Chapter 27

Probability Distributions

INTRODUCTION

Probability distribution is the theoretical counterpart of frequency distribution, and plays an important role in the theoretical study of populations. A probability model can be developed, for a given idealized conditions in a game of chance by incorporating all the factors that have a bearing on this game. In building such model, the empirical data of frequency distribution, A.M., variance etc. are to be taken into account. In the discrete case we consider discrete uniform distribution, Binomial, Hypergeometric, Poisson distributions. The continuous probability distributions we study are uniform distribution, normal distribution, exponential, gamma, Weibull distributions which are of great practical importance.

Recall that in a random experiment, the outcomes (or results) are governed by chance mechanism and the sample space S of such a random experiment consists of all outcomes of the experiment. When the elements (outcomes/events) of the sample space are non-numeric, they can be quantified by assigning a real number to every event of the sample space. This assignment rule, known as the random variable (R.V.), provides the power of abstraction and thus discards unimportant finest-grain description of the sample space.

A **random variable** X on a sample space S is a function $X: S \to R$ from S to the set of real numbers R, which assigns a real number $X(s)$ to each sample point s of S (refer Fig. 27.1).

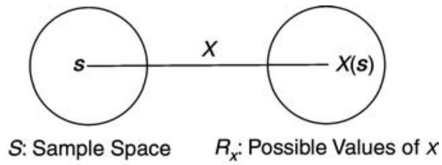


Fig. 27.1

Range space R_X: is the set of all possible values of X is a subset of real numbers R.

Although X is called a random "variable" note that it is infact a "single-valued function".

Notation: If R.V. is denoted by X, then x (corresponding small letter) denotes one of its values.

Discrete

A R.V. X is said to be discrete R.V. if its set of possible outcomes, the sample space S, is countable (finite or an unending sequence with as many elements as there are whole numbers).

Continuous

A R.V. X is said to be continuous R.V. if S contains infinite numbers equal to the number of points on a line segment.

27.1 PROBABILITY DISTRIBUTIONS

Discrete Probability Distributions

Each event in a sample space has certain probability (or chance) of occurrence (or happening). A formula representing all these probabilities which a discrete R.V. assumes, is known as the discrete probability distribution.

Example: Let X denote the discrete R.V. which denotes the minimum of the two numbers that appear in a single throw of a pair of fair dice. Then X is a function from the sample space S consisting of 36 ordered pairs $\{(1, 1), (1, 2), \ldots, (6, 6)\}$ to a subset of real numbers $\{1, 2, 3, 4, 5, 6\}$.

The event minimum 5 can appear in the following cases (occurrences) (5, 5), (5, 6), (6, 5). Thus R.V. X assigns to this event of the sample space a real number 3. The probability of such an event happening is $\frac{3}{36}$ since there are 36 exhaustive cases. This is represented as

$$P(X = x_i) = p_i = f(x_i) = P(X = 5) = f(5) = \frac{3}{36}.$$

Calculating in a similar way the other probabilities, the distribution of probabilities of this discrete R.V. is denoted by the discrete probability distribution as follows:

$X = x_i$	1	2	3	4	5	6
$\left.\begin{array}{l} P(X = x_i) \\ = f(x_i) \\ = p_i \end{array}\right\}$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

Discrete probability distribution, probability function or probability mass function of a discrete R.V. X is the function $f(x)$ satisfying the following conditions:

i. $f(x) \geq 0$

ii. $\sum_x f(x) = 1$

iii. $P(X = x) = f(x)$.

Thus probability distribution is the set of ordered pairs $(x, f(x))$, i.e., outcome x and its probability (chance) $f(x)$.
Cumulative distribution or simply *distribution* of a discrete R.V. X is $F(x)$ defined by

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \quad \text{for } -\infty < x < \infty.$$

It follows that

$$F(-\infty) = 0, \quad F(+\infty) = 1,$$

$$p(x_j) = P(X = x_j) = F(x_j) - F(x_{j-1}).$$

Continuous Probability Distributions

For a continuous R.V. X, the function $f(x)$ satisfying the following, is known as the probability density function (P.D.F.) or simply density function (Fig. 27.2).

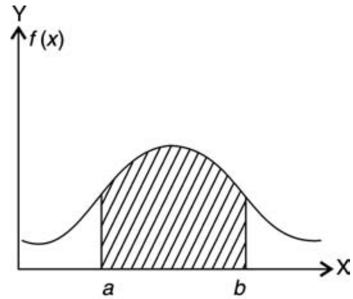


Fig. 27.2

i. $f(x) \geq 0$

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

iii. $P(a < X < b) = \int_a^b f(x)dx =$ area under $f(x)$ between ordinates $x = a$ and $x = b$.

Note 1: $P(a < X < b) = P(a \leq X < b)$
$= P(a < X \leq b) = P(a \leq X \leq b)$

i.e., inclusion or non-inclusion of end points, does not change the probability, which is *not* the case in the discrete distributions.

Note 2: Probability at a point,
$P(X = a) = \int_{a-\Delta x}^{a+\Delta x} f(x)dx$.

Cumulative Distribution

For a continuous R.V. X, with P.D.F. $f(x)$, the cumulative distribution $F(x)$ is defined as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \ -\infty < x < \infty$$

It follows that

$$F(-\infty) = 0, \quad F(+\infty) = 1, \quad 0 \leq F(x) \leq 1$$

for $-\infty < x < \infty$.

$$f(x) = \frac{dF(x)}{dx} = F'(x) \geq 0 \text{ and}$$

$$P(a < X < b) = F(b) - F(a).$$

Expectation

The behaviour of a R.V. (either discrete or continuous) is completely characterized by the distribution function $F(x)$ or density $f(x)$ [$P(x_i)$ in discrete case]. Instead of a function, a more compact description can be made by a single numbers such as mean (expectation), median and mode known as measures of central tendency of the R.V. X.

Expectation or mean or expected value

Expectation or mean or expected value of a random variable X, denoted by $E(X)$ or μ, is defined as

$$E(X) = \begin{cases} \sum_i x_i f(x_i), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f(x)dx, & \text{if } X \text{ is continuous} \end{cases}$$

Note 1: x is median if $P(X < x) \le \frac{1}{2}$ and $P(X > x) \le \frac{1}{2}$.

Note 2: x is mode for which $f(x)$ or $P(x_i)$ attains its maximum.

Variance

Variance characterizes the variability in the distributions, since two distributions with *same* mean can still have different dispersion of data about their means.

Variance of R.V. X is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (X-\mu)^2 f(x), \text{ for } X \text{ discrete}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx,$$

for X continuous.

Standard deviation (S.D.) denoted by σ, is the positive square root of variance.

Result: $\sigma^2 = E(X^2) - \mu^2$

$$\text{Since } \sigma^2 = \sum_x (x-\mu)^2 f(x) = \sum_x (x^2 - 2\mu x + \mu^2) f(x)$$

$$= \sum x^2 f(x) - 2\mu \sum x f(x) + \mu^2 \sum f(x)$$

$$= E(X^2) - 2\mu\cdot\mu + \mu^2\cdot 1 = E(X^2) - \mu^2$$

Since $\mu = \sum x f(x)$, $\sum f(x) = 1$.

Similar result follows for continuous R.V. X, with $\sum$ replaced by integration from $-\infty$ to ∞.

Note 1: In a gambling game, expected value E of the game is considered to be the value of the game to the player. Game is favourable to the player if $E > 0$, unfavourable if $E < 0$, fair if $E = 0$.

Note 2: Mathematical expectation $E = a_1p_1 + a_2p_2 + \cdots + a_kp_k$ where the probabilities of obtaining the amounts $a_1, a_2, \ldots$ or a_k are $p_1, p_2, \ldots p_k$ respectively.

27.2 CHEBYSHEV'S THEOREM

Theorem: *Let μ and σ be the mean and standard deviation of a random variable X with probability density $f(X)$. Then the probability that X will assume a value within k standard deviations of the mean is at least $1 - \frac{1}{k^2}$, for any positive constant k. Symbolically*

$$P(\mu - k\sigma < X < \mu + k\sigma) = P(|X-\mu| < k\sigma) \ge 1 - \frac{1}{k^2}$$

Proof: By definition

$$\sigma^2 = \text{variance} = E\Big[(X-\mu)^2\Big] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

$$= \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx$$

Since the second integral on the R.H.S. is non-negative, we get an inequality of the form

$$\sigma^2 \ge \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx$$

For the first integral $x \le \mu - k\sigma$ and for the second integral $x \ge \mu + k\sigma$. In either case, we have

$$x - \mu \le -k\sigma \quad \text{or} \quad x - \mu \ge k\sigma \quad \text{i.e.,} \quad |x-\mu| \ge k\sigma$$
$$\text{or} \quad (x-\mu)^2 \ge k^2\sigma^2.$$

Replacing $(x-\mu)^2$ by $k^2\sigma^2$ in the two integrals,

$$\sigma^2 \ge \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)dx$$

Rewriting

$$P(|X-\mu| \geq k^2\sigma^2) = \int_{-\infty}^{\mu-k\sigma} f(x)dx + + \int_{\mu+k\sigma}^{\infty} f(x)dx \leq \frac{1}{k^2}$$

By complimentation rule,

$$\int_{\mu-k\sigma}^{\mu+k\sigma} f(x)dx = 1 - \left[\int_{-\infty}^{\mu-k\sigma} f(x)dx + \int_{\mu+k\sigma}^{\infty} f(x)dx\right]$$

Hence $P(|X-\mu| < k\sigma) = P(\mu - k\sigma < X < \mu + k\sigma)$

$$= \int_{\mu-k\sigma}^{\mu+k\sigma} f(x)dx \geq 1 - \frac{1}{k^2}$$

Note 1: Put $k\sigma = C > 0$ then

$$P\left\{|X-\mu| \geq C\right\} \leq \frac{\sigma^2}{C^2}$$

i.e., $$P\left\{|X-E(X)| \geq C\right\} \leq \frac{\text{var }(X)}{C^2}$$

or $$P\left\{|X-\mu| < C\right\} \geq 1 - \frac{\sigma^2}{C^2}$$

i.e., $$P\left\{|X-E(X)| < C\right\} \geq \frac{1-\text{var }(X)}{C^2}$$

Note 2: Chebyshev's theorem (1853) is "distribution-free" since it is applicable to any unknown distribution and gives only the lower bound for the probability.

WORKED OUT EXAMPLES

Chebyshev's theorem

Example 1: Determine the smallest value of k in the Chebyshev's theorem for which the probability is (a) at least 0.95 (b) at least 0.99.

Solution: From Chebyshev's theorem, we have the probability as

$$P\Big(|X-\mu| < k\sigma\Big) \geq 1 - \frac{1}{k^2}$$

a. Probability $= 0.95 \geq 1 - \frac{1}{k^2}$ or $k^2 \leq \frac{1}{0.05}$
$\therefore k = \sqrt{20} = 4.472$

b. Probability $= 0.99 \geq 1 - \frac{1}{k^2}$ or $k^2 \leq \frac{1}{0.01}$
$\therefore k = 10$

Example 2: Find the probability, using Chebyshev's theorem, that the number of driving licences X issued by Road Transport Authority (R.T.A.) in a specific month is between 64 and 184 if the number of driving licences issued X is a random variable with $\mu = 124$ and $\sigma = 7.5$.

Solution: $X = \mu \pm k\sigma$,

For 64, $64 = 124 - k(7.5) \quad \therefore \quad k = 8$

For 184, $184 = 124 - k(7.5) \quad \therefore \quad k = 8$

$$P(|X-\mu| < \sigma) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{8^2} = 0.984375$$

Example 3: Suppose the amount of thiamine in a slice of 'modern' bread is a random variable X with $\mu = 0.260$ mg and $\sigma = 0.005$ mg. Using Chebyshev's theorem, between what values must be the thiamine content of (a) at least $\frac{35}{36}$ of all slices of 'modern' bread (b) at least $\frac{143}{144}$ of all slices of the 'modern' bread, lies.

Solution: By Chebyshev's theorem

$$P\Big(|X-\mu| < k\sigma\Big) \geq 1 - \frac{1}{k^2}$$

a. Given $\frac{35}{36} = 0.972 \geq 1 - \frac{1}{k^2} \quad \therefore \quad k \leq 6$,
$\mu = 0.26, \sigma = 0.005$

For $\mu - k\sigma = 0.260 - 6(0.005) = 0.230$

For $\mu + k\sigma = 0.260 + 6(0.005) = 0.290$

Thus at least $\frac{35}{36}$ of all slices of bread will contain thiamine between 0.230 and 0.290.

b. Given $\frac{143}{144} = 0.993 \geq 1 - \frac{1}{k^2} \quad \therefore k \leq 12$

For $\mu - k\sigma = 0.260 - 12(0.005) = 0.200$

For $\mu + k\sigma = 0.260 + 12(0.005) = 0.320$

Thus at least $\frac{143}{144}$ of all slices of bread will contain thiamine between 0.200 and 0.320.

EXERCISE

Chebyshev's theorem

1. Suppose X is a random variable such that $E(X) = 3$ and $E(X^2) = 13$. Calculate a lower bound for the probability that X lies between -2 and 8 using Chebyshev's theorem.

   Hint: $E(X) = \mu = 3, \sigma^2 = E(X^2) - \left\{E(X)\right\}^2 = 13 - 9 = 4, \sigma = 2.$

$P\{3-2k < X < 3+2k\} \geq 1 - \frac{1}{k^2}$, choose $k = \frac{5}{2}$.

Ans. $P(-2 < X < 8) \geq 1 - \frac{1}{\left(\frac{5}{2}\right)^2} = \frac{21}{25} = 0.84$

2. The number of patients requiring I.C.U. in a hospital in a random variable with $\mu = 18$ and $\sigma = 2.5$. Determine the probability that there will be between 8 and 28 patients.

Ans. $k = \frac{28-18}{2.5} = \frac{18-8}{2.5} = 4$, the probability is at least $1 - \frac{1}{4^2} = \frac{15}{16}$

3. Let X be a random variable with an unknown probability distribution, with mean $\mu = 8$ and variance $\sigma^2 = 9$. Determine
(i) $P(|X-8| \geq 6)$ (ii) $P(-4 < X < 20)$

Ans. **i.** $P\Big(|X-8| \geq 6\Big) = 1 - P\Big(|X-8| < 6\Big)$
$= 1 - P(-6 < (X-8) < 6)$
$= 1 - P\Big[8-(2)(3) < X < 8+(2)(3)\Big] \leq \frac{1}{4}$

ii. $P(-4 < X < 20) =$
$P\Big[8-(4)(3) < X < 8+(4)(3)\Big] \geq \frac{15}{16}$.

4. Let X be the discrete random variable denoting the number appearing in a single throw of a fair die. Let $E(X) = \mu$

Using Chebyshev's theorem prove that

$$P\Big[|X-\mu| > 2.5\Big] < 0.47$$

while the actual probability is zero.

Ans. $E(X) = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2}$, $E(X^2) = \frac{91}{6}$

$$\text{var}(X) = E(X^2) - \Big[E(X)\Big]^2 = \frac{91}{6} - \frac{49}{4} = 2.9167$$

Choose $k = 2.5$

$$P\Big\{|X-\mu| > 2.5\Big\} < \frac{2.9167}{6.25} = 0.47$$

Actual probability $= p = P\{|X-3.5| > 2.5\}$

$= P\,\{X$ lies outside the limits $(3.5-2.5, 3.5+2.5)\}$

$= P\Big\{X$ lies outside $(1, 6)\Big\} = 0$

since impossible event.

5. Apply Chebyshev's theorem to calculate
(i) $P(5 < X < 15)$ (ii) $P(|X-10| \geq 3)$
(iii) $P(|X-10| < 3)$

for a random variable X with mean $\mu = 10$ and variance $\sigma^2 = 4$.

Hint:

i. $\mu - k\sigma = 10 - k2 = 5$
$\therefore k = \frac{5}{2}$, $\mu + k\sigma = 10 + k2 = 15$
$P(5 < X < 15) = P(10-2k < X < 10+2k) \geq 1 - \frac{1}{k^2} = 1 - \frac{4}{25} = \frac{21}{25}$

ii. $|X-10| \geq 3$ or $-3 < X-10 < 3$ or $10-3 < x < 10+3$
$7 < x < 13 : \mu + k\sigma = 10 + 2k = 13$ and $10 - 2k = 7$
Choose $k = \frac{3}{2}$
$P(|X-10| \geq 3) \leq \frac{1}{k^2} = \frac{4}{9}$

iii. $P(|X-10| < 3) \leq 1 - \frac{4}{9} = \frac{5}{9}$

Ans. (i) at least $\frac{21}{25}$ (ii) at most $\frac{4}{9}$ (iii) at least $\frac{5}{9}$

Theoretical probability distributions

Generally, frequency distributions are formed from the observed or experimental data. However, frequency distributions of certain populations can be deduced mathematically by fitting a theoretical probability distributions under certain assumptions.

Example: The shoes-industry should know the 'sizes' of foot of the population, the food industry the 'tastes' (menu) of the population, etc.

Three such important theoretical probability distributions in order of their discovery are:

i. Binomial (due to James Bernoulli, 1700).
ii. Normal (due to De-Moivre 1733) also credited to Laplace (1774), Gauss (1809).
iii. Poisson (due to S.D. Poisson 1837).

