda^3}{3!}+\ldots\ldots\right]-\lambda^2$$

$$= \lambda e^{-\lambda}\left[1+\frac{(1+1)\lambda}{1!}+\frac{(1+2)\lambda^2}{2!}+\frac{(1+3)\lambda^3}{3!}+\ldots\ldots\right]-\lambda^2$$

$$= \lambda e^{-\lambda}\left[\left(1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\ldots\ldots\right)+\left(\frac{\lambda}{1!}+\frac{2\lambda^2}{2!}+\frac{3\lambda^3}{3!}+\ldots\ldots\right)\right]-\lambda^2$$

$$= \lambda e^{-\lambda}\left[e^{\lambda}+\lambda\left(1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\ldots\ldots\right)\right]-\lambda^2$$

$$= \lambda e^{-\lambda}[e^{\lambda}+\lambda e^{\lambda}]-\lambda^2 = \lambda e^{-\lambda}\,.\,e^{\lambda}(1+\lambda)-\lambda^2 = \lambda(1+\lambda)-\lambda^2 = \lambda.$$

Hence, the variance of the Poisson distribution is also λ.

Thus, the mean and the variance of the Poisson distribution are each equal to the parameter λ.

Note. The mean and the variance of the Poisson distribution can also be derived from those of the binomial distribution in the limiting case when $n \to \infty$, $p \to 0$ and $np = \lambda$.

Mean of binomial distribution is np.

$\therefore$ Mean of Poisson distribution $= \underset{n\to\infty}{\text{Lt}}\ np = \underset{n\to\infty}{\text{Lt}}\ \lambda = \lambda$

Variance of binomial distribution is $npq = np\,(1-p)$

$\therefore$ Variance of Poisson distribution $= \underset{n\to\infty}{\text{Lt}}\ np\,(1-p) = \underset{n\to\infty}{\text{Lt}}\ \lambda\left(1-\frac{\lambda}{n}\right) = \lambda.$

ILLUSTRATIVE EXAMPLES

Example 1. *If the variance of the Poisson distribution is 2, find the probabilities for $r = 1, 2, 3, 4$ from the recurrence relation of the Poisson distribution.*

Sol. λ, the parameter of Poisson distribution = Variance = 2

Recurrence relation for the Poisson distribution is

$$P(r+1) = \frac{\lambda}{r+1}\,P(r) = \frac{2}{r+1}\,P(r) \qquad \ldots(1)$$

Now $\quad P(r) = \dfrac{\lambda^r e^{-\lambda}}{r!} \quad \Rightarrow \quad P(0) = \dfrac{e^{-2}}{0!} = e^{-2} = 0.1353$

Putting $r = 0, 1, 2, 3$ in (1), we get

$$P(1) = 2P(0) = 2 \times 0.1353 = 0.2706\,; \qquad P(2) = \frac{2}{2}\,P(1) = 0.2706$$

$$P(3) = \frac{2}{3}\,P(2) = \frac{2}{3} \times 0.2706 = 0.1804\,; \qquad P(4) = \frac{2}{4}\,P(3) = \frac{1}{2} \times 0.1804 = .0902.$$

Example 2. *Assume that the probability of an individual coalminer being killed in a mine accident during a year is $\frac{1}{2400}$. Use Poisson's distribution to calculate the probability that in a mine employing 200 miners there will be at least one fatal accident in a year.*

Sol. Here $p = \frac{1}{2400}, n = 200$; $\therefore$ $\lambda = np = \frac{200}{2400} = \frac{1}{12} = 0.083$

$$\therefore \quad P(r) = \frac{\lambda^r e^{-\lambda}}{r!} = \frac{(0.083)^r e^{-.083}}{r!}$$

P(at least one fatal accident) = 1 – P (no fatal accident)

$$= 1 - P(0) = 1 - \frac{(0.083)^0 e^{-0.83}}{0!} = 1 - .92 = 0.08.$$

Example 3. *Data was collected over a period of 10 years, showing number of deaths from horse kicks in each of the 200 army corps. The distribution of deaths was as follows :*