Discrete probability distributions: Binomial, Poisson, geometric, negative binomial, hypergeometric, multinomial, multivariate hypergeometric distributions.

Continuous probability distributions: Uniform (rectangular), normal, Gamma, exponential, χ^2, Beta, bivariate normal, 't', 'F', distributions.

Worked Out Examples

Discrete probability distributions

Example 1: Prove that (a) $E(kX) = kE(X)$ (b) $E(X+k) = E(X)+k$ (c) $E(X+Y) = E(X)+E(Y)$.

Solution:

a. $E(kX) = \frac{\sum k f_i x_i}{\sum f_i} = k\frac{\sum f_i x_i}{\sum f_i} = kE(X)$

b. $E(X+k) = \frac{\sum f_i(x_i+k)}{\sum f_i} = \frac{\sum f_i x_i}{\sum f_i} + k\frac{\sum f_i}{\sum f_i}$
$= E(X) + k$

c. $E(X+Y) = \frac{\sum f_i(x_i+y_i)}{\sum f_i} = \frac{\sum f_i x_i}{\sum f_i} + \frac{\sum f_i y_i}{\sum f_i}$
$= E(X) + E(Y)$

Note 1: Above results can be proved for continuous case by 'replacing' $\sum$ by $\int_{-\infty}^{\infty}$.

Note 2: Above results are rewritten in 'μ' notation as (a)$\mu_{kX} = k\mu_X$ (b) $\mu_{X+k} = \mu_X + k$ (c)$\mu_{X+y} = \mu_X + \mu_Y$.

Example 2: Prove that (a) Var $(X+k) =$ Var (X) (b) Var $(kX) = k^2$ Var (X). Hence $\sigma_{X+k} = \sigma_X$ and $\sigma_{kX} = |k|\sigma_X$.

Solution:

a. Var $(X+k) = \sum(x_i+k)^2 f(x_i) - \mu_{X+k}^2$ by using the result Var $(X) = E(X^2) - \mu_X^2$

$$= \sum(x_i^2 + k^2 + 2kx_i)f(x_i) - (\mu_X + k)^2$$

$$= \sum x_i^2 f_i + k^2\sum f_i + 2k\sum x_i f_i - (\mu_X^2 + k^2 + 2\mu_X k)$$

$$= \sum x_i^2 f_i + k^2 + 2k\mu_X - \mu_X^2 - 2k\mu_X - k^2$$

$$= (\text{Var}\ (X) + \mu_X^2) - \mu_X^2 = \text{Var}\ (X)$$

b. Var $(kX) = \sum(kx_i)^2 f_i - \mu_{kx}^2$

$$= k^2\sum x_i^2 f_i - (k\mu_X)^2 = k^2\left(\sum x_i^2 f_i - \mu_X^2\right)$$

$$= k^2\text{Var}(X).$$

Example 3: Determine the discrete probability distribution, expectation, variance, S.D. of a discrete random variable (D.R.V.) X which denotes the minimum of the two numbers that appear when a pair of fair dice is thrown once.

Solution: The total number of cases are $6 \times 6 = 36$. The minimum number could be 1, 2, 3, 4, 5, 6, i.e., $X(s) = X(a, b) = \min\{a, b\}$. The number 6 will appear only in one case (6, 6), so

$$f(6) = P(X = 6) = P(\{(6, 6)\}) = \frac{1}{36}.$$

For minimum 5, favourable cases are (5, 5), (5, 6), (6, 5) so

$$f(5) = P(X = 5) = \frac{3}{36}.$$

For minimum 4, favourable cases are (4, 4), (4, 5), (4, 6), (5, 4) so

$$f(4) = P(X = 4) = \frac{5}{36}.$$

For minimum 3: (3, 3), (3, 4), (3, 5), (3, 6), (6, 3), (5, 3), (4, 3) so

$$f(3) = P(X = 3) = \frac{7}{36}.$$

For minimum 2: (2, 2), (3, 3), (2, 4), (2, 5), (2, 6), (6, 2), (5, 2), (4, 2), (3, 2) so

$$f(2) = P(X = 2) = \frac{9}{36}.$$

Similarly,

$$f(1) = P(X = 1) = \frac{11}{36}.$$

Thus the required discrete probability distribution

$X = x_i$	1	2	3	4	5	6
$P(X = x_i)$ $= f(x_i)$ $= f_i$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

Mean = Expectation = $E(X) = \sum x_i f_i$

$$E(X) = 1\cdot\frac{11}{36} + 2\cdot\frac{9}{36} + 3.\frac{7}{36} + 4\cdot\frac{5}{36} + 5\cdot\frac{3}{36} + 6\cdot\frac{1}{36} = 2.5$$

$$\text{Var}\ (X) = \sum x^2 f_i - \mu^2$$

$$= 1\cdot\frac{11}{36} + 4\cdot\frac{9}{36} + 9\cdot\frac{7}{36} + 16\cdot\frac{5}{36} + 25\cdot\frac{3}{36} + 36\cdot\frac{1}{36} - (2.5)^2$$

$\sigma^2 = 1.9745$, so $\sigma =$ S.D. $= 1.4$.

Example 4: A player tosses 3 fair coins. He wins Rs. 500 if 3 heads occur, Rs. 300 if 2 heads occur, Rs. 100 if one head occurs. On the other hand, he loses Rs. 1500 if 3 tails occur. Find the value of the game to the player. Is it favourable?

Solution: Let $X =$ D.R.V. = number of heads occurring in 3 tosses of a fair coin. The sample space S is

$$S = \{H, T\} \times \{H, T\} \times \{H, T\}$$
$$= \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Probability of all 3 heads $= P(X = 3) = \frac{1}{8}$
Probability of all 3 tails $= P(X = 0) = \frac{1}{8}$
Probability of 2 heads $= P(X = 2) = \frac{3}{8}$,
$P(X = 1) = \frac{3}{8}$
Discrete probability distribution is

$X = x_i$	0	1	2	3
$P(X = x_i)$ $= f(x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Expected value of the game

$$= 500 \cdot \frac{1}{8} + 300 \cdot \frac{3}{8} + 100 \cdot \frac{3}{8} - 1500 \cdot \frac{1}{8}$$
$$= \frac{200}{8} = 25 \text{ rupees.}$$

Game is favourable to the player since $E > 0$.

Continuous probability distributions

Example 5: Suppose a continuous R.V. x has the probability density

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

(a) Find k (b) Find $P(0.1 < x < 0.2)$ $(c) P(x > 0.5)$ Using distribution function, determine the probabilities that (d) x is less than 0.3 (e) between 0.4 and 0.6 (f) Calculate mean and variance for the probability density function.

Solution:

a. Since $\int_{-\infty}^{\infty} f(x)dx = 1$ so

$$\int_{-\infty}^{\infty} f(x) = \int_0^1 k(1 - x^2)dx = k(x - \tfrac{x^3}{3})\Big|_0^1$$
$$= \tfrac{2}{3}k = 1 \quad \therefore k = \tfrac{3}{2}.$$

b. $P(0.1 < x < 0.2) = \int_{0.1}^{0.2} k(1 - x^2)dx$
$$= \tfrac{3}{2}\left(x - \tfrac{x^3}{3}\right)\Big|_{0.1}^{0.2} = 0.1465.$$

c. $P(x > 0.5) = \int_{0.5}^{\infty} f(x)dx = \int_{0.5}^{1} f(x)dx$
$$= \tfrac{3}{2}\left(x - \tfrac{x^3}{3}\right)\Big|_{0.5}^{1} = 0.3125.$$

d. Distribution function:

$F(x) = \int_{-\infty}^{x} f(t)dt$ so

$$F(x) = \int_0^x \tfrac{3}{2}(1 - x^2)dx = \tfrac{3}{2}\left(x - \tfrac{x^3}{3}\right)$$

$$F(x < 0.3) = \int_{-\infty}^{0.3} f(t)dt = \tfrac{3}{2}\left(x - \tfrac{x^3}{3}\right)\Big|_0^{0.3}$$
$$= 0.4365$$

e. $F(0.4 < x < 0.6) = F(b) - F(a)$
$= F(0.6) - F(0.4)$
$$= \tfrac{3}{2}\left(x - \tfrac{x^3}{3}\right)\Big|_{0.4}^{0.6} = 0.224.$$

f. Mean $=\mu= \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 x\{\tfrac{3}{2}(1 - x^2)\}dx$
$$= \tfrac{3}{2}\left(\tfrac{x^2}{2} - \tfrac{x^4}{4}\right)\Big|_0^1 = \tfrac{3}{8}$$

Variance $= \int_{-\infty}^{\infty}(x - \mu)^2 f(x)dx$
$$= \int_0^1 \left(x - \tfrac{3}{8}\right)^2 \tfrac{3}{2}(1 - x)^2dx = \tfrac{19}{320}$$

or variance $= \int_0^1 x^2\{k(1 - x^2)\}dx - \mu^2$
$$= k\left(\tfrac{x^3}{3} - \tfrac{x^5}{5}\right)\Big|_0^1 - \mu^2 = \tfrac{19}{320}.$$

Example 6: The daily consumption of electric power (in millions of kW-hours) is a R.V. having the

$$\text{P.D.F.} \qquad f(x) = \begin{cases} \tfrac{1}{9}xe^{-x/3}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

If the total production is 12 million kW-hours, determine the probability that there is power cut (shortage) on any given day.

Solution: Probability that the power consumed is between 0 to 12 is

$$P(0 \le x \le 12) = \int_0^{12} f(x)dx = \int_0^{12} \frac{1}{9} x e^{-x/3} dx$$

$$= -\frac{x}{3} e^{-x/3} - e^{-x/3} \Big|_0^{12} = 1 - 5e^{-4}$$

Power supply is inadequate if daily consumption exceeds 12 million kW, i.e.,

$$P(x > 12) = 1 - P(0 \le x \le 12) = 1 - [1 - 5e^{-4}] = 5e^{-4}$$

$$= 0.0915781$$

Example 7: (a) Find the mean and variance of a uniform (rectangular) distribution (b) Determine its cumulative distribution function (Fig. 27.3).

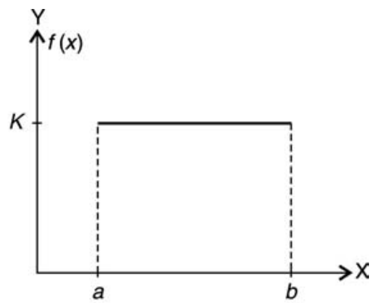


Fig. 27.3

Solution: The uniform distribution is defined by $f(x) = k = \text{constant in } (a, b)$
$= 0$ elsewhere

Since $$1 = \int_{-\infty}^{\infty} f(x)dx = \int_a^b k dx = k(b-a),$$

so $$k = \frac{1}{b-a}$$

Thus $$f(x) = \begin{cases} \frac{1}{b-a}, & \text{for } a < x < b \\ 0, & \text{elsewhere} \end{cases}$$

$$\text{Mean} = \mu = \int_a^b x f(x) dx = \int_a^b x \cdot \frac{dx}{b-a}$$

$$= \frac{1}{b-a} \cdot \frac{x^2}{2} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)}$$

$$\mu = \frac{b+a}{2}$$

$$\text{Variance} = \int_a^b x^2 f(x) dx - \mu^2 = \frac{1}{b-a} \cdot \int_a^b x^2 dx - \mu^2$$

$$\sigma^2 = \frac{1}{b-a} \frac{x^3}{3} \Big|_a^b - \mu^2 = \frac{b^3 - a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2$$

$$= \frac{(b-a)^2}{12}$$

Cumulative distribution $F(x)$:

i. When $x \le a$, $F(x) = \int_{-\infty}^{x} = \int 0 = 0$
so $F(x) = 0$

ii. When $a < x < b$,
$F(x) = \int_{-\infty}^{x} = \int_{-\infty}^{a} 0 \cdot dx + \int_a^x \frac{1}{b-a} dx$
$F(x) = \frac{x-a}{b-a}$

iii. When $x \ge b$, $F(x) = \int_{-\infty}^{x} = \int_{-\infty}^{a} 0$
$+ \int_a^b \frac{1}{b-a} dx + \int_b^x 0 = \frac{b-a}{b-a} = 1.$

Thus

$$f(x) = \begin{cases} 0, & \text{when } x \le a \\ \frac{x-a}{a-b}, & \text{when } a < x < b \\ 1, & \text{when } x \ge b. \end{cases}$$

EXERCISE

Probability distributions

1. Calculate μ, σ^2, σ for

a.

x_i	2	3	8
f_i	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

b.

x_i	−1	0	1	2	3
f_i	.3	.1	.1	.3	.2

Ans. **a.** $\mu = 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{2} + 8 \cdot \frac{1}{4} = 4,$
$\sigma^2 = (2-4)^2 \frac{1}{4} + (3-4)^2 \frac{1}{2} + (8-4)^2 \frac{1}{4}$
$= 5.5$
$\sigma = 2.3$ [or $\sigma^2 = 4 \cdot \frac{1}{4} + 9 \cdot \frac{1}{2} + 64 \cdot \frac{1}{4} - 4^2$
$= 5.5$]

b. $\mu = -1 \cdot (.3) + 0 \cdot (.1) + 1 \cdot (0.1)$
$+ 2 \cdot (0.3) + 3 \cdot (0.2) = 1.0$
$\sigma^2 = 4(.3) + 0(.1) + 1(.1) + 4(.3) + 9(.2)$
$= 2.4, \sigma = 1.5$

2. Determine the expected number of families to have (a) 2 boys and 2 girls (b) at least one boy (c) no girls (d) at most two girls, out of 800 families with 4 children each. Assume equal probabilities for boys and girls.

Ans. (a) 37.5% (b) 93.75% (c) 6.25% (d) 68.75%

Hint: X: No boys in a family. P.D.F.

X	0	1	2	3	4
$p(x_i)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Percentage of families $= \frac{6}{16} \times 100 = 37.5\%$ for (a) etc.

3. A box contains 8 items of which 2 are defective. A person draws 3 items from the box. Determine the expected number of defective items he has drawn.

Ans. $0 \cdot \frac{20}{56} + 1 \cdot \frac{30}{56} + 2 \cdot \frac{6}{56} = \frac{3}{4}$

Hint:

X	0	1	2
$f(x_i)$	$\frac{6c_3 2c_0}{8c_3}$	$\frac{6c_2 2c_1}{8c_3}$	$\frac{6c_1 2c_2}{8c_3}$.

Here X is the number of defectives.

4. A stake of Rs. 44 is to be won between 2 players A and B, whoever gets 6 in a throw of die alternatively. Determine their respective expectations if A starts the game.

Ans. $E(A) = \frac{6}{11} \times 44 =$ Rs. 24,

$E(B) = \frac{5}{11} \times 44 =$ Rs. 20

Hint: $P(A \text{ wins}) = \frac{1}{6} + \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} + \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} + \cdots = \frac{1}{6} \frac{1}{1-(\frac{5}{6})^2} = \frac{6}{11}$

$p =$ probability of $6 = \frac{1}{6}$, prob no $6 = \frac{5}{6}$.

5. A person wins Rs. 80 if 3 heads occur, Rs. 30 if 2 heads occur, Rs. 10 if only one head occurs in a single toss of 3 fair coins. If the game is to be fair, how much should he lose if no heads occur?

Ans. Rs. 200

Hint: $0 =$ Expectation $= 80 \cdot \frac{1}{8} + 30 \cdot \frac{3}{8} + 10 \cdot \frac{3}{8} - x \cdot \frac{1}{8}$

6. Find the mean and variance of P.D.F.

$$f(x) = \begin{cases} \frac{1}{4} e^{-x/4} & \text{for } x > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

Ans. $\mu = \int_0^\infty \frac{1}{4} e^{-x/4} dx = \frac{e^{-x/4}}{-1/4} = 4 \cdot 1 = 4$

$\sigma^2 = \int_0^\infty (x-4)^2 \frac{1}{4} e^{-x/4} dx = \int_0^\infty x^2 \frac{1}{4} e^{-x/4} dx - \mu^2 = 96 - 16 = 80$

7. If P.D.F. $f(x) = k(x+3)$ in (2, 8), determine (a) $P(3 < x < 5)$ (b) $P(x \geq 4)$ (c) $P(|x-5| < 0.5)$

Ans. $k = \frac{1}{48}$, (a) $\frac{7}{24}$ (b) $\frac{3}{4}$

(c) $P(4.5 < x < 5.5) = \frac{1}{6}$.

8. Find the mean and variance of the "exponential" distribution $f(x) = \frac{1}{b} e^{-x/b}$ for $x > 0, b > 0$.

Ans. $\mu = b, \sigma^2 = b^2$

Hint: $\mu = \int_0^\infty x \cdot \frac{1}{b} e^{-x/b} dx = b\Gamma(2) = b$,

$\sigma^2 = \int_n^\infty x^2 \frac{1}{b} e^{-x/b} dx - \mu^2 = b^2 \Gamma(3) - b^2$

$= 2b^2 - b^2 = b^2$

27.3 DISCRETE UNIFORM DISTRIBUTION

Discrete uniform distribution is the simplest of all discrete probability distribution. The discrete random variable assumes each of its values with the same (equal) probability. In this equiprobable or uniform space each sample point is assigned equal probabilities.

Example 1: In the tossing of a fair die, each sample point in the sample space $\{1, 2, 3, 4, 5, 6\}$ is assigned with the same (uniform) probability $\frac{1}{6}$ i.e. $p(x) = \frac{1}{6}$ for $x = 1, 2, 3, 4, 5, 6$. In particular, if the sample space S contains k points, then probability of each point is $\frac{1}{k}$. Thus in the discrete uniform distribution, the discrete random variable X assigns equal probabilities to all possible values of X. Therefore the probability mass function $f(X)$ has the form

$$f(X) = \frac{1}{k} \text{ for } X = x_1, x_2, \ldots, x_k \qquad (1)$$

or equivalently

X	x_1	x_2	x_3	$\ldots$	x_k
$f(X)$	$\frac{1}{k}$	$\frac{1}{k}$	$\frac{1}{k}$	$\ldots$	$\frac{1}{k}$

Here the constant K which is the *parameter* completely determines the discrete uniform distribution (1). The mean and variance of (1) are given by

$$\mu = E(X) = \sum_{i=1}^{k} x_i f(x_i) = \sum x_i \cdot \frac{1}{k} = \frac{\sum_{i=1}^{k} x_i}{k}$$

and

$$\sigma^2 = E((X-\mu)^2) = \sum_{i=1}^{k} (x_i-\mu)^2 f(x_i)$$
$$= \sum_{i=1}^{k} \frac{(x_i-\mu)^2}{k}$$
$$\sigma^2 = \frac{1}{k}\sum_{i=1}^{k} ()x_i-\mu)^2$$

The discrete uniform distribution is of particular importance in lotteries.

Worked Out Examples

Example 2: If a ticket is drawn from a box containing 10 tickets numbered 1 to 10 inclusive. Find the probability that the number x drawn is (a) less that 4 (b) even number (c) prime number (d) find the mean and variance of the random variable X.

Solution: (a) since each ticket has the same probability for being drawn, the probability distribution is discrete uniform distribution given by $f(x) = \frac{1}{10}$ for $x = 1, 2, 3, \ldots, 10$. Now

$$P(x<4) = \sum_{x=0}^{3} P(x) = \sum_{x=0}^{3} \frac{1}{10} = \frac{1}{10}(1+1+1) = \frac{3}{10}$$

(b) 2, 4, 6, 8, 10 are even numbers each with probability $\frac{1}{10}$ probability of even number $= \frac{5}{10} = \frac{1}{2}$ (c) 2, 3, 5, 7 are prime prob. of prime $= \frac{4}{10} = \frac{2}{5}$. (d) Mean $= E(x) = \sum_{x=1}^{10} xP(x) = \sum_{x=1}^{10} x \cdot \frac{1}{10} = \frac{1}{10}(1+2+3+\cdots+10)$

Mean $= \mu = \frac{1}{10} \cdot \frac{10(10+1)}{10\ \ 2} = \frac{11}{2} = 5 \cdot 5$

Variance $= \sigma^2 = \sum_{x=1} (x-\mu)^2 p(x)$

$$= \frac{1}{10}[(1-5\cdot5)^2 + (2-5\cdot5)^2 + (3-5\cdot5)^2 + \ldots + (10-5\cdot5)^2]$$
$$\sigma^2 = 8.25$$

(or $\sigma^2 = E(x^2) - \{E(x)\}^2 = \sum x^2 \cdot P(x) - (5\cdot5)^2$

$$= \frac{1}{10} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{1}{10}\frac{10(10+1)(20+1)}{6} - (5\cdot5)^2$$

$38.5 - 30.25 = 8.25$)

Exercise

1. Deteremine the probability that an odd number appears in the toss of a fair die.

Ans. $\frac{3}{6}$

2. Find the probability that at least one head appears in the throw of three fair coins.

Ans. $\frac{7}{8}$

3. If a card is selected at random from an ordinary pack of 52 cards, find the probability that (a) card is a spade (b) card is a face card i.e., Jack, queen or king (c) card is a spade face card.

Ans. $\frac{13}{52}$ (b) $\frac{12}{52}$ (c) $\frac{3}{52}$

Hint: (a) 13 spade cards available. So $\binom{13}{1}\Big/\binom{52}{1}$

(b) 12 face cards so $\binom{12}{1}\Big/\binom{52}{1}$

(c) 3 spade face cards so $\binom{3}{1}\Big/\binom{52}{1}$.

Note: Each card has the same probability $\frac{1}{52}$.

4. Find the probability that a card drawn at random from 50 cards numbered 1 to 50 is (a) prime, (b) ends in the digit 2, (c) divisible by 5.

Ans. (a) $\frac{3}{10}$ (b) $\frac{1}{10}$ (c) $\frac{1}{5}$

5. Two marbles are drawn from a box containing 4 red and 8 black marbles. Find the probability that (a) both are red (b) both are black (c) at least one is red.

Ans. (a) $\binom{4}{2}\Big/\binom{12}{2} = \frac{1}{11}$,

(b) $\binom{8}{2}\Big/\binom{12}{2} = \frac{14}{33}$ (c) $1 - \frac{14}{33} = \frac{19}{33}$

6. If two cards are drawn from an ordinary pack of 52 cards, determine the probability that (a) both are spades (b) one is a spade and one is a heart.

Ans. (a) $\binom{13}{2}\Big/\binom{15}{2} = \frac{78}{1326} = \frac{1}{17}$

(b) $(13.13) \Big/ \binom{15}{2} = \frac{13}{102}$

27.4 BINOMIAL DISTRIBUTION

Binomial distribution (B.D.) due to James Bernoulli (1700) is a discrete probability distribution. The Bernoulli process has the following properties:

- **i.** An experiment is repeated n number of times, called n trials where n is a fixed integer.
- **ii.** The outcome of each trial is classified into two mutually exclusive (dichotomus) categories arbitrarily called a "success" and a "failure".
- **iii.** Probability of success, denoted by p, remains constant for all trials.
- **iv.** The outcomes are independent (of the outcomes of the previous trials).

Each trial in the Bernoulli process is known as Bernoulli trial.

The binomial random variable X is the *number* of successes in n Bernoulli trials. X is discrete since X takes only integer values (we 'count' the number of successes).

Binomial distribution is thus the probability distribution of this discrete random variable X, and is given by

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, 2, \ldots, n \qquad (1)$$

where n is the number of trials and p is the probability of success in any trial. The probability of x successes is p^x and remaining failures is q^{n-x}. This can happen in n_{C_x} ways. By multiplication rule, the probability of x successes in n trials is $\binom{n}{x} p^x q^{n-x}$.

Note that the $(n+1)$ terms of the binomial expansion

$$(q+p)^n = \binom{n}{0} q^n + \binom{n}{1} pq^{n-1} + \binom{n}{2} p^2 q^{n-2} + \cdots + \binom{n}{n} p^n$$

$$= b(0; n, p) + b(1; n, p) + b(2; n, p) + \cdots + b(n; n, p)$$

$$= \sum_{x=0}^{n} b(x; n, p)$$

correspond to various values of $b(x; n, p)$ for $x = 0, 1, 2, \ldots, n$.

Since $p + q = 1$, it follows that

$$\sum_{x=0}^{n} b(x; n, p) = 1.$$

B.D. is characterized by the parameter p and the number of trials n.

The mean μ of B.D. is np and the variance σ^2 of B.D. is npq. (see Worked Out Example 1 on page 27.11)

The binomial sums

$$B(r; n, p) = \sum_{x=0}^{r} b(x; n, p)$$

are tabulated (see A2 to A7)

since

$$\frac{n_{C_{x+1}}}{n_{C_x}} = \frac{n!}{(x+1)!(n-x-1)!} \frac{x!(n-x)!}{n!} = \left(\frac{n-x}{x+1}\right)$$

The recurrence relation for B.D. is

$$b(x+1; n, p) = \left(\frac{n-x}{x+1}\right)\left(\frac{p}{q}\right) b(x; n, p).$$

WORKED OUT EXAMPLES

Binomial distribution

Example 1: Find the (a) mean and (b) variance of B.D.

Solution:

a. mean $= \mu =$ expectation $= \sum_{x=0}^{n} x P(x)$

$$= \sum_{x=0}^{n} x b(x; n, p) = \sum_{x=0}^{n} x \frac{x!}{x!(n-x)!} p^x q^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(n-x)!(x-1)!} p^{x-1} q^{n-x}, \text{ put } x - 1 = x^*$$

$$= np \sum_{x^*=0}^{n-1} \frac{(n-1)!}{(n-x^*-1)x^*!} p^{x^*} q^{n-x^*-1}$$

$$= np(p+q)^{n-1} = np \quad \text{since } p + q = 1.$$

b. Variance $\sigma^2 = \sum_{x=0}^{n} (x-\mu)^2 p(x)$

$$= \sum_{x=0}^{n} (x^2 - 2\mu x + \mu^2) p(x)$$

$$= \sum_{x=0}^{n} x^2 p(x) - 2\mu \sum_{x=0}^{n} x p(x) + \mu^2 \sum_{x=0}^{n} p(x)$$

$$= \sum_{x=0}^{n} x^2 p(x) - 2\mu \cdot np + \mu^2 \cdot 1$$

$$\because \sum x p(x) = \mu, \quad \sum p(x) = 1$$

$$= \sum_{x=0}^{n} x^2 p(x) - n^2 p^2 \quad \text{since } \mu = np$$

Consider

$$\sum_{x=0}^{n} x^2 p(x) = \sum_{x=1}^{n} x^2 \frac{n!}{(n-x)!x!} p^x q^{n-x}$$

$$= \sum_{x=1}^{n} [x(x-1) + x] \frac{n!}{(n-x)!x!} p^x q^{n-x}$$

$$= \sum_{x=2}^{n} \frac{x(x-1)n!}{(n-x)!x!} p^x q^{n-x} + \sum_{x=1}^{N} x \frac{n!}{(n-x)!x!} p^x q^{n-x}$$

$$= \sum_{x=2}^{n} \frac{n!}{(n-x)!(x-2)!} p^x q^{n-x} + np$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(n-x)!(x-2)!} p^{x-2} q^{n-x} + np$$

$$= n(n-1)p^2 \cdot (q+p)^{n-2} + np$$

$$= n(n-1)p^2 + np \quad \text{since } p+q=1$$

Thus

$$\sigma^2 = n(n-1)p^2 + np - n^2p^2 = np - np^2$$

$$\sigma^2 = np(1-p) = npq.$$

Example 2: Determine the probability of getting 9 exactly twice in 3 throws with a pair of fair dice.

Solution: In a single throw of a pair of fair dice, 9 can occur in 4 ways: (6, 3), (3, 6), (5, 4), (4, 5) out of $6 \times 6 = 36$ ways. Thus
p = probability of occurrence of 9 in one throw= $\frac{4}{36} = \frac{1}{9}$.
n = number of trials = 3.
Probability of getting 9 exactly twice in 3 throws

$$= b\left(2; 3, \frac{1}{9}\right) = 3_{C_2} \left(\frac{1}{9}\right)^2 \left(\frac{8}{9}\right)^{3-2} = 3 \cdot \frac{1}{9} \cdot \frac{1}{9} \cdot \frac{8}{9} = \frac{8}{243}.$$

Example 3: Out of 800 families with 5 children each, how many would you expect to have (a) 3 boys (b) 5 girls (c) either 2 or 3 boys. Assume equal probabilities for boys and girls.

Solution: Probability of boy $= P(B) = p = \frac{1}{2}$, and probability of girl $= P(G) = q = \frac{1}{2}$.
n = number of trials = 5, X = no. of boys in a family

a. Probability of a family having 3 boys

$$= P(X=3) = 5_{C_3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{5-2}$$

$$= \frac{5!}{2!3!} \left(\frac{1}{2}\right)^5 = \frac{10}{32} = \frac{5}{16}$$

Expected number of families having 3 boys out of 5 children = $800\left(\frac{5}{16}\right) = 250$, i.e., 250 families have 3 boys out of 5 children.

b. $P(X=0) = P$ (all girls) $= \left(\frac{1}{2}\right)^5 = \frac{1}{32}$

Expectation $= 800 \times \frac{1}{32} = 25$.

c. $P(X=2) = P(2 \text{ boys}) = 5_{C_2} \left(\frac{1}{2}\right)^5 = \frac{5}{8}$

Expectation $= 800 \times \frac{5}{8} = 500$.

Example 4: Determine the probability distribution of the number of bad eggs in a box of 6 chosen at random if 10% of eggs are bad, in a large consignment.

Solution: Probability of a bad egg $= p = \frac{10}{100} = 0.1$. Let X = number of bad eggs, $n = 6$. The required B.D. $= b(x; 6, 0.1) = 6_{C_x} (.1)^x (.9)^{6-x}$, for $x = 0, 1, 2, 3, 4, 5, 6$.

X:	0	1	2	3	4	5	6
$P(X)$:	.5311	.35429	0.098	0.015	0.001215	0.000054	0

Example 5: Assume that 50% of all engineering students are good in mathematics. Determine the probabilities that among 18 engineering students (a)

exactly 10 (b) at least 10 (c) at most 8 (d) at least 2 and at most 9, are good in maths.

Solution: Let X = number of engineering students who are good in maths:
p = prob of good in maths $= \frac{50}{100} = \frac{1}{2}$, $n = 18$
$b(x; n, p) = b(x; 18, \frac{1}{2}) = 18_{C_x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{18-x}$

a. Exactly 10 students out of 18 are good in maths

$$P(X = 10) = 18_{C_{10}} \left(\frac{1}{2}\right)^{10} \left(\frac{1}{2}\right)^{8} = .1670$$

From tables (A2 to A7)

$$P(X = 10) = \sum_{x=0}^{10} b\left(x; 18, \frac{1}{2}\right) - \sum_{x=0}^{9} b\left(x; 18, \frac{1}{2}\right)$$

$$= .7597 - .5927 = .1670$$

b. $$P(X \geq 10) = \sum_{x=10}^{18} 18_{C_x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{18-x}$$

$$= \sum_{x=0}^{18} b\left(x; 18, \frac{1}{2}\right) - \sum_{x=0}^{9} b\left(x; 18, \frac{1}{2}\right)$$
$$= 1 - .5927 = .4073$$

c. $$P(X \leq 8) = \sum_{x=0}^{8} 18_{C_x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{18-x} = .4073$$

from table with $n = 18$, $x = 8$, $p = \frac{1}{2}$.

d. $$P(2 \leq x \leq 9) = \sum_{x=2}^{9} 18_{C_x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{18-x}$$

$$= \sum_{x=0}^{9} b\left(x; 18, \frac{1}{2}\right) - \sum_{x=0}^{1} b\left(x; 18, \frac{1}{2}\right).$$
$$= .5927 - .0007 = .5920.$$

Example 6: The probability of a man hitting a target is $\frac{1}{3}$. (a) If he fires 5 times, what is the probability of his hitting the target at least twice ? (b) How many times must he fire so that the probability of his hitting the target at least once is more than 90%?

Solution: Probability of hitting $= p = \frac{1}{3}$

probability of no hit (or failure) $= q = \frac{2}{3}$

a. X = number of hits (successes), $n = 5$

$$P(X \geq 2) = \sum_{x=2}^{5} 5_{C_x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{5-x}$$

$$= 1 - \sum_{x=0}^{1} 5_{C_x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{5-x}$$

$$= 1 - \left(\frac{2}{3}\right)^5 - 5_{C_1} \left(\frac{1}{3}\right) \left(\frac{2}{3}\right)^4 = \frac{131}{243}.$$

b. The probability of not hitting the target is q^n in n trials (fires). Thus to find the smallest n for which the probability of hitting at least once $1 - q^n$ is more than 90%.

i.e., $$1 - q^n > 0.9$$

or $$1 - \left(\frac{2}{3}\right)^n > 0.9 \quad \text{i.e.,} \quad \left(\frac{2}{3}\right)^n < 0.1$$

For $n = 6$, $2^6 = 64 < (0.1)3^6 = 72.9$ this is true. In other words, he must fire 6 times.

Example 7: If X be a binomially distributed random variable with $E(X) = 2$ and Var $(X) = \frac{4}{3}$, find the distribution of X.

Solution: We know that $E(X)$ = mean $= np = 2$ and Var (X) = Variance $= npq = \frac{4}{3}$. or $\frac{npq}{np} = \frac{4}{3}\frac{1}{2} = \frac{2}{3}$ or $q = \frac{2}{3}$ so $p = \frac{1}{3}$. Thus $n\frac{1}{3} = 2$ or $n = 6$.
Hence the B.D. is $b(x; 6, \frac{1}{3})$

x_i	0	1	2	3	4	5	6
$f(x_i)$	$\frac{64}{729}$	$\frac{192}{729}$	$\frac{240}{729}$	$\frac{160}{729}$	$\frac{60}{729}$	$\frac{12}{729}$	$\frac{1}{729}$

Example 8: Fit a binomial distribution to the following data:

X	0	1	2	3	4
f	30	62	46	10	2

Solution: Here n = no. of trials = 4 and
N = total frequency $= \sum_{i=0}^{4} f_i = 30 + 62 + 46 + 10 + 2 = 150.$

Mean of the binomial distribution is

$$\mu = np = \frac{\sum f_i x_i}{\sum f_i} = \frac{0(30)+1(62)+2(46)+3(10)+4(12)}{150}$$

$$np = 4p = \frac{192}{150} \quad \therefore p = \frac{192}{600} = 0.32$$

Thus the binomial distribution that fits the given data is $b(x; 4, 0.32) = 4_{C_x}(.32)^x(.68)^{4-x} = p(x)$

x :	0	1	2	3	4
$P(x)$:	.2138	0.4	0.2866	0.0866	0.0133
Expected frequency = $N \times P(x)$ = $(150)P(x)$	32	60	43	13	2

Hint: Use the recurrence relation

$$b(x+1; n, p) = \left(\frac{n-x}{x+1}\right)\frac{p}{q} b(x; n, p)$$

EXERCISE

Binomial distribution

1. A fair coin is tossed 6 times. Find the probability of getting (a) exactly 2 heads (b) at least four heads (c) no heads (d) at least one head.