No. of deaths :	*0*	*1*	*2*	*3*	*4*	*Total*
Frequency :	*109*	*65*	*22*	*3*	*1*	*200*

Fit a Poisson distribution to the data and calculate the theoretical frequencies :

Sol. Mean of given distribution $= \frac{\Sigma fx}{\Sigma f} = \frac{65 + 44 + 9 + 4}{200} = \frac{122}{200} = 0.61$

This is the parameter (m) of the Poisson distribution.

$\therefore$ Required Poisson distribution is $N \cdot \frac{m^r e^{-m}}{r!}$ where $N = \Sigma f = 200$

$$= 200e^{-0.61} \cdot \frac{(0.61)^r}{r!} = 200 \times 0.5435 \frac{(0.61)^r}{r!} = 108.7 \times \frac{(0.61)^2}{r!}.$$

Scanned by CamScanner

Sol. Since the number of demands for a car is distributed as a Poisson distribution with mean $m = 1.5$.

$\therefore$ Proportion of days on which neither car is used

$= $ Probability of there being no demand for the car

$$= \frac{m^0 e^{-m}}{0!} = e^{-1.5} = 0.2231$$

Proportion of days on which some demand is refused

$=$ probability for the number of demands to be more than two

$$= 1 - P(x \le 2) = 1 - \left(e^{-m} + \frac{me^{-m}}{1!} + \frac{m^2 e^{-m}}{2!}\right)$$

$$= 1 - e^{-1.5}\left(1 + 1.5 + \frac{(1.5)^2}{2}\right) = 1 - 0.2231\,(1 + 1.5 + 1.125)$$

$$= 1 - 0.2231 \times 3.625 = 1 - 0.8087375 = 0.1912625.$$

Example 5. *Six coins are tossed 6400 times. Using the Poisson distribution, determine the approximate probability of getting six heads x times.*

Sol. Probability of getting one head with one coin $= \frac{1}{2}$.

$\therefore$ The probability of getting six heads with six coins $= \left(\frac{1}{2}\right)^6 = \frac{1}{64}$

$\therefore$ Average number of six heads with six coins in 6400 throws $= np = 6400 \times \frac{1}{64} = 100$

$\therefore$ The mean of the Poisson distribution $= 100$.

Approximate probability of getting six heads x times when the distribution is Poisson

$$= \frac{m^x e^{-m}}{x!} = \frac{(100)^x \cdot e^{-100}}{(100)!}.$$

TEST YOUR KNOWLEDGE

1. Fit a Poisson distribution to the following :

x :	0	1	2	3	4
f :	192	100	24	3	1

2. If the probability of a bad reaction from a certain injection is 0.001, determine the chance that out of 2000 individuals more than two will get a bad reaction.
3. If X is a Poisson variate such that $P(X = 2) = 9P(X = 4) + 90P(X = 6)$, find the standard deviation.
4. If a random variable has a Poisson distribution such that $P(1) = P(2)$, find
   (*i*) mean of the distribution (*ii*) $P(4)$.
5. Suppose that X has a Poisson distribution. If $P(X = 2) = \frac{2}{3} P(X = 1)$ find, (*i*) $P(X = 0)$ (*ii*) $P(X = 3)$.
6. A certain screw making machine produces on average 2 defective screws out of 100, and packs them in boxes of 500. Find the probability that a box contains 15 defective screws.
7. The incidence of occupational disease in an industry is such that the workmen have a 10% chance of suffering from it. What is the probability that in a group of 7, five or more will suffer from it ?

8. Fit a Poisson distribution to the following and calculate theoretical frequencies :

x :	0	1	2	3	4
f :	122	60	15	2	1

9. Fit a Poisson distribution to the following data given the number of yeast cells per square for 400 squares :

No. of cells per sq. :	0	1	2	3	4	5	6	7	8	9	10
No. of squares :	103	143	98	42	8	4	2	0	0	0	0

10. Show that in a Poisson distribution with unit mean, mean deviation about mean is $\left(\frac{2}{e}\right)$ times the standard deviation.

11. In a certain factory turning razor blades, there is a small chance of 0.002 for any blade to be defective. The blades are supplied in packets of 10. Use Poisson distribution to calculate the approximate number of packets containing no defective, one defective and two defective blades respectively in a consignment of 10000 packets.

12. The probability that a man aged 35 years will die before reaching the age of 40 years may be taken as 0.018. Out of a group of 400 men, now aged 35 years, what is the probability that 2 men will die within the next 5 years ?

13. Suppose a book of 585 pages contains 43 typographical errors. If these errors are randomly distributed throughout the book, what is the probability that 10 pages, selected at random, will be free from errors ?

14. If x has a Poisson distribution with parameter λ, and if $P(x = 0) = 0.2.$, evaluate $P(x > 2)$ *(M.G.U. Dec. 2007)*

15. If 3% of the electric bulbs manufactured by a company are defective, find the probability that in a sample of 100 bulbs (*a*) 0, (*b*) 1, (*c*) 2, (*d*) 3, (*e*) 4 bulbs will be defective.

16. The probabilities of a poisson variate taking values 3 and 4 are equal. Calculate the probabilities of the variate taking the values 0 and 1. *(M.G.U. Jan. 2007)*

17. Fit a Poisson distribution for the following data and hence calculate the theoretical frequencies :

x :	0	1	2	3	4	5
f :	142	156	69	27	5	1

(M.G.U. May 2010)

Answers

1. $320 \times \frac{e^{0.503}(9.503)^r}{r!}$ **2.** 0.32 **3.** 1 **4.** (*i*) 2 (*ii*) $\frac{2}{3e^2}$

5. (*i*) e^{-4} (*ii*) $4e^{-4}$ **6.** $\frac{(10)^{15}e^{-10}}{(15)!} = 0.035$ **7.** 0.0008

8. $121.36 \times \frac{(0.5)^r}{r!}$, where $r = 0, 1, 2, 3, 4$

Theoretical frequencies are 121, 61, 15, 3, 0 respectively

9. Theoretical frequencies are 109, 142, 92, 40, 13, 3, 1, 0, 0, 0, 0

11. 9802, 196, 2 **12.** 0.01936 **13.** 0.4795

14. $P(x > 2) = 1 - (0.2)\left(1 + \ln 5 + \frac{(\ln 5)^2}{2}\right)$.

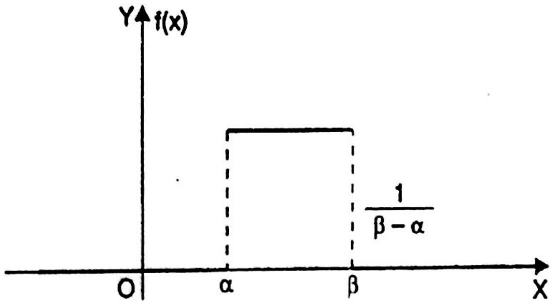
15. (*a*) 0.04979 (*b*) 0.1494 (*c*) 0.2241 (*d*) 0.2241 (*e*) 0.1680

16. $P(0) = e^{-4}$, $P(1) = 4e^{-4}$.

17. $P(x) = \frac{147.1517}{x!}$, $x = 0, 1, 2, 3, 4, 5$ theoretical frequencies are 147, 147, 74, 25, 6, 1.