Ans. (a) $\frac{15}{64}$ (b) $\frac{11}{32}$ (c) $\frac{1}{64}$ (d) $\frac{63}{64}$

Hint:
a. $b(2; 6, \frac{1}{2}) = 6_{C_2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^4$
b. $\sum_{x=4}^{6} b\left(x; 6, \frac{1}{2}\right) = \frac{15}{64} + \frac{6}{64} + \frac{1}{64}$
c. $1 - q^6 = 1 - \left(\frac{1}{2}\right)^6$

2. A fair die is tossed 7 times. Determine the probability that a 5 or a 6 appears (a) exactly 3 times (b) never occurs.

Ans. (a) $\frac{560}{2187}$ (b) $\frac{2059}{2187}$

Hint: (a) $b\left(3; 7, \frac{1}{3}\right)$ (b) $1 - q^7 = 1 - \left(\frac{2}{3}\right)^7$

3. Team A has probability $\frac{2}{3}$ of winning whenever it plays. If A plays 4 games, find the probability that A wins (i) exactly 2 games (ii) at least 1 game (iii) more than half of the games.

Ans. (i) $P(2) = b\left(2; 4, \frac{2}{3}\right) = \frac{8}{27}$ (ii) $1 - q^4 = 1 - \left(\frac{1}{3}\right)^4 = \frac{80}{81}$ (iii) $P(3)+P(4) = \frac{32}{81} + \frac{16}{81} = \frac{16}{27}$

4. How many dice must be thrown so that there is a better than even chance of obtaining a six?

Ans. 4 dice

Hint: Find n such that $\left(\frac{5}{6}\right)^n < \frac{1}{2}$.

5. A man hits a target with probability $\frac{1}{4}$.
(i) Determine the probability of hitting at least twice when he fires 7 times (ii) How many times must he fire so that the probability of his hitting the target at least once is greater than $\frac{2}{3}$?

Ans. **i.** $1 - P(0) - P(1) = 1 - \frac{2187}{16384} - \frac{5103}{16384} = \frac{4547}{8192}$
ii. $n = 4$

Hint: Find n such that $1 - q^n > \frac{2}{3}$.

6. The probability that a pen manufactured by a company will be defective is 0.1. If 12 such pens are examined, find the probability that (a) exactly two (b) at least two (c) none, will be defective.

Ans. (a) 0.2301 (b) 0.3412 (d) 0.2833

7. 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?

Ans. 323

Hint: Mean $= 2 = np = 20p$, $p = 0.1$,

$$P(X \geq 3) = 1 - \sum_{x=0}^{2} b(x; 20, 0.1) = 0.323$$

Expected number $= 1000 \times 0.323 = 323$.

8. The probability that a patient recovers from a disease is 0.4. If 15 persons have such a disease, determine the probability that (a) exactly 5 survive (b) at least 10 survive (c) from 3 to 8 survive.

Ans. **a.** $P(X=5) = b(5; 15, 0.4) = \sum_{x=0}^{5} b(x, 15, 0.4) - \sum_{x=0}^{4} b(x; 15, 0.4) = 0.4032 - 0.2173$

$= 0.1859.$

b. $P(X \geq 10) = 1 - P(X < 10)$

$= 1 - \sum_{x=0}^{9} b(x; 15, 0.4) = 1 - 0.9662$

$= 0.0338.$

c. $P(3 \leq X \leq 8) = \sum_{x=3}^{\infty} = \sum_{x=0}^{\infty} - \sum_{x=0}^{2}$

$= 0.9050 - 0.0271 = 0.8779.$

9. A manufacturer of fax machine claims that only 10% of his machines require repairs within one year. If 5 of 20 of his machines required repaires within 1 year, does this tend to support or refute the claim?

Ans. Reject (refute) the claim since probability is very small.

Hint: $\sum_{x=5}^{20} b(x; 20, 0.10) = 1 - \sum_{x=0}^{4} b(x; 20, 0.10)$
$= 1 - 0.9568 = 0.0432.$

10. Two dice are thrown 120 times. Find the average number of times in which the number on first dice exceeds the number on the second dice.

Ans. $E(X) = np = 120\left(\frac{5}{12}\right) = 50$

Hint: Successful are (2, 1), (3, 1), (3, 2), (4, 1), (4, 2), (4, 3), (5, 1), (5, 2), (5, 3), (5, 4), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), $p = \frac{15}{6.6} = \frac{5}{12}$.

11. A communication system consists of n components, each of which will independently function with probability p. The total system will be able to operate effectively if at least one half of its components function. For what values of p is a 5-component system more likely to operate effectively than a 3-component system?

Ans. $p \geq \frac{1}{2}$

Hint: $P(X \geq 3) = \sum_{x=3}^{5} b(x; 5, p) \geq P(X \geq 2)$

$= \sum_{x=2}^{3} b(x; 3, p).$

Fitting of binomial distribution

Fit a B.D. to the following data:

12.

x:	0	1	2	3	4	5	6
f:	5	18	28	12	7	6	4

Ans. 4 15 25 22 11 3 0

Hint: $n = 6$, $p = 0.4$, $N = 80$

13.

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

Ans. 6.9 19.1 24 17.8 8.6 2.9 .7 .1 0 0 0

Hint: $n = 10$, $N = 80$, $p = 0.2175$

14.

x:	0	1	2	3	4	5
f:	38	144	342	287	164	25

Ans. 33.2 161.9 316.2 308.7 150.7 29.4

Hint: $n = 5$, $p = 0.494$

15. Seven coins are tossed and number of heads noted. The experiment is repeated 128 times with the following data:

No. of heads	0	1	2	3	4	5	6	7
Frequencies	7	6	19	35	30	23	7	1

Fit a binomial distribution assuming
i. coin is unbiased
ii. nature of coin is not known.

Ans. **i.** 1, 7, 21, 35, 35, 21, 7, 1

Hint: $p = \frac{1}{2}$, $N = 128$.

ii. 1, 8, 23, 36, 34, 19, 6, 1

Hint: $n = 7$, $N = 128$,.
$np = \frac{433}{128} = 3.3828$, $p = 0.48326$

27.5 HYPERGEOMETRIC DISTRIBUTION

The binomial distribution quite frequently arises from a random experiment in which sampling is done *with* replacement. In contrast, the hypergeometric distribution arises from random experiments in which sampling is done *without* replacement. It is very useful in quality control and analysis of the opinion surveys. Consider a population of N units in which each unit is classified into two dischotomus classes (arbitrarily known as "success" and "failure") according to whether the unit does or does not posses a certain property under consideration. Let k be the number of successes and $N - k$ be the failures in the population.

Draw a random sample of size n *without* replacement from the population. Let X be the discrete random variable which denotes the number of successes in the sample. Then the probability distribution of x is known as hypergeometric distribution and is given by $h(x; N, n, k) = \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$,

$x = 0, 1, 2, \ldots, n$ (1)

The integer x should lie in the interval $\max\ (0, n - N + k) \le x \le \min\ (n, k)$.

Here $\binom{N}{n}$ is the number of ways of choosing a sample of size n from a population N, $\binom{k}{x}$ is the number of ways in which x successes is chosen from a total of k successes and finally $\binom{N-k}{n-x}$ gives the number of ways of getting $(n - x)$ failures out of the (remaining) $N - k$ failures. The hypergeometric distribution (1) has N, n, k as the three parameters.

Book work I Prove that mean of hypergeometric distribution is $\dfrac{nk}{N}$.

Proof: Mean $= E(x) = \sum_{x=0}^{n} x \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} =$

$$k\sum_{x=1}^{n} \frac{(k-1)!}{(x-1)!(k-x)!} \frac{\binom{N-k}{n-x}}{\binom{N}{n}} = k\sum_{x=1}^{n} \frac{\binom{k-1}{x-1}\binom{N-k}{n-x}}{\binom{N}{n}}$$

Now $\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N}{n}\binom{N-1}{n-1}$ so

$$E(x) = \frac{nk}{N} \sum_{y=0}^{n-1} \frac{\binom{k-1}{y}\binom{N-1-(k-1)}{n-1-y}}{\binom{N-1}{n-1}}$$

$$= \frac{nk}{N} \cdot 1 = \frac{nk}{N}$$

Here $y = x - 1$ and the summation represents the sum of all probabilities in hypergeometric experiment $(n - 1)$ sample is drawn from a population of $(N - 1)$ containing $(k - 1)$ successes.

Book work II Prove that the variance of hypergeometric distribution is $\dfrac{nk(N-k)(N-n)}{N^2(N-1)}$.

Proof: Var $(X) = E(X^2) - \{E(X)\}^2$

Now $E(X^2) = E\{X(X-1) + X\} = E\{X(X-1)\} + E(X)$

Consider

$$E(X(X-1)\} = \sum_{x=0}^{n} x(x-1)\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= \sum\nolimits_{x=0}^{n} x(x-1) \cdot \frac{k!}{x!(k-x)!} \frac{\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= \frac{k(k-1)}{\binom{N}{n}} \sum_{x=2}^{n} \frac{(k-2)!}{(x-2)!(k-2-(x-2))!} \times$$

$$\times \binom{N-2-(k-2)}{n-2-(x-2)}$$

Now

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N(N-1)\cdot(N-2)!}{n(n-1)\cdot(n-2)!(N-n)!}$$

So,

$$E\{X(X-1)\} = \frac{k(k-1)n(n-1)}{N(N-1)} \times$$

$$\times \sum_{y=0}^{n-2} \frac{\binom{k-2}{y}\binom{N-2-(k-2)}{n-2-y}}{\binom{N-2}{n-2}}$$

$$= \frac{k(k-1)n(n-1)}{N(N-1)} \cdot 1$$

Here $y = x - 2$ and the summation is one because it is the sum of the probabilities for $y = 0$ to $n - 2$. Thus

$$\text{var}(X) = E(X^2) - \{E(X)\}^2 = E\{X(X-1)\} + E(X) - \{E(X)\}^2$$

$$= \frac{k(k-1)n(n-1)}{N(N-1)} + \frac{nk}{N} - \frac{n^2k^2}{N^2} =$$

$$= \frac{nk}{N^2(N-1)}\{N(N-1)+N(k-1)(n-1)-(N-1)nk\}$$

$$= \frac{nk}{N^2(n-1)}\{N^2 - Nn - Nk + nk\}.$$

$$\sigma^2 = \text{var}(X) = \frac{nk(N-k)(N-n)}{N^2(N-1)}$$

$$= \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right)$$

Introducing $p = \frac{k}{N}$ which is the proportion successes in the population, the mean and variance of the hypergeometric distribution are written as

$$\mu = n \cdot \frac{k}{N} = np \qquad \text{and}$$

$$\sigma^2 = \left(\frac{N-n}{N-1}\right) n \cdot p(1-p) = \left(\frac{N-n}{N-1}\right) npq.$$

Observe that the mean of the hypergeometric distribution and the mean of binomial distribution are same, while the variances differ by the factor $\left(\dfrac{N-n}{N-1}\right)$, known as "*finite population correction factor*", which tends 1 as $N \to \infty$. Thus binomial distribution may be viewed as a large population edition of the hypergeometric distribution, since sampling from the finite population with replacement amounts to sampling from the infinite population (without replacement).

Approximation of the Hypergeometric Distribution by the Binomial Distribution

Book work III Show that hypergeometric distribution tends to binomial distribution as $N \to \infty$ and $\frac{k}{N} \to p$.

Proof: Consider

$$h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= \frac{k!}{x!(k-x)!} \frac{(N-k)!}{(n-x)!(N-k-n+x)!} \frac{n!(N-n)!}{N!}$$

$$= \frac{k(k-1)(k-2)\ldots(k-(x-1))}{x!} \times$$

$$\times \cdot \frac{(N-k)(N-k-1)\ldots(N-k-(n-k-1))}{(n-k-1)!} \times$$

$$\times \frac{n!}{N(N-1)(N-2)\ldots(N-(n-1))}$$

$$= \frac{n!}{x!(n-x)!} \times$$

$$\left[\left(\frac{k}{N}\right)\left(\frac{k}{N} - \frac{1}{N}\right)\left(\frac{k}{N} - \frac{2}{N}\right)\cdots\left(\frac{k}{N} - \frac{(x-1)}{N}\right)\right]$$

$$\times \frac{\left(1 - \frac{k}{N}\right)\left(1 - \frac{k}{N} - \frac{1}{N}\right)\ldots\left(1 - \frac{(n-k-1)}{N}\right)}{\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)\ldots\left(1 - \frac{n-1}{N}\right)}$$

Letting $N \to \infty$ and putting $\frac{k}{N} = p$ we get

$$\lim_{N\to\infty} h(x; N, n, k) = \binom{n}{x} p \cdot (p-0)(p-0)\ldots(p-0) \times (1-p)(1-p-0)\ldots(1-p)$$

$$= \binom{n}{x} p^k (1-p)^k = b(x; p, 1-p)$$

WORKED OUT EXAMPLES

Example 1: Out of 60 applicants to a university 40 are from the south. If 20 applicants are selected at random, find the probability that (a) 10 (b) not more than 2, are from south.

Solution: The total number of ways in which 20 applicants are selected from 60 is $\binom{60}{20}$.

(a) The number of ways in which 10 applicants from south are selected from 40 south applicants is $\binom{40}{10}$.

Now out of the total 20 selected the remaining 10 non-south applicants will be selected from 20(60–40 south) non-south applicants, which is $\binom{20}{10}$. Thus the probability that 10 out of 20 applicants are from south is $\dfrac{\binom{40}{10}\binom{20}{10}}{\binom{60}{20}} = \dfrac{\left(\frac{40!}{10!30!}\right)\left(\frac{20!}{10!10!}\right)}{\left(\frac{60!}{20!40!}\right)} =$ 0.0373613

or population size $N = 60$,
sample size $= n = 20$
number of south applicants $k = 40$. Let x denote the number of applicants from the south. Then the hypergeometric distribution is $h(x; N, n, k) = \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$ for $x = 0, 1, 2, \ldots n$.

Here $$h(10; 60, 20, 40) = \frac{\binom{40}{10}\binom{60-40}{20-10}}{\binom{60}{20}} =$$

0.03736

(b) Probability that $x \leq 2$ is

$$P(x \leq 2) = \sum_{x=0}^{2} h(x; N, n, k) = \sum_{x=0}^{2} h(x; 60, 20, 40)$$

$$= h(0; 60,20,40) + h(1; 60,20,40) + h(2; 60,20,40)$$

$$= \frac{\binom{40}{0}\binom{60-40}{20-0}}{\binom{60}{20}} + \frac{\binom{40}{1}\binom{60-40}{20-1}}{\binom{60}{20}} + \frac{\binom{40}{2}\binom{60-40}{20-2}}{\binom{60}{20}}$$

$$= \frac{(40c_0)(20c20) + (40c1)(20c19) + (40c2)(20c18)}{60c20}$$

Example 2: Solve (a) of the above problem using Binomial approximation. Explain the accuracy obtained. What is the finite population correlation factor.

Solution: As $N \to \infty$, Binomial approximation is $P(X = x) = \binom{n}{x} p^x q^{n-x}$.

Here the probability of (south) success is $p = \frac{40}{60} = \frac{2}{3}$ and probability of (non-south) failure is $q = \frac{20}{60} = \frac{1}{3}$ and $n = 20$ and $x = 10$.

Now using binomial approximation we have $P(X = x)$ = probability of 10 south applicants out of 20 applicants $= \binom{20}{10}\left(\frac{2}{3}\right)^{10}\left(\frac{1}{3}\right)^{10}$

$$= 0.0542591$$

From hypergeometric distribution the probability is 0.03736. The accuracy is less since $N = 60$ is not very large more so compared with $n = 20$. Finite population correction factor $= \dfrac{N-n}{N-1} = \dfrac{60-20}{60-1} =$ 0.6779

Example 3: Find the mean and variance and standard deviation of the random variable X in (a) of above example.

Solution: Mean $= \dfrac{nk}{N} = \dfrac{20(40)}{(60)} = 1.3333$

Variance $= \dfrac{N-n}{N-1} \cdot n \cdot \dfrac{k}{N} \cdot \left(1 - \dfrac{k}{N}\right)$

$$= \left(\frac{60-20}{60-1}\right) \cdot 20\frac{(40)}{(60)} \cdot \left(1 - \frac{40}{60}\right) = 0.30131$$

Standard deviation = 0.5489

EXERCISE

1. Find the probability of selecting 5 cards of which 3 are red and 2 are black from an ordinary deck of 52 playing cards.