4.11. UNIFORM DISTRIBUTION

The uniform distribution is a continuous distribution. The probability density function of the uniform distribution is given by $f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} \end{cases}$ for $\alpha < x < \beta$ elsewhere α and β are the parameters in the function. The graph of uniform distribution is as follows :



The mean of uniform distribution is

$$\mu = \int_{\alpha}^{\beta} x \frac{1}{\beta - \alpha}\, dx = \frac{\alpha + \beta}{2}$$

The variance of uniform distribution is $\sigma^2 = \dfrac{1}{12}(\beta - \alpha)^2$

Proof : Let $\mu_2^1 = \displaystyle\int_{\alpha}^{\beta} x^2 \frac{1}{\beta - \alpha}\, dx = \frac{\alpha^2 + \alpha\beta + \beta^2}{3}.$

The variance $\sigma^2 = \displaystyle\int_{\alpha}^{\beta} (x - \mu)^2 \frac{1}{\beta - \alpha}\, dx$

$$= \int_{\alpha}^{\beta} x^2 \frac{1}{\beta - \alpha}\, dx + \mu^2 \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha}\, dx - \mu \int_{\alpha}^{\beta} 2x \frac{1}{\beta - \alpha}\, dx$$

$$= \frac{\alpha^2 + \beta\alpha + \beta^2}{3} + \mu^2 - 2\mu\left(\frac{\alpha + \beta}{2}\right)$$

$$= \frac{\alpha^2 + \beta\alpha + \beta^2}{3} + \mu^2 - 2\mu^2$$

$$= \frac{\alpha^2 + \beta\alpha + \beta^2}{3} - \mu^2$$

$$= \frac{\alpha^2 + \beta\alpha + \beta^2}{3} - \left(\frac{\alpha + \beta}{2}\right)^2$$

$$= \frac{1}{12}(\beta - \alpha)^2.$$

4.12. EXPONENTIAL DISTRIBUTION

The exponential distribution is a continuous distribution. A continuous random variable X assuming non-negative values is said to have an exponential distribution with parameter $\alpha > 0$, if its probability density function is given by

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

The mean of the distribution

$$\mu = \int_0^\infty x \alpha e^{-\alpha x} \, dx = \frac{1}{\alpha}$$

The variance of the distribution

$$\sigma^2 = \int_0^\infty (x - \mu)^2 \alpha e^{-\alpha x} \, dx$$

$$= \int_0^\infty x^2 \alpha e^{-\alpha x} \, dx - \mu^2$$

$$= \frac{2}{\alpha^2} - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}.$$

4.13. NORMAL DISTRIBUTION

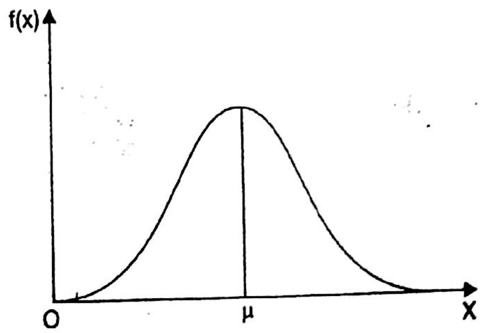
The normal distribution is a continuous distribution. It can be derived from the binomial distribution in the limiting case when n, the number of trials is very large and p, the probability of a success, is close to $\frac{1}{2}$. The general equation of the normal distribution is given by

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

where the variable x can assume all values from $-\infty$ to $+\infty$. μ and σ, called the parameters of the distribution, are respectively the mean and the standard deviation of the distribution and $-\infty < \mu < \infty$, $\sigma > 0$. x is called the normal variate and $f(x)$ is called probability density function of the normal distribution.

If a variable x has the normal distribution with mean μ and standard deviation σ, we briefly write $x : N(\mu, \sigma^2)$.

The graph of the normal distribution is called the *normal curve*. It is bell-shaped and symmetrical about the mean μ. The two tails of the curve extend to $+\infty$ and $-\infty$ towards the positive and negative directions of the x-axis respectively and gradually approach the x-axis without ever metting it. The curve is unimodal and the mode of the normal distribution coincides with its mean μ. The line $x = \mu$ divides the area under the normal curve above x-axis into two equal parts. Thus, the median of the distribution also coincides with its mean and mode. The area under the normal curve between any two given ordinates $x = x_1$ and $x = x_2$ represents the probability of values falling into the given interval. The total area under the normal curve above the x-axis is 1.