Ans. $\binom{26}{3}\binom{26}{2}\Big/\binom{52}{5} = 0.3251$

2. Determine the probability that exactly one defective is found in a sample of 5 from a lot of 40 components containing 3 defectives (in the entire lot).

Ans. $\binom{3}{1}\binom{37}{4}\Big/\binom{40}{5} = 0.3011$

3. Find the mean and variance of the random variable X in the above example 2.

Ans. Mean = $\frac{5 \cdot 3}{40} = 0.375$, $\sigma^2 = 0.3113$

4. (a) Determine the probability that exactly 3 computers are defective out of 10 computers purchased from a lot of 5000 computers containing 1000 defective computers. (b) Use binomial approximation and explain the accuracy. (c) what is the finite population correction factor.

Ans. (a) $h(3; 5000, 10, 1000) = 0.2015$
(b) $h(3; 5000, 10, 1000) \simeq b(3; 10, 0.2)$
$= \sum_{x=0}^{3} b(x; 10, 0.2) - \sum_{x=0}^{2} b(x; 10, 0.2) = 0.8791 - 0.6778 = 0.2015$
since $N = 5000$ large compared to sample $n = 10$, the accuracy achieved is high.
(c) $\dfrac{N-n}{N-1} = \dfrac{5000-10}{5000-1} = 0.9982$

5. TV's are shipped in lots of 50. A shipment is accepted if a sample of 5 TV's inspected from this lot, does not contain any defective TV. If one or more are found defective, the entire shipment is rejected . Suppose the lot of 50 contains 3 defective TV's. Determine the probability that 100% inspection is required?

Ans. $P(X \geq 1) = 1 - P(X = 0)$
$= 1 - \left\{\binom{3}{0}\binom{47}{5} \Big/ \binom{50}{5}\right\} = 0.28$ where X is the number of defective TV's.

6. Suppose a shipment of 100 cars contain 25 defectives. Determine the probability that 2 cars out of a sample of 10 cars drawn from the lot are defective. Use binomial approximation also.

Ans. $h(2; 10, 25, 100) = 0.292$
$b(2; 10, 0.25) = 0.282$

7. Find the probability of getting 3 blue marbles if 5 marbles are drawn (one after the other without replacement) from a box containing 6 blue and 4 red marbles.

Ans. $\binom{6}{3}\binom{4}{2} \Big/ \binom{10}{5} = \dfrac{10}{21} = 0.4762$

8. A committee of 2 is chosen from five faculty members out of which 3 are doctorate and 2 are post-graduates. If X denotes the number of doctorates in the committee, obtain the probability distribution of X.

Ans.

X	0	1	2
$P(x)$	$\frac{2}{20}$	$\frac{12}{20}$	$\frac{6}{20}$

Hint: $h(x; 5, 2, 3)$ for $x = 0, 1, 2$.

9. Let X be the number of defective motors in a sample of 6 drawn from a lot of 12 motors containing four defectives. Compute (a) $P(X = 1)$ (b) $P(X \geq 4)$ (c) $P(1 \leq X \leq 3)$.

Ans. (a) 0.242 (b) 0.030 (c) 0.939

10. A box contains 12 red and 8 black marbles. If 5 marbles are drawn successively (without replacement) from the box, find the probability that (a) 3 are red and 2 are black (b) at least 3 are red (c) all the 5 are of the same colour.

Ans. (a) $\dfrac{385}{969}$ (b) $\dfrac{682}{969}$ (c) $\dfrac{53}{969}$

Hint: (a) $\binom{12}{3}\binom{8}{2} \Big/ \binom{20}{35}$
(b) $\sum_{x=3}^{5} \binom{12}{x}\binom{8}{5-x} \Big/ \binom{20}{5}$
(c) $\left\{\binom{12}{5}\binom{8}{0} + \binom{12}{0}\binom{8}{5}\right\} \Big/ \binom{20}{5}$

27.6 POISSON DISTRIBUTION

Poisson* distribution is the discrete probability distribution of a discrete random variable x, which has no upper bound. It is defined for non-negative values of x as follows:

$$f(x, \lambda) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{for } x = 0, 1, 2, \ldots \quad (1)$$

Here $\lambda > 0$ is called the parameter of the distribution. Note that in binomial distribution the number of successes (occurrence of an event) out of a total definite number of n trials is determined, whereas in Poisson distribution the number of successes at a random point of time and space is determined.

* Simeon Denis Poisson (1781–1840) French mathematician.

Poisson distribution (P.D.) is suitable for 'rare' events for which the probability of occurrence p is very small and the number of trials n is very large. Also binomial distribution can be approximated by poisson distribution when $n \to \infty$ and $p \to 0$ such that $\lambda = np =$ constant.

Examples of rare events:

i. Number of printing mistakes per page.
ii. Number of accidents on a highway.
iii. Number of defectives in a production centre.
iv. Number of telephone calls during a particular (odd) time.
v. Number of bad (dishonoured) cheques at a bank.

Result 1: Since $\sum_{x=0}^{\infty} f(x, \lambda) = \sum_{x=0}^{\infty} p(X = x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1$, Therefore (1) is a probability function.

Result 2: Arithmetic mean of Poisson distribution

$$\overline{X} = E(X) = \sum_{x=0}^{\infty} x P(X = x) = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!}$$

$$= \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda$$

Thus the parameter λ is the A.M. of P.D.

Result 3: Variance of Poisson distribution

$$= E\left[(x - \overline{X})^2\right] = \sum_{x=0}^{\infty} (x - \overline{X})^2 P(X = x)$$

$$= \sum \left(x^2 + \overline{X}^2 - 2\overline{X}x\right) P = \sum x^2 P + \overline{X}^2 \sum P$$

$$-2\overline{X} \sum x P$$

$$= \sum x^2 P + \overline{X}^2 - 2\overline{X}\,\overline{X} = \sum x^2 P + \lambda^2 - 2\lambda^2$$

$$= \sum x^2 P - \lambda^2.$$

But

$$\sum x^2 P = \sum_{x=0}^{\infty} [x(x-1) + x]\, e^{-\lambda} \frac{\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x \cdot x(x-1)}{x!} + e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x \cdot x}{x!}$$

$$= \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} = \lambda^2 + \lambda$$

Thus

$$\text{Variance} = \sum x^2 p - \lambda^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda$$

Hence the variance of P.D. = mean of P.D.

Result 4: Recurrence formula

$$\frac{P(x+1)}{P(x)} = \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!} \frac{x!}{e^{-\lambda}\lambda^x} = \frac{\lambda}{x+1}$$

Thus

$$P(x+1) = \left(\frac{\lambda}{x+1}\right) P(x).$$

Result 5: Poisson distribution function

$$F(x; \lambda) = \sum_{k=0}^{x} \frac{e^{-\lambda}\lambda^k}{k!}$$

has been tabulated (see A8 to A11)

Then $\quad f(x; \lambda) = F(x; \lambda) - F(x-1; \lambda).$

Theorem: *Prove that Poisson distribution is the limiting case of binomial distribution for very large trials with very small probability, i.e.,* $f(x; \lambda) = \lim_{\substack{n \to \infty \\ p \to 0}} b(x; n, p)$ *such that* $\lambda = np =$ *constant.*

Proof: Put $p = \frac{\lambda}{n}$ in binomial distribution

$$b(x; n, p) = \frac{n!}{x!(n-x)!} \cdot \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n(n-1)(n-2)\cdots(n-(x-1))}{x!} \cdot \frac{\lambda^x}{n^x} \times$$

$$\times \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n^x \cdot 1 \cdot \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{(x-1)}{n}\right)}{x!} \frac{\lambda^x}{n^x} \times$$

$$\times \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Taking the limit as $n \to \infty$, we have

$$\lim_{n\to\infty} b(x; n, p) = \frac{1}{x!}\lambda^x \cdot e^{-\lambda} = \frac{\lambda^x e^{-\lambda}}{x!}$$

since

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right) = 1\cdot 1\cdots 1 = 1$$

and

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-x} = \left[\left(1-\frac{\lambda}{n}\right)^{n/\lambda}\right]^{\lambda}\times\left[1-\frac{\lambda}{n}\right]^{-x} = e^{-\lambda}.$$

Note: $\lim_{n\to\infty}\left(1-\frac{x}{n}\right)^n = e^{-x}$

Thus binomial probabilities for large n and small p are often approximated by means of poisson distribution with mean $\lambda = np$.

Example: For $n = 3000$, $p = 0.005$, the probability of 18 successes by binomial distribution is given by $b(18; 3000, .005) = 3000_{C_{18}}(.005)^{18}(.995)^{2982}$ which involves prohibitive amount of work. Instead using Poisson distribution as an approximation, we get $\lambda = 3000 \times .005 = 15$. Probability of 18 success $= f(18, 15) = 0.8195$ from table (A8 to A11).

General rule: Poisson approximation to B.D. is used whenever $n \geq 20$ and $p \leq 0.05$. For $n \geq 100$, approximation is excellent provided $\lambda = np \leq 10$.

WORKED OUT EXAMPLES

Poisson distribution

Example 1: A distributor of bean seeds determines from extensive tests that 5% of large batch of seeds will not germinate. He sells the seeds in packets of 200 and guarantees 90% germination. Determine the probability that a particular packet will violate the guarantee.

Solution: The probability of a seed not germinating $= p = \frac{5}{100} = 0.05$

$\lambda =$ mean number of seeds, in a sample of 200, which do not germinate

$= np = 200 \times 0.05 = 10$

Let $X =$ R.V. $=$ number of seeds that do not germinate

A packet will violate guarantee if it contains more than 20 non-germinating seeds.
Probability that the guarantee is violated

$$= P(X > 20) = 1 - P(X \leq 20) = 1 - \sum_{x=0}^{20}\frac{e^{-10}10^x}{x!}$$

$$= 1 - F(20, 10) = 1 - .9984 = 0.0016$$

where cummulative distribution function F is read for x=20 and λ=10 from the tables (A8 to A11).

Example 2: The average number of phone calls/minute coming into a switch board between 2 and 4 PM is 2.5. Determine the probability that during one particular minute there will be (a) 0 (b) 1 (c) 2 (d) 3 (e) 4 or fewer (f) more than 6 (g) at most 5 (h) at least 20 calls.

Solution: $\lambda = 2.5$, $f(x;\lambda) = f(x; 2.5) = \frac{(2.5)^x(e^{-2.5})}{x!}$
Let X = R.V. = number of phone calls/minute during that (odd) 2 and 4 PM.

a. $f(0; 2.5) = e^{-2.5} = .08208$

b. $f(1; 2.5) = .2052$

c. $f(2; 2.5) = .2565$

d. $f(3; 2.5) = .2138$

e. $P(X \leq 4) = \sum_{x=0}^{4} f(x; 2.5) = F(4; 2.5) = .8912$

(read from tables A8 to A11)

f. $P(X > 6) = 1 - P(X \leq 6) = 1 - \sum_{x=0}^{6} f(x; 2.5)$

$= 1 - F(6; 2.5) = 1 - .9858 = 0.0142$

g. $P(X \leq 5) = \sum_{x=0}^{5} f(x; 2.5) = F(5; 2.5) = .9580$

h. $P(X \geq 2.0) = 1 - P(X \leq 19) = 1 - \sum_{x=0}^{19} f(x; 2.5)$

$= 1 - F(19; 2.5) = 1 - 1 = 0.$

Example 3: Suppose that on the average one person in 1000 makes a numerical error in preparing income tax return (ITR). If 10000 forms are selected at random and examined, find the probability that 6, 7 or 8 of the forms will be in error.

Solution: Let X = R.V. = number ITR forms containing a numerical error. Essentially this is a binomial experiment with 10000 trials and probability (of success) $p = \frac{1}{1000} = 0.001$. So by B.D. probability of 6, 7 or 8 error forms = $P(X = 6, 7 \text{ or } 8)$

$$= P(6) + P(7) + P(8) = \sum_{x=6}^{8} b(x; 10000, 0.001)$$

$$= \sum_{x=6}^{8} 10000_{C_x} (.001)^x (.999)^{10000-x}$$

which involves cumbersome lengthy calculations. Since n is large and p is small, approximate the binomial probabilities by Poisson distribution with $\lambda = np = 10000 \times \frac{1}{1000} = 10$.

Probability of 6, 7 or 8 error ITR forms = $P(X = 6, 7 \text{ or } 8)$

$$= \sum_{x=6}^{8} f(x; 10) = \sum_{x=6}^{\infty} \frac{e^{-10}(10)^x}{x!}$$

$$= e^{-10}\left[\frac{10^6}{6!} + \frac{10^7}{7!} + \frac{10^8}{8!}\right] = .2657$$

Instead, using tables A8 to A11, we get the result in a simpler way

$$= F(8; 10) - F(5; 10) = .3328 - .0671 = .2657.$$

Example 4: Fit a Poisson distribution to the following data:

X_i:	0	1	2	3	4
Observed frequencies f_i	30	62	46	10	2

Solution: To fit a Poisson distribution, determine the only parameter λ of the distribution from the given data. Since λ is the arithmetic mean,

$$\lambda = \frac{\sum_{i=0}^{4} f_i X_i}{\sum f_i X_i} = \frac{0 \times 30 + 1 \times 62 + 2 \times 46 + 3 \times 10 + 4 \times 2}{150}$$

$$= \frac{192}{150} = 1.28$$

Thus the Poisson distribution that "fits" to the given data is $P(X) = \frac{e^{-1.28}(1.28)^X}{X!}$.

Here total frequency $N = \sum_{i=0}^{4} f_i = 150$

Expected frequency = (Total frequency) × Probability

X_i:	0	1	2	3	4
$P(X_i)$:	0.27803	.35588	.22776	.09718	.031097
$(N)(P(X_i))$ =Expected frequency	41.7045 ≈ 42	53.382 ≈ 53	34.164 ≈ 34	14.577 ≈ 15	4.6646 ≈ 5

EXERCISE

Poisson distribution

1. Determine the probability that 2 of 100 books bound will be defective if it is known that 5% of books bound at this bindery are defective. (a) use B.D. (b) use Poisson approximation to B.D.

Ans. **a.** $b(2; 100, 0.5) = \binom{100}{2}(0.05)^2(.95)^{98} = 0.081$

b. $f(2; 5) = \frac{5^2 e^{-5}}{2!} = 0.084$ with $\lambda = np = 100(0.05) = 5$

2. Find the probabilities that $0, 1, 2, 3, 4, \ldots$ of 3840 generators fail if the probability of failure is $\frac{1}{1200}$.

Ans.

x:	0	1	2	3	4	5	6	7	8	9	10
$f(x, 3.2)$:	0.041	.130	.209	.223	.178	.114	.06	.028	.011	.004	.002

Hint: $\lambda = 3840 \times \frac{1}{1200} = 3.2$. Use tables (A8 to A11) and the identity $f(x; \lambda) = F(x; \lambda) - F(x - 1; \lambda)$.

3. On an average, 1.3 gamma particles/millisecond come out of a radioactive substance, determine (a) mean (b) variance

(c) probability of more than one gamma particles emanate from the substance.

Ans. (a)(b): $\lambda = \sigma^2 = 1.3$ (c) $1 - P(X = 0) = 1 - e^{-1.3} = 0.727$

4. Determine the probability p that there are 3 defective items in a sample of 100 items if 2% of items made in this factory are defective.

Ans. $p = f(3; 2) = \frac{2^3 e^{-2}}{3!} = 0.180$ with $\lambda = np = 100(0.02) = 2$

5. Suppose 300 misprints are distributed randomly throughout a book of 500 pages. Find the probability P that a given page contains (i) exactly two misprints (ii) two or more misprints.

Ans. **i.** $f(2; 0.6) = \frac{(0.6)^2 e^{-0.6}}{2!} = 0.0988 \approx 0.1$

ii. $P = 1 - P(0 \text{ or } 1 \text{ misprint})$
$= 1 - (0.549 + 0.329) = 0.122$

6. In a factory producing blades, the probability of any blade being defective is 0.002. If blades are supplied in packets of 10, determine the number of packets containing (a) no defective (b) one defective and (c) two defective blades respectively in a consignment of 10000 packets.

Ans. **a.** $10000 \times P(0) = 10000 \times e^{-0.02}$
$= 10000 \times .9802 = 9802$

i.e., 9802 packets do not have any defective blades.

b. $10000 \times (0.02)(.9802) = 196$

c. $10000 \times \frac{(.02)^2}{2!} \cdot 9802 = 2$

Hint: $\lambda = np = 10 \times 0.002 = 0.02.$

7. A manufacturer of cotter pins knows that 5% of his product is defective. Pins are sold in boxes of 100. He guarantees that not more than 10 pins will be defective. Determine the probability that a box will fail to meet the guarantee.

Ans. $P(X > 10) = 1 - P(X \leq 10)$
$= 1 - \sum_{x=0}^{10} \frac{e^{-5} 5^x}{x!} = 1 - F(10, 5)$
$= 1 - .9863 = 0.0137$

Hint: $\lambda = np = 100 \times 0.05 = 5$

8. On an average 20 red blood cells are found in a fixed volume of blood for a normal person. Determine the probability that the blood sample of a normal person will contain less than 15 red cells.