4.14. BASIC PROPERTIES OF THE NORMAL DISTRIBUTION

The probability density function of the normal distribution is given by

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

(*i*) $f(x) \geq 0$ (*ii*) $\int_{-\infty}^{\infty} f(x)\, dx = 1,$

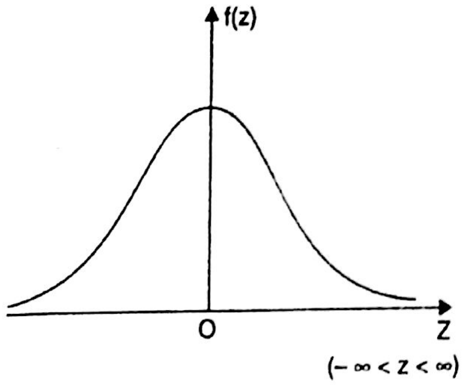
i.e., the total area under the normal curve above the x-axis is 1.

(*iii*) The normal distribution is symmetrical about its mean.

(*iv*) It is a unimodal distribution. The mean, mode and median of this distribution coincide.

4.15. STANDARD FORM OF THE NORMAL DISTRIBUTION

If X is a normal random variable with mean μ and standard deviation σ, then the random variable $Z = \frac{X-\mu}{\sigma}$ has the normal distribution with mean 0 and standard deviation 1. The random variable Z is called the *standardized* (or *standard*) normal random variable.



The probability density function for the normal distribution in standard form is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

It is free from any parameter. This helps us to compute areas under the normal probability curve by making use of standard tables.

Note 1. If $f(z)$ is the probability density function for the normal distribution, then

$$P(z_1 \leq Z \leq z_2) = \int_{z_1}^{z_2} f(z)\, dz = F(z_2) - F(z_1), \quad \text{where} \quad F(z) = \int_{-\infty}^{z} f(z)\, dz = P(Z \leq z)$$

The function F(z) defined above is called the *distribution function* for the normal distribution.

Note 2. The probabilities $P(z_1 \leq Z \leq z_2)$, $P(z_1 < Z \leq z_2)$, $P(z_1 \leq Z < z_2)$ and $P(z_1 < Z < z_2)$ are all regarded to be the same.

Note 3. $F(-z_1) = 1 - F(z_1)$.

ILLUSTRATIVE EXAMPLES

Example 1. *A sample of 100 dry battery cells tested to find the length of life produced the following results :*

$$\bar{x} = 12 \text{ hours}, \sigma = 3 \text{ hours}.$$

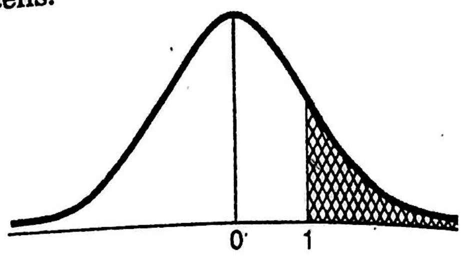
Assuming the data to be normally distributed, what percentage of battery cells are expected to have life

(*i*) *more than 15 hours* (*ii*) *less than 6 hours*

(*iii*) *between 10 and 14 hours ?*

Sol. Here x denotes the length of life of dry battery cells.