Ans. $P(X < 15) = \sum_{x=0}^{14} \frac{e^{-20}(20)^x}{x!} = F(14, 20)$
$= 0.105$

Hint: $\lambda = 20.$

9. Two shipments of computers are received. The first shipment contains 1000 computers with 10% defectives and the second shipment contains 2000 computers with 5% defectives. One shipment is selected at random. Two computers are found good. Find the probability that the two computers are drawn from the first shipment.

Ans. 0.183

Hint: $q_1 = 0.1,\ p_1 = 0.9,\ q_2 = 0.05,$
$p_2 = 0.95$

$\lambda_1 = n_1 p_1 = (1000)(0.9) = 900,$

$\lambda_2 = n_2\, p_2 = (2000)(.95) = 1900$

C: two computers good, A: first shipment,

B: second shipment.

$$P(A/C) = \frac{P(A)P(C/A)}{P(A)P(C/A) + P(B) \cdot P(C/B)}$$

where

$$P(A) = P(B) = \frac{1}{2},\ P(C/A) = \frac{e^{-900}(900)^2}{2!},$$

$$P(C/B) = \frac{e^{-1900}(1900)^2}{2!}.$$

10. Given that the probability of an accident in an industry is 0.005 and assuming the accidents are independent (a) determine the probability that in any given period of 400 days, there will be an accident one day? (b) What is the probability that there are at most three days with an accident?

Ans. (a) $P(X = 1) = e^{-2} 2^1 = 0.271$

(b) $P(X \leq 3) = \sum_{x=0}^{3} \frac{e^{-2}2^x}{x!} = 0.857$

Hint: $\lambda = np = 400(0.005) = 2$.

11. If one in every 1000 of computers produced is defective, determine the probability that a random sample of 8000 will yield fewer than 7 defective computers?

Ans. $P(X < 7) = \sum_{x=0}^{6} b(x; 8000, 0.001)$

$\simeq \sum_{x=0}^{6} f(x; 8) = 0.3134$

Hint: B.D. calculation is very hard, approximate it by P.D. with $\lambda = np = (8000)(0.001) = 8$.

12. Suppose the average number of telephone calls coming into a telephone exchange between 10 AM to 11 AM is 2, while between 11 AM to 12 noon is 6, determine the probability that more than five calls come in between 10 AM to 12 noon, assuming that calls are independent.

Ans. $P(x > 5) = 1 - P(x \leq 5) = 1 - \sum_{x=0}^{5} \frac{e^{-8}8^x}{x!} = 1 - 0.1912 = 0.8088$

Hint: P.D. is additive: $X = X_1 + X_2$, $\lambda = \lambda_1 + \lambda_2 = 2 + 6 = 8$.

Fitting of Poisson distribution

Fit a Poisson distribution to the following data:

1.

x:	0	1	2	3	4	5	6	7	8
Observed frequency f_i	56	156	132	92	37	22	4	0	1

Ans. 69.6 137.25 135.33 88.95 43.85 17.29 5.68 1.60 0.3942

Hint: $\lambda = \frac{\sum f_i x_i}{N} = \frac{986}{500} = 1.972$.

2.

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

Ans. 121 61 15 2 0

Hint:
$\lambda = \sum \frac{f_i x_i}{N} = \frac{60+36+6+4}{200} = 0.5; e^{-.5} = 0.61$

3.

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

Ans. 147.15 147.15 73.58 24.53 6.13 1.23

Hint: $\lambda = \frac{\sum f_i x_i}{\sum f_i} = \frac{400}{400} = 1$.

4. Determine the number of pages expected with 0, 1, 2, 3, and 4 errors in 1000 pages of a book if on the average two errors are found in five pages.

Ans.

x:	0	1	2	3	4
$P(x)$:	.6703	.26812	.053624	.0071	.00071
Expected number of pages	670	268	54	7	1

Hint: $\lambda = 2/5 = 0.4$, $e^{-0.4} = .6703$,

Expected number of pages $= 1000 \times P(x)$.

5.

x:	0	1	2	3	4
f:	109	65	22	3	1

Ans. 108.7 66.3 20.2 4.1 0.7

Hint: $\lambda = \frac{65+44+9+4}{200} = \frac{122}{200} = 0.61$.

27.7 POISSON PROCESS

Poisson process is a random process in which the number of events (or successes) x occurring in a time interval of length say T is counted. It is a continuous parameter, discrete state process. By dividing T into n equal parts of length Δt, we have $T = n \cdot \Delta t$.

Assume that

1. The probability of success (or occurrence of an event) in a given time interval is proportional to the length of the interval, i.e., $p \propto \Delta t$ or $p = \alpha \Delta t$ where α is the proportionality constant.

2. The occurrences of events are independent, i.e., probability of success in an interval of time (or space) does not depend on the what happened prior to that time or any other interval.

3. The probability of more than one success during a small time interval Δt is negligible.

As $n \to \infty$, the probability of x successes (or occurrence of an event) during a time interval T is governed by the Poisson distribution with the parameter

$$\lambda = n \cdot p = \left(\frac{T}{\Delta t}\right)(\alpha \Delta t) = \alpha T$$

Thus α is the average (mean) number of successes (occurrences) per unit time.

Worked Out Examples

Poisson process

Example: Average rate of arrival of persons in a queue is 1.5 per minute. Determine the probability that (a) at most four persons will arrive in any given minute (b) at least five will arrive during an interval of 2 minutes (c) at most 20 will arrive during an interval of 6 minutes.

Solution: α = arrival rate = 1.5.
Let X be number of persons arriving in queue

a. T = time interval = 1 minute

$$\lambda = \alpha T = (1.5)(1) = 1.5$$

$$P(X \le 4) = \sum_{x=0}^{4} f(x; 1.5) = F(4; 1.51) = .981.$$

b. T = time interval = 2 minutes

$$\lambda = \alpha T = (1.5)(2) = 3.0$$

$$P(X \ge 5) = 1 - P(X < 5) = 1 - \sum_{x=0}^{4} f(x, 3)$$
$$= 1 - F(4, 3) = 1 - .815 = .185.$$

c. T = time interval = 6 minutes

$$\lambda = \alpha T = (1.5)(6) = 9$$

$$P(X \le 20) = \sum_{x=0}^{20} f(x; 9) = F(20, 9) = 1.0.$$

Exercise

Poisson process

1. The average rate of phone calls received is 0.6 calls per minute at an office. Determine the probability that (a) there will be one or more calls in a minute (b) there will be at least three calls during 4 minutes.

Ans. **a.** $f(x \ge 1; 0.6) = 1 - F(0, 0.6)$
$= 1 - .549 = .451$

b. $f(x \ge 3; 2.4) = 1 - F(2, 2.4)$
$= 1 - .570 = .430$

2. On an average six bad cheques per day are received by a bank. Find the probability that the bank will receive (a) on any given day four bad cheques (b) 10 bad cheques on any two consecutive days.

Ans. **a.** $f(4; 6) = e^{-6}6^4/4! = 0.135$

b. $f(10; 12) = F(10, 12) - F(9, 12)$
$= .347 - .242 = 0.105$

3. At an airport, the average number of aeroplanes arriving is 10. There are only 15 runways in the airport. Determine the probability that an aeroplane will be refused landing on any given day.

Ans. $P(X \ge 15) = 1 - \sum_{x=0}^{15} f(x, 10) = 1 - F(15, 10)$
$= 1 - .09513 = .0487$

4. The number of e-mails received by a computer is at the rate of two per 3 minutes. Determine the probability that five or more e-mails are received in a duration of 9 minutes.

Ans. $\sum_{x=5}^{\infty} f(x, 6) = 1 - F(4; 6) = 1 - 0.285$
$= 0.7149$

Hint: By reproductive property of Poisson process
$\lambda = \lambda_1 + \lambda_2 + \lambda_3 = 2 + 2 + 2 = 6$ for R.V. X_1, X_2, X_3 in 3 minutes duration, so $X = X_1 + X_2 + X_3$.

5. On an average two emergency cases are received in a week (7 days) period at a hospital. Determine the probability that there are
   a. three or less emergency cases in 2 weeks period
   b. exactly eight emergency cases in 3 weeks period.

Ans. **a.** $F(3; 4) = \sum_{x=0}^{3} f(x; 4) = 0.4335$

b. $f(8;6) = F(8;6) - F(7;6)$
$= .8472 - .7440 = 0.1032.$

27.8 CONTINUOUS UNIFORM DISTRIBUTION

The probability density function $f(X)$ of a continous random variable X having *uniform* distribution over the interval $[a, b]$ is given by

$$U[a,\, b] = f(X) = \begin{cases} K, \text{ constant } a \le x \le b \\ 0, \text{ otherwise} \end{cases}$$

Here X is uniformly distributed over the interval $[a, b]$. Since $\int_{-\infty}^{\infty} f(x)dx = 1$ we have $\int_a^b K\,dx = 1$ or $(b-a)K = 1$ or $K = \frac{1}{b-a}$. So the constant k is the reciprocal of the length of the interval. Thus the continuous uniform distribution takes the form

$$U[a, b] = f(X) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

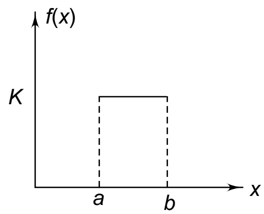


Fig. 27.4

The density (1) is uniform in value over the interval $[a, b]$. The uniform distribution (1) is also known as *rectangular* distribution since the graph of the distribution is rectangular. The constants a and b, which are known as the parameters, completely determines the distribution (1). Since

$$E(X^K) = \int_a^b x^K f(x)dx = \int_a^b x^K \left(\frac{1}{b-a}\right) dx$$
$$= \frac{1}{b-a}\left(\frac{b^{K+1} - a^{K+1}}{K+1}\right)$$

we get the mean and variance of the uniform distribution as

$$\mu = E(X) = \frac{b^2 - a^2}{(b-a)^2} = \frac{b+a}{2}$$

and

$$\sigma^2 = E(X^2) - \mu^2 = \frac{b^3 - a^3}{(b-a)\cdot 3} - \left(\frac{b+a}{2}\right)^2$$
$$= \frac{(b-a)^2}{12}$$

Solving $a = \mu - \sqrt{3}\sigma$, $b = \mu + \sqrt{3}\sigma$
The cumulative distribution function $F(x)$:

(i) when $x \le a$, $F(x) = \int_{-\infty}^{x} 0\, dx = 0$
(ii) when $a \le x \le b$, $F(x) = \int_{-\infty}^{x} = \int_{-\infty}^{a} 0 + \int_a^x \frac{1}{b-a} \cdot dx = 0 + \frac{1}{b-a}(x-a) = \frac{x-a}{b-a}$
(iii) when $x > b$, $F(x) = \int_{-\infty}^{x} = \int_{-\infty}^{a} 0 + \int_a^b \frac{1}{b-a} dx + \int_b^x 0 = \frac{1}{b-a} \cdot (b-a) = 1$

Thus

$$F(x) = \begin{cases} 0 & \text{when } x < a \\ \frac{x-a}{b-a} & \text{when } a \le x \le b \\ 1 & \text{when } x > b \end{cases}$$

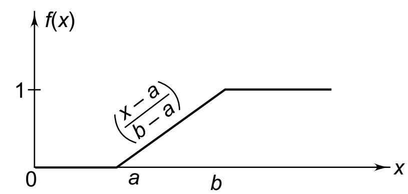


Fig. 27.5

Now for any subinterval $[c, d]$ where $a \le c < d \le b$. The probability that x lies in the interval $[c, d]$ is given by

$$P(c \le X \le d) = \int_c^d f(x)dx$$
$$= \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}.$$

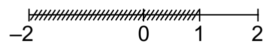
Thus the probability depends only on the length of the interval $(d - c)$ but not on the location of that interval in $[a, b]$. Therefore in continuous uniform distribution, the probability is *same* (uniform) for *all* subintervals having the *same length*.

WORKED OUT EXAMPLES

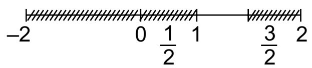
Example 1: If X is uniformly distributed in $-2 \leq x \leq 2$, find (a) $P(X < 1)$ (b) $P|X - 1| \geq 1/2)$.

Solution: (a) Since $X < 1$, it lies in the interval $[-2, 1]$, of length 3. Then

$$P(X < 1) = \frac{1-(-2)}{2-(-2)} = \frac{3}{4}$$



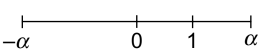
(b) If $|X - 1| \geq \frac{1}{2}$ then $X \geq \frac{3}{2}$ and $x \leq \frac{1}{2}$ i.e. x lies in the two intervals $[\frac{3}{2}, 2]$ and $[-2, \frac{1}{2}]$. So



$$\begin{aligned} P\left(|X-1| \geq \frac{1}{2}\right) &= P\left(-2 \leq X \leq \frac{1}{2}\right) \\ &\quad + P\left(\frac{3}{2} \leq X \leq 2\right) \\ &= \frac{\frac{1}{2}-(-2)}{2-(-2)} + \frac{2-\frac{3}{2}}{2-(-2)} \\ &= \frac{\frac{5}{2}+\frac{1}{2}}{4} = \frac{3}{4} \end{aligned}$$

Example 2: If X is uniformly distributed in $[-\alpha, \alpha]$ with $\alpha > 0$ then determine α such that $P(X > 1) = \frac{1}{3}$.

Solution: If $\alpha < 1$, then $P(X > 1)$ should be zero since X lies outside the given interval $[-\alpha, \alpha]$. Therefore α must be greater than 1. Now $P(X > 1) = \frac{\alpha-1}{\alpha-(-\alpha)} = \frac{\alpha-1}{2\alpha} = \frac{1}{3}$ (given). So $1 - \frac{1}{\alpha} = \frac{2}{3}$ or $\alpha = 3$.



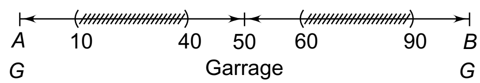
Example 3: A bus travels between two cities A and B which are 100 miles apart. If the bus has a breakdown, the distance X of the point of breakdown from city A has a uniform distribution $U[0, 100]$.

(a) There are service garages in the city A, city B and midway between cities A and B. If a breakdown occurs, a tow truck is sent from the garage closest to the point of breakdown. What is the probability that the tow truck has to travel more than 10 miles to reach the bus.

(b) Would it be more "efficient" if the three service garages were placed at 25, 50 and 75 miles from city A? Explain.

Solution:

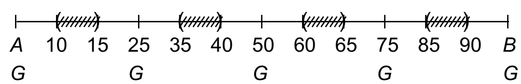
(a) If the bus breaksdown in the intervals [10, 40] miles or [60, 90] miles, then the bus have to be towed for more than 10 miles.



So probability that the bus has to be towed for more than 10 miles = probability that X lies in the intervals [10, 40] or [60, 90]. Thus the required probability is given by

$$\begin{aligned} &P(10 < X < 40 \text{ or } 60 < X < 90) \\ &= P(10 < X < 40) + P(60 < X < 90) \\ &= \frac{40-10}{100-0} + \frac{90-60}{100-0} \\ &= \frac{3}{10} + \frac{3}{10} = \frac{3}{5} \end{aligned}$$

(b) Suppose three garages are placed at 25, 50 75 miles from city A.



In this case, the bus is to be towed for more than 10 miles if the bus breaksdown in any one of the four intervals (10, 15), (35, 40), (60, 65) or (85, 90) miles. Probability is given by

$$\begin{aligned} &P(10 < X < 15 \text{ or } 35 < X < 40 \\ &\text{or} \qquad 60 < X < 65 \text{ or } 85 < X < 90) \end{aligned}$$

$= P(10 < X < 15) + P(35 < X < 40)$

$+P(60 < X < 65) + P(85 < x < 90)$

$$= \frac{15-10}{100-0} + \frac{40-35}{100-0} + \frac{65-60}{100-0} + \frac{90-85}{100-0}$$

$$= \frac{20}{100} = \frac{1}{5}.$$

Since the probability is small $\left(\frac{1}{5}\right)$ compared to $\left(\frac{3}{5}\right)$ in the case a, b is more "effective".

EXERCISE

1. A point is chosen at random from the line segment [0, 2]. What is the probability that the chosen point lies (a) $1 \le X \le \frac{3}{2}$ (b) $X \ge \frac{3}{2}$ (c) $X \le 1$ (d) $x \ge 3$

Ans. (a) $\left(\frac{3}{2} - 1\right)\frac{1}{2} = \frac{1}{4}$ (b) $\left(2 - \frac{3}{2}\right)\frac{1}{2} = \frac{1}{4}$ (c) $(1-0)\frac{1}{2} = \frac{1}{2}$ d) 0.

Hint: $f(x) = \frac{1}{2-0} = \frac{1}{2}$

2. If X is uniformly distributed in $[-\alpha, \alpha]$ with $\alpha > 0$. Then find α such that $P(X < \frac{1}{2}) = 0.7$.

Ans. $\alpha = \frac{5}{4}$

3. If a conference room cannot be reserved for more than 4 hours, find the probability that a given conference lasts more than three hours.

Ans. $\frac{1}{4}$

Hint:: $f(x) = \frac{1}{4}$, $P(X \ge 3) = (4-3)\frac{1}{4} = \frac{1}{4}$.