Also $$z = \frac{x - \bar{x}}{\sigma} = \frac{x - 12}{3}.$$

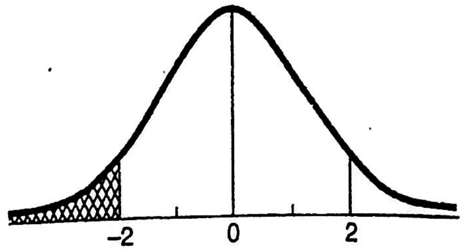


(*i*) When $x = 15, z = 1$

$\therefore \quad P(x > 15) = P(z > 1)$

$= P(0 < z < \infty) - P(0 < z < 1)$

$= 0.5 - 0.3413 = 0.1587 = 15.87\%.$



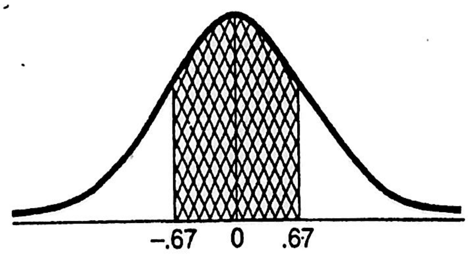
(*ii*) When $x = 6, \ z = -2$

$\therefore \quad P(x < 6) = P(z < -2)$

$= P(z > 2) = P(0 < z < \infty) - P(0 < z < 2)$

$= 0.5 - 0.4772 = 0.0228 = 2.28\%.$

(*iii*) When $x = 10, z = -\frac{2}{3} = -0.67$



When $x = 14, z = \frac{2}{3} = 0.67$

$P(10 < x < 14)$

$= P(-0.67 < z < 0.67)$

$= 2P(0 < z < 0.67) = 2 \times 0.2487$

$= 0.4974 = 49.74\%.$

Example 2. *In a normal distribution, 31% of the items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution.*

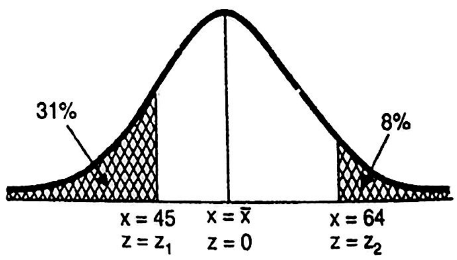
Sol. Let $\bar{x}$ and σ be the mean and S.D. respectively.

31% of the items are under 45.

$\Rightarrow$ Area to the left of the ordinate $x = 45$ is 0.31

When $x = 45$, let $z = z_1$

$P(z_1 < z < 0) = 0.5 - 0.31 = 0.19$



From the tables, the value of z corresponding to this area is 0.5

$\therefore \quad z_1 = -0.5 [z_1 < 0]$

When $x = 64$, let $z = z_2$.

$P(0 < z < z_2) = 0.5 - 0.08 = 0.42$

From the tables, the value of z corresponding to this area is 1.4.

$z_2 = 1.4$

Since $$z = \frac{x - \bar{x}}{\sigma}$$

$$-0.5 = \frac{45 - \bar{x}}{\sigma} \quad \text{and} \quad 1.4 = \frac{64 - \bar{x}}{\sigma}$$

$\Rightarrow \quad 45 - \bar{x} = -0.5\sigma \quad \quad ...(1)$

and $\quad 64 - \bar{x} = 1.4\sigma \quad \quad ...(2)$

Subtracting $\quad -19 = -1.9\sigma \quad \therefore \quad \sigma = 10$

From (1), $\quad 45 - \bar{x} = -0.5 \times 10 = -5 \quad \therefore \quad \bar{x} = 50.$

Example 3. *If X has normal distribution with mean* μ *and variance* σ^2, *find* $P[\mu - \sigma < X < \mu + \sigma]$. (M.G.U. Dec 2007)

Sol. Let X follows $N(\mu, \sigma)$.

Then $\dfrac{X-\mu}{\sigma} = T$ follows $N(0, 1)$

When $\mu - \sigma < X < \mu + \sigma$, the variable T lies between -1 and 1.

So $\quad p[\mu - \sigma < X < \mu + \sigma]$

$$= p\left[-1 < \frac{X-\mu}{\sigma} < 1\right]$$

$$= \int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt = .6826$$

TEST YOUR KNOWLEDGE

1. In a certain examination, the percentage of candidate passing and getting distinctions were 45 and 9 respectively. Estimate the average marks obtained by the candidates, if the minimum pass and distinction marks being 40 and 75 respectively. (Assume the distribution of marks to be normal).
2. A random variable has a normal distribution with $\sigma = 10$. If the probability is 0.8212 that it will take on a value less than 82.5. What is the probability that it will take on a value greater than 58.3.
3. Fit a Binomial distribution to the following data :

x :	0	1	2	3	4
f :	4	30	36	25	5

4. The mean weight of 500 students is 52 kgs and SD is 15 kgs . Assume the weights are normally distributed find how many students weight lie between 40 kgs and 50 kgs.
5. The mean height of 500 male students in a certain college is 151 cm and the standard deviation is 15 cm. Assuming the heights are normally distributed, find how many students have heights between 120 and 155 cm ?
6. An aptitude test for selecting officers in a bank is conducted on 1000 candidates. The average score is 42 and the standard deviation of score is 24. Assuming normal distribution for the scores, find

   (*i*) The number of candidates whose scores exceed 60

   (*ii*) The number of candidates whose scores lie between 30 and 60.
7. 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 ?
8. Let X denote the number of scores on a test. If X is normally distributed with mean 100 and standard deviation 15, find the probability that X does not exceed 130.
9. It is known from the past experience that the number of telephone calls made daily in a certain community between 3 p.m. and 4 p.m. have a mean of 352 and a standard deviation of 31. What percentage of the time will there be more than 400 telephone calls made in this community between 3 p.m. and 4 p.m. ?
10. Students of a class were given a mechanical aptitude test. Their marks were found to be normally distributed with mean 60 and standard deviation 5. What per cent of students scored

    (*i*) more than 60 marks ? (*ii*) less than 56 marks ?

    (*iii*) between 45 and 65 marks ?

11. In an examination taken by 500 candidates, the average and the standard deviation of marks obtained (normally distributed) are 40% and 10%. Find approximately :

 (*i*) How many will pass, if 50% is fixed as a minimum ?

 (*ii*) What should be the minimum if 350 candidates are to pass ?

 (*iii*) How many have scored marks above 60% ?

12. If X is normally distributed with mean 5 and standard deviation 2, find $p(X > 8)$.

13. In a distribution exactly normal, 10.03% of the items are under 25 kgs weight and 97% of the items are under 70 kgs weight. What are the mean and standard deviation of the distribution ?

14. The mean and standard deviation of marks obtained by 1000 students in an examination are respectively 34.4 and 16.5. Assuming the normality of the distribution find the approximate number of students expected to obtain marks between 30 and 60.

15. In a normal distribution, 7% of the items are under 35 and 10% of the items are above 55. Find the mean and variance of the distribution. (*M.G.U. May 2010*)

16. In a normal distribution 9% of the items are under 30 and 85% are under 65. Find the mean and standard deviation. (*M.G.U. Nov. 2009*)

17. A normal population has mean 0.1 and SD, $\sigma = 2.1$. Find the probability that the mean of a sample of size 900 will be negative.

Answers

1. Average = 36.4 **2.** $\mu = 73.48$, Required probability = .9357

3. $\overline{X} = 1.97$, $f(x) = {}^4C_x\,(.4925)^x\,(.5075)^{4-x}$ **4.** 118.

5. 300 **6.** (*i*) 252 (*ii*) 533 **7.** $\bar{x} = 50.3$, $\sigma = 10.33$

8. 0.9772 **9.** 6.06% **10.** (*i*) 50% (*ii*) 21.2% (*iii*) 84%

11. (*i*) 79 (*ii*) 35% (*iii*) 11. **12.** 0.0668

13. $\mu = 47.5$, $\sigma = 15.578$ **14.** 543

15. mean $\mu = 45.69$, variance $\sigma^2 = 52.128$

16. mean $\mu = 49.7058$, standard deviation $\sigma = 14.706$ **17.** Probability = 0.0764.