4. The daily amount X of coffee, in liters dispensed by a machine is uniformly distributed with $a = 7, b = 10$. Determine the probability that the amount of coffee dispensed by the machine will be (a) at most 8.8 (b) more than 7.4 but less than 9.5 (c) at least 8.5 litres.

Ans. a) 0.6 b) 0.7 (c) 0.5

Hint: $f(x) = \frac{1}{10-7} = \frac{1}{3}$

5. The driving time X from house to bus station is uniformly distributed $U[10, 50]$. If it takes 2 minutes to board the bus, determine the probability that person catches the 6.00 pm bus if he starts at 5.43 pm at his house.

Ans. $P(X \le 15) = \frac{15-10}{50-10} = \frac{1}{8} = 0.125$

Hint: Maximum time to catch bus is $6.0 - 5.43 - 0.2 = 15$ minutes.

6. Find the third and fourth moment about the mean of a uniform distribution.

Ans. $0, (b-a)^4/80$

Hint: $\mu_r = E\{(X-\mu)^r\} = \int_a^b \left\{x - \frac{(b+a)}{2}\right\}^r \frac{1}{b-a}dx = \begin{cases} 0 & \text{if } r \text{ is odd} \\ \frac{1}{r+1}\left(\frac{b-a}{2}\right)^r & \text{if } r \text{ is even} \end{cases}$

7. A bus arrives every 10 minutes at a bus stop. Assuming waiting time X for bus is uniformly distributed find the probability that a person has to wait for the bus (a) for more than 7 minutes (b) between 2 and 7 minutes

Ans. (a) $\frac{10-7}{10-0} = \frac{3}{10} = 0.3$ (b) $\frac{7-2}{10-0} = \frac{5}{10} = \frac{1}{2} = 0.5$

8. If X is uniformly distributed with mean 1 and variance $\frac{4}{3}$ find $P(X < 0)$.

Ans. $\frac{1}{4}$

Hint: Mean $= \frac{b+a}{2} = 1$, variance $= \frac{(b-a)^2}{12} = \frac{4}{3}$, $a = -1, b = 3$ $f(x) = \frac{1}{3-(-1)} = \frac{1}{4}$, $P(X < 0) = \frac{0-(-1)}{1}\frac{1}{4} = \frac{1}{4}$.

27.9 NORMAL DISTRIBUTION

Normal probability distribution or simply normal distribution is the probability distribution of a continuous random variable X, known as normal random variable or normal variate. It is given by

$$N(\overline{X}, \sigma) = f(X) = Y(X) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(X-\overline{X})^2/\sigma^2} \quad (1)$$

Here $\overline{X}$ = Arithmetic mean, σ = standard deviation, are the two parameters of the continuous distribution (1). Normal distribution (N.D.) is also known as Gaussian distribution (due to Karl Friedrich Gauss and also credited to de Moivre and Laplace). This theoretical distribution (1) is most important, simple, useful and is the corner stone of modern statistics because (a) discrete probability distributions such as Binomial, Poisson, Hypergeometric can be approximated by N.D. (b) sampling distributions 't', F, χ^2 tend to be normal for large samples and (c) it is applicable in statistical quality control in industry.

Properties of Normal Distribution (N.D.)

1. The graph of the N.D. $y = f(X)$ in the XY-plane is known as normal curve (N.C.). N.C. is (a) symmetric about y-axis (b) it is bell shaped (c) the mean, median and mode coincide and therefore N.C. is unimodal (has only one maximum point). (d) N.C. has inflection points at $\overline{x} \pm \sigma$. (e) N.C. is asymptotic to both positive x-axis and negative x-axis (see Fig. 27.6).

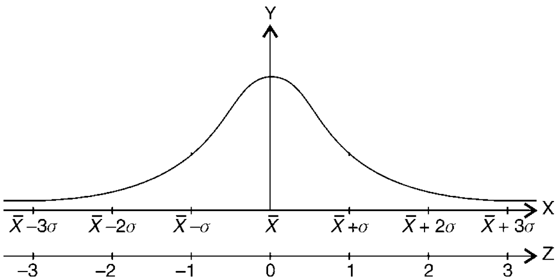


Fig. 27.6

2. Area under the normal curve is unity.

3. Probability that the continuous random variable X lies between X_1 and X_2 is denoted by probability $(X_1 \le X \le X_2)$ and is given by

$$P(X_1 \le X \le X_2) = \int_{X_1}^{X_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma}\right)^2} dx \tag{2}$$

Since (2) depends on the two parameters $\overline{x}$ and σ, we get different normal curves for different values of $\overline{x}$ and σ and it is an impractiable task to plot all such normal curves. Instead, by introducing

$$Z = \frac{x - \overline{x}}{\sigma}$$

the R.H.S. integral in (2) becomes independent (dimensionless) of the two parameters $\overline{x}$ and σ. Here Z is known standard (or standardized) variable (variate).

4. Change of scale from x-axis to z-axis.

$$P(X_1 \le X \le X_2) = \int_{X_1}^{X_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma}\right)^2} dx$$

$$P(Z_1 \le Z \le Z_2) = \int_{Z_1}^{Z_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-Z^2/2} \sigma\, dZ$$

$$= \int_{Z_1}^{Z_2} \frac{1}{\sqrt{2\pi}} e^{-Z^2/2} dZ \tag{3}$$

where

$$Z_1 = \frac{X_1 - \overline{X}}{\sigma}, \quad Z_2 = \frac{X_2 - \overline{X}}{\sigma}.$$

5. Error function or probability integral is defined as

$$P(Z) = \frac{1}{\sqrt{2\pi}} \int_0^Z e^{-Z^2/2} dZ \tag{4}$$

Now (3) can be written using (4) as

$$P(Z_1 \le Z \le Z_2) = \int_{Z_1}^{Z_2} \frac{1}{\sqrt{2\pi}} e^{-Z^2/2} dZ$$

$$= P(Z_2) - P(Z_1) \tag{5}$$

Normal distribution $N(\overline{x}, \sigma)$ transformed by the standard variable Z is given by

$$N(0, 1) = Y(Z) = \frac{1}{\sqrt{2\pi}} e^{-Z^2/2}$$

with mean 0 and standard deviation 1. $N(0, 1)$ is known as "Standard Normal Distribution" and its normal curve as standard normal curve (Fig. 27.7). The probability integral (4) is tabulated for various values of Z varying from 0 to 3.9 and is known as normal table (see A12). Thus the entries in the normal table gives (represents) the area under the normal curve between the ordinates $Z = 0$ to Z (shaded in the figure). Since normal curve is symmetric about y-axis, the area from 0 to $-Z$ is same as the area from 0 to Z. For this reason, normal table is tabulated only for positive values of Z. Hence the determination of normal probabilities (3) reduce to the determination of areas under

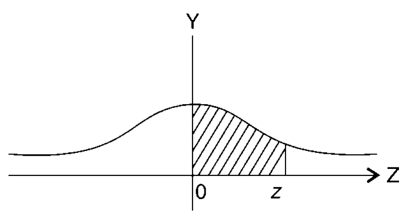


Fig. 27.7

the normal curve by (5) (see Fig. 27.7). Therefore

$$P(X_1 \le X \le X_2) = P(Z_1 \le Z \le Z_2) = P(Z_2) - P(Z_1)$$
$$= (\text{Area under the N.C. from 0 to} Z_2)$$
$$- (\text{Area under the N.C. from 0 to} Z_1)$$

6. Area under the N.C. is distributed as follows:
68.27% area lies between $\overline{X} - \sigma$ to $\overline{X} + \sigma$
i.e., between $-1 \le Z \le 1$
94.45% area lies between $\overline{X} - 2\sigma$ to $\overline{X} + 2\sigma$
i.e., between $-2 \le Z \le 2$
99.73% area lies between $\overline{X} - 3\sigma$ to $\overline{X} + 3\sigma$
i.e., between $-3 \le Z \le 3$

Note: 50% area in the Z-interval $(-.745, +.745)$
99% area in the Z-interval $(-2.58, +2.58)$

Arithmetic Mean of Normal Distribution

By definition
the A.M. of a continuous distribution $f(x)$ is given by

$$\text{A.M.} = \frac{\int_{-\infty}^{\infty} x\, f(x)dx}{\int_{-\infty}^{\infty} f(x)dx}.$$

Consider the normal distribution with B, C as the parameters, i.e., $N(B, C) = f(x) = \frac{1}{c\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-B}{c}\right)^2}$. Then

$$\text{A.M.} = \overline{X} = \int_{-\infty}^{\infty} x \cdot \frac{1}{c\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-B}{c}\right)^2} dx$$

since $\int_{-\infty}^{\infty} f(x)dx$ = area under the normal curve $= 1$
Put $\frac{x-B}{c} = z$ so $x = B + cz$, $dx = cdz$

$$\text{So} \quad \overline{X} = \int_{-\infty}^{\infty} (B + cz) \frac{1}{c\sqrt{2\pi}} e^{-\frac{1}{2}z^2} cdz$$
$$= B \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + \frac{c}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} dz$$
$$= B + \frac{c}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} d\left(\frac{z^2}{2}\right)$$

since $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 1$.

$$= B + \frac{c}{\sqrt{2\pi}} \left.\frac{e^{-\frac{z^2}{2}}}{-1}\right|_{-\infty}^{\infty} = B + 0$$

So $\overline{X} = B$.

Variance for Normal Distribution

By definition

$$\text{Variance} = \int_{-\infty}^{\infty} (x - \overline{x})^2 f(x)dx$$
$$= \int_{-\infty}^{\infty} x^2 f(x)dx + \overline{x}^2 \int_{-\infty}^{\infty} f(x)dx$$
$$-2\overline{x} \int_{-\infty}^{\infty} x\, f(x)dx$$
$$= \int_{-\infty}^{\infty} x^2\, f(x)dx + \overline{x}^2 - 2\overline{x}\overline{x}$$

since $\int_{-\infty}^{\infty} f(x)dx = 1$ and $\int_{-\infty}^{\infty} x\, f(x)dx = \overline{x}$.
Consider the first integral in the R.H.S.

$$\int_{-\infty}^{\infty} x^2 f(x)dx = \int_{-\infty}^{\infty} x^2 \frac{1}{c\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\overline{x}}{c}\right)^2} dx$$

Put $\frac{x-\overline{x}}{c} = z$ so $x = \overline{x} + cz$, $dx = cdz$

$$\int_{-\infty}^{\infty} x^2 f(x)dx = \int_{-\infty}^{\infty} (\overline{x} + cz)^2 \frac{1}{c\sqrt{2\pi}} e^{-\frac{z^2}{2}} cdz$$
$$= \frac{1}{\sqrt{2\pi}} \left[c^2 \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} dz \right.$$
$$\left. + \overline{x}^2 \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz + 2c\overline{x} \int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} dz \right]$$
$$= \frac{-c^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} zd\left(e^{-\frac{z^2}{2}}\right) + \overline{x}^2 \cdot 1 + 2c\overline{x} \cdot 0$$
$$= \frac{-c^2}{\sqrt{2\pi}} \left. z e^{-\frac{z^2}{2}} \right|_{-\infty}^{\infty} + \frac{c^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz + \overline{x}^2$$
$$= 0 + c^2 \cdot 1 + \overline{x}^2$$

Substituting this value

$$\text{Variance} = \int_{-\infty}^{\infty} x^2 f(x)dx - \overline{x}^2 = [c^2 + \overline{x}^2] - \overline{x}^2 = c^2$$

Thus the standard deviation (s.d.), of N.D. is c.

Book Work: Show that the area under the normal curve is unity.

Proof: Normal probability distribution is given by

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\overline{x})^2/2\sigma^2}$$

Then the area A under the normal curve is

$$A = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\overline{x})^2/2\sigma^2} dx$$

Put $\frac{x-\overline{x}}{\sigma} = z$, so $dx = \sigma dz$

or $$A = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2} \sigma dz = \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-z^2/2} dz$$

or $$A \cdot A = \left[\sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-x^2/2} dx\right]\left[\sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-y^2/2} dy\right]$$

Here x, y are dummy variables (Fig. 27.8).

$$A^2 = \frac{2}{\pi} \int_0^{\infty}\int_0^{\infty} e^{-(x^2+y^2)/2} dx\, dy$$

Put $x = r\cos\theta$, $y = r\sin\theta$, $J =$ Jacobian $= r$
Limits for $r: 0$ to ∞, $\theta = 0$ to 2π (to cover the first quadrant $0 < x < \infty$, $0 < y < \infty$)

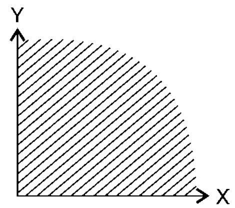


Fig. 27.8

So $$A^2 = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \int_0^{\infty} e^{-\frac{r^2}{2}} r\, dr\, d\theta$$

$$= \int_0^{\infty} e^{-\frac{r^2}{2}} d\left(\frac{r^2}{2}\right) = \left.\frac{e^{-\frac{r^2}{2}}}{-1}\right|_0^{\infty} = 1$$

Thus $A =$ area under the normal curve $= 1$.

Book Work: Prove that for normal distribution the mean deviation from the mean equals to $\frac{4}{5}$ of standard deviation approximately.

Proof: Let $\overline{x}$ and σ be the mean and standard deviation of the normal distribution. Then by definite mean deviations from the mean $= \int_{\infty}^{\infty} |x-\overline{x}| f(x) dx$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |x-\overline{x}| e^{-\frac{(x-\overline{x})^2}{(2\sigma^2)}} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma |z| e^{-\frac{1}{2}z^2} \sigma dz$$

where $z = \frac{x-\overline{x}}{\sigma}$ and $dx = \sigma dz$.

$$= \sigma\sqrt{\frac{2}{\pi}} \int_0^{\infty} z e^{-\frac{z^2}{2}} dz = \sigma\sqrt{\frac{2}{\pi}}\left[-e^{-\frac{z^2}{2}}\right]_0^{\infty} = \sqrt{\frac{2}{\pi}}\sigma$$

$$= 0.7979\sigma \approx 0.8\sigma = \frac{8}{10}\sigma = \frac{4}{5}\sigma$$

Fitting of Normal Distribution

Given any frequency distribution, a normal distribution (i.e., a normal curve) can be fitted to it using $N(\overline{X}, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\overline{X}}{\sigma}\right)^2}$.

Here $\overline{X} =$ A.M. and $\sigma =$ s.d. are calculated from the given frequency distribution.

Procedure

Consider a frequency distribution (F.D.)

$$\begin{array}{ll} L_1 - U_1 & f_1 \\ L_2 - U_2 & f_2 \\ \vdots & \\ n - U_n & f_n \end{array}$$

and $N =$ total frequency $= \sum_{i=1}^{n} f_i$

Let $\overline{X}$ and σ be the A.M. and S.D. of the F.D.

Here L_i, U_i are the true lower and upper limits of the ith class.

I. Compute standard variable $z_i = \frac{X_i - \overline{X}}{\sigma}$ for each of the true lower limit X_i of the n classes (there will be $n+1$ such quantities).

II. Compute area under N.C. (from normal table A12) from 0 to z_i.

III. Normal probability of a class is obtained by taking the difference between the successive areas calculated in step II. (when z_i's are of opposite sign, add the successive areas).

IV. Expected or theoretical frequencies are obtained by multiplying probabilities in III by N, the total frequency of the F.D.

Worked Out Examples

Normal distribution

Example 1: Find the area A under the normal curve:

a. to the left of $z = -1.78$

b. to the left of $z = 0.56$

c. to the right of $z = -1.45$

d. corresponding to $z \geq 2.16$

e. corresponding to $-0.80 \leq z \leq 1.53$

f. to the left of $z = -2.52$ and to right of $z = 1.83$

Solution: Refer to normal table (A12)

a. $A = 0.5 -$ Area (0 to -1.78) (Fig. 27.9)

$= 0.5 -$ Area (0 to 1.78) due to symmetry

$= 0.5 - 0.4625 = 0.0375$ (from table)

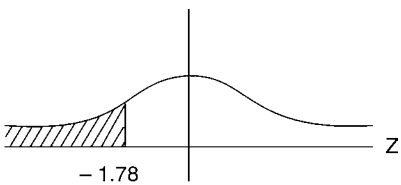


Fig. 27.9

b. $A = 0.5 +$ Area from 0 to 0.56 (Fig. 27.10)

$= 0.5 + 0.2123$ (from table)

$= 0.7123$

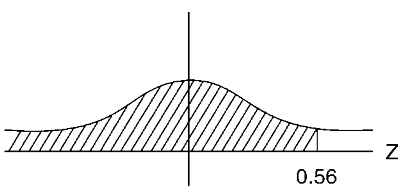


Fig. 27.10

c. $A = 0.5 +$ Area from 0 to -1.45 (Fig. 27.11)

$= 0.5 +$ Area from 0 to 1.45

$= 0.5 + 0.4265 = 0.9265$

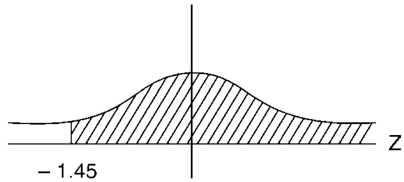


Fig. 27.11

d. $A = 0.5 - A$ (0 to 2.16)

$= 0.5 - 0.4846 = 0.0154$ (Fig. 27.12)

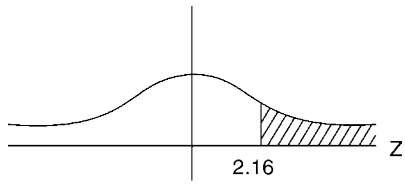


Fig. 27.12

e. $A =$ Area from (0 to -0.8)

$+$Area from (0 to 1.53)

$=$ Area from (0 to 0.8)

$+$Area from (0 to 1.53)

$= 0.4370 + 0.2881 = 0.7251$ (Fig. 27.13)

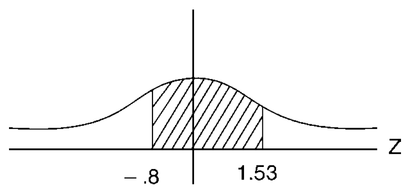


Fig. 27.13

f. $A = [0.5 - A(0 \text{ to } 2.52)] + [0.5 - A(0, 1.83)]$

$= (0.5 - 0.4941) + (0.5 - 0.4664)$

$= 0.0059 + 0.0336$

$= 0.0395$ (Fig. 27.14)

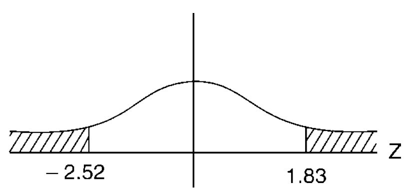


Fig. 27.14

Example 2: If z is normally distributed with mean 0 and variance 1, find

a. $P(z \geq -1.64)$

b. $P(-1.96 \leq z \leq 1.96)$

c. $P(z \leq 1)$

d. $P(z \geq 1)$

Solution:

a. $P(z \geq -1.64) = 0.5 + A(0 \text{ to } -1.64)$ (Fig. 27.15)

$$= 0.5 + A(0 \text{ to } 1.64)$$
$$= 0.5 + 0.4495 = 0.9495$$

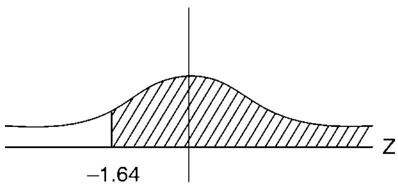


Fig. 27.15

b. $P(-1.96 \leq z \leq 1.96)$

$= 2A(0 \text{ to } 1.96)$ by symmetry

$= 2(0.4750) = 0.9500$ (Fig. 27.16)

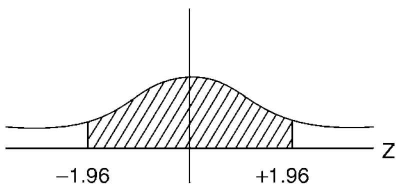


Fig. 27.16

c. $P(z \leq 1) = 0.5 + A(0 \text{ to } 1)$

$$= 0.5 + 0.3413$$
$$= 0.8413 \text{ (Fig. 27.17)}$$

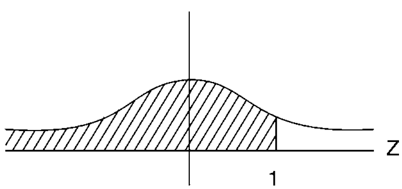


Fig. 27.17

d. $P(z \geq 1) = 0.5 - A(0 \text{ to } 1)$

$$= 0.5 - 0.3413 = 0.1587 \text{ (Fig. 27.18).}$$

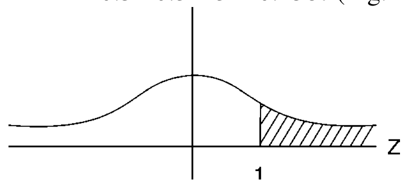


Fig. 27.18

Example 3: Determine the value of z such that (a) area to the right of z is 0.2266 (b) area to the left of z is 0.0314.

Solution: Here the areas (entries of the normal table) are given, the values of z (1st column) are determined.

a. Since area $0.2266 < \frac{1}{2}$ is to the right of z, z must be positive such that area from 0 to z is $0.5 - 0.2266 = 0.2734$. From normal table for area 0.2734, the value of z is 0.75 (Fig. 27.19).

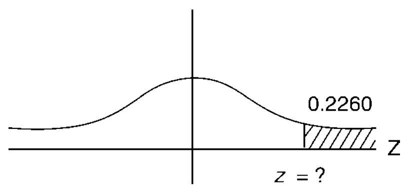


Fig. 27.19

b. Since area $0.134 < \frac{1}{2}$ is to the left of z, z must be negative. So determine z such that area from 0 to z is $0.5 - 0.134 = 0.4686$. From table A12, $z = -1.86$ (Fig. 27.20).

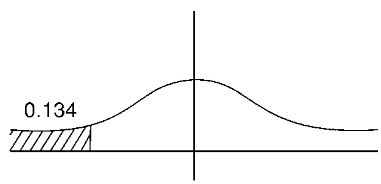


Fig. 27.20

Example 4: Find the (a) mean and (b) standard deviation of an examination in which grades 70 and

88 correspond to standard scores of -0.6 and 1.4 respectively.

Solution: Standard variable $z = \frac{X-\overline{X}}{\sigma}$.

$$\text{Here} \qquad -0.6 = \frac{70-\overline{X}}{\sigma} \quad \text{so} \quad \overline{X} - 0.6\sigma = 70$$

$$1.4 = \frac{88-\overline{X}}{\sigma} \quad \text{so} \quad \overline{X} + 1.4\sigma = 88$$

Solving $\overline{X} = 75.4$, $\sigma = 9$ are the mean and standard deviation.

Example 5: Determine the minimum mark a student must get in order to receive an A grade if the top 10% of the students are awarded A grades in an examination where the mean mark is 72 and standard deviation is 9.

Solution: The 0.1 area to the right of z corresponds to the top 10% of the students (see Fig. 27.21). From table if area from 0 to z is 0.4, then $z = 1.28$. Given $\overline{X} = 72$, $\sigma = 9$, we have $1.28 = z = \frac{X-\overline{X}}{\sigma} = \frac{X-72}{9}$,

$$X = 72 + 11.52 = 83.52 \simeq 84$$

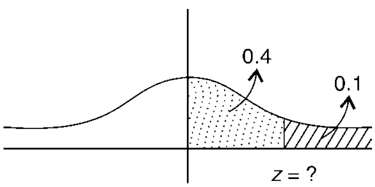


Fig. 27.21

So a student must get a minimum (or more) of 84 marks to get an A grade.

Example 6: Find the mean and standard deviation of a normal distribution in which 7% of the items are under 35 and 89% are under 63 (see Fig. 27.22).

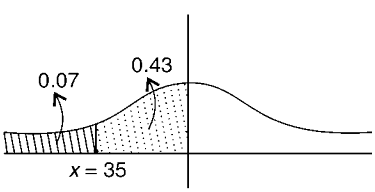


Fig. 27.22

Solution: Let X be the continuous random variable. Given that $P(X < 35) = 0.07 < \frac{1}{2}$. So z must be negative such that area from 0 to z is $0.5 - 0.07 = 0.43$. From normal table $z = -1.48$.
Given that $P(X < 63) = 0.89 > \frac{1}{2}$. So z must be positive such that area from 0 to z is $0.89 - 0.5 = 0.39$ (Fig. 27.23).

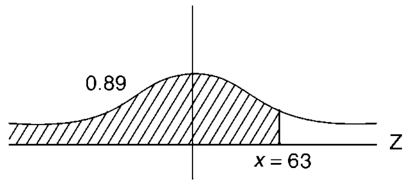


Fig. 27.23

From normal table $z = 1.23$.
Since $z = \frac{X-\overline{X}}{\sigma}$, we have

$$-1.48 = \frac{35-\overline{X}}{\sigma} \quad \text{or} \quad \overline{X} - 1.48\sigma = 35$$

$$1.23 = \frac{63-\overline{X}}{\sigma} \quad \text{or} \quad \overline{X} + 1.23\sigma = 63$$

Solving the arithmetic mean $\overline{X} = 50.3$ and standard deviation $\sigma = 10.33$.

Example 7: When the mean of marks was 50% and S.D. 5% then 60% of the students failed in an examination. Determine the 'grace' marks to be awarded in order to show that 70% of the students passed. Assume that the marks are normally distributed.

Solution: Let X be the marks obtained in the exam. Given $\overline{X} = 0.5$, $\sigma = \text{s.d.} = 0.05$.

Before grace marks were awarded, 60% failed. Since 60% failure corresponds 0.6 area, z_1 must be positive (Fig. 27.24). Determine z_1 such that the area to its left is 0.6. The value of z_1 for which the area is 0.1 is 0.25.

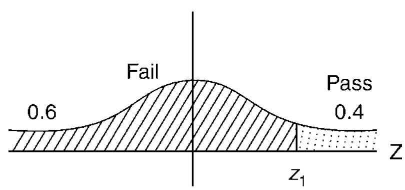


Fig. 27.24

$$0.25 = z_1 = \frac{X_1 - 0.5}{0.05} \quad \text{so} \quad X_1 = 0.5125$$

After grace marks were awarded, 70% passed examination. The area 0.70 $\left(> \frac{1}{2}\right)$ corresponds pass students (Fig. 27.25). Determine z_2 such that the area to its right is 0.7. So z_2 must be negative and from table, $z_2 = -0.52$. Then

$$z_2 = -0.52 = \frac{X - 0.5}{0.05} \quad \text{or} \quad X_2 = 0.4740$$

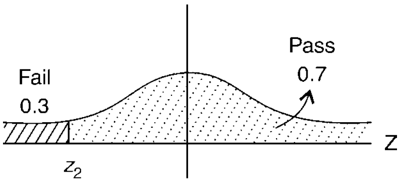


Fig. 27.25

Thus the minimum pass mark for a student is 51.25 before grace while the minimum pass mark is 47.40 after grace. So grace marks awarded is 51.25 − 47.40 = 3.85.

Example 8: Assume that the 'reduction' of a person's oxygen consumption during a period of Transcendenta Meditation (T.M.) is a continuous random variable X normally distributed with mean 37.6 cc/mt and s.d. 4.6 cc/mt. Determine the probability that during a period of T.M. a person's oxygen consumption will be reduced by (a) at least 44.5 cc/mt (b) at most 35.0 cc/mt (c) anywhere from 30.0 to 40.0 cc/mt.

Solution: $z = \frac{X - \overline{X}}{\sigma} = \frac{X - 37.6}{4.6}$

a. For $X = 44.5$, $z = \frac{44.5 - 37.6}{4.6} = 1.5$ (Fig. 27.26)

$$P(X \geq 44.5) = P(z \geq 1.5) = 0.5 - 0.4332 = 0.068$$

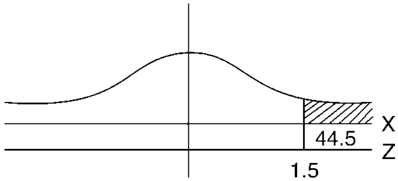


Fig. 27.26

b. For X=35.0, $z = \frac{35.0 - 37.6}{4.6} = -0.5652$ (Fig. 27.26)

$$\begin{aligned} P(X \leq 35) &= P(z \leq -0.5652) \\ &= 0.5 - 0.2157 \\ &= 0.2843. \end{aligned}$$

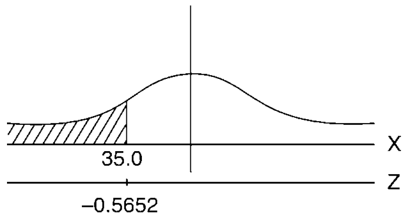


Fig. 27.27

c. For $X_1 = 30$, $z_1 = \frac{30 - 37.6}{4.6} = -1.6521$

For $X_2 = 40$, $z_2 = \frac{40 - 37.6}{4.6} = 0.52173$ (Fig. 27.26)

$$\begin{aligned} P(30 \leq X \leq 40) &= P(-1.6521 \leq z \leq 0.52173) \\ &= 0.4505 + 0.1985 \\ &= 0.6490. \end{aligned}$$

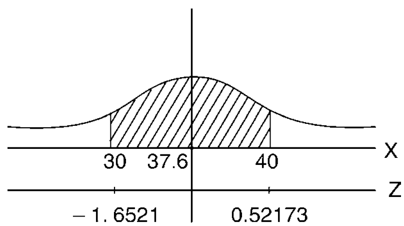


Fig. 27.28

Example 9: The marks X obtained in mathematics by 1000 students in normally distributed with mean 78% and s.d. 11% (Fig. 27.29). Determine (a) how many students got marks above 90%? (b) what was the highest mark obtained by the lowest 10% of students? (c) semi-inter quartile range (d) within what limits did the middle 90% of students lie?

Solution: Here $z = \frac{X - \overline{X}}{\sigma} = \frac{X - 0.78}{0.11}$

a. For $X = 0.9$, $z = \frac{0.9 - 0.78}{0.11} = 1.09$.

$P(X > 0.9) = P(z > 1.09) = 0.5 - 0.3621 = 0.1379$

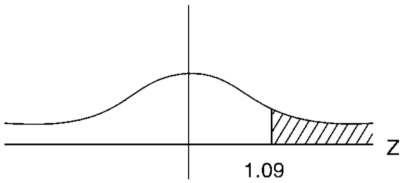


Fig. 27.29

Number of students with marks above 90%

$= 1000 \times P(X > 0.9) = 1000 \times 0.1379 = 137.9 \approx 138.$

b. The lowest 10% students constitute 0.1 area $\left(< \frac{1}{2}\right)$ of extreme left tail. So z_1 must be negative. From table $0.4 = 0.5 - 0.1 =$ $0.5-$ Area 0.1 from 0 to z_1 so $z_1 = -1.28$.

Thus $\quad -1.28 = z_1 = \dfrac{X - 0.78}{0.11}$ or $X = 0.6392$

(see Fig. 27.30)

Thus the highest mark obtained by the lowest 10% of students is $63.92 \approx 64\%$.

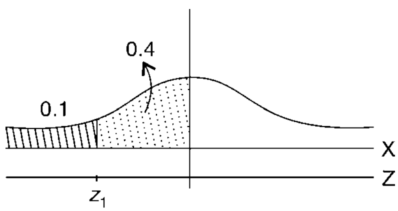


Fig. 27.30

c. Quartiles Q_1, Q_2, Q_3 divide the area into four equal parts. The value of z_1 corresponding to the first quartile Q_1 is such that the area to its left is 0.25. From table $z_1 = -0.67$. Similarly, $z_3 = 0.67$ corresponding to Q_3. Now $-0.67 = z_1 = \frac{X_1 - 0.78}{0.11}$. So the quartile mark is $X_1 = 0.7063 = 70.63\%$. Similarly, $X_3 = 85.37\%$. Thus the semi-inter quartile range $= \frac{Q_3 - Q_1}{2} = \frac{85.37 - 70.63}{2} = 7.37$ (Fig. 27.31).

d. Middle 90% correspond to 0.9 area, leaving 0.05 area on both sides. Then the corresponding z's are ± 1.64 (refer Fig. 27.32).

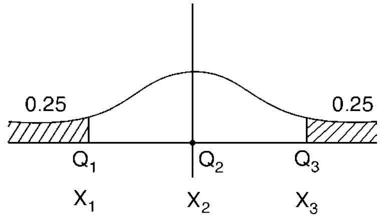


Fig. 27.31

$$1.64 = z_2 = \frac{X_2 - 0.78}{0.11} \quad \text{so} \quad X_2 = 96.04$$

$$-1.64 = z_1 = X_1 - 0.78 \quad \text{so} \quad X_1 = 59.96$$

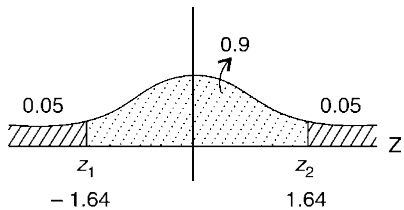


Fig. 27.32

Thus the middle 90% have marks in between 60 to 96.

Example 10: Fit a normal distribution to the following data (frequency distribution):

S. No.	Class	Observed frequency f_i
1	5–9	1
2	10–14	10
3	15–19	37
4	20–24	36
5	25–29	13
6	30–34	2
7	35–39	1

$$\text{Total frequency} = \sum_{i=1}^{7} f_i = 100$$

Solution:

1	2	3	4	5	6	7	8
S. No.	Class	Frequency f_i	True lower class limit X_i	Standard variate $z_i = \frac{X_i-20}{5}$	Area from 0 to z_i	Area for class (Probability P)	Expected or theoretical frequency $= NP = 100P$
1	5–9	1	4.5	-3.1	0.4990	0.0169	$1.69 \approx 2$
2	10–14	10	9.5	-2.1	0.4821	0.1178	$11.78 \approx 12$
3	15–19	37	14.5	-1.1	0.3643	0.3245	$32.45 \approx 32$
			19.5	-0.1	0.0398		
4	20–24	36	24.5	0.9	0.3159	0.3557	$35.57 \approx 36$
5	25–29	13	29.5	1.9	0.4713	0.1554	$15.54 \approx 16$
6	30–34	2	34.5	2.9	0.4981	0.0268	$2.68 \approx 3$
7	35–39	1	39.5	3.9	0.5000	0.0019	$0.19 \approx 0$
		Total Frequency $N = 100$					

Note: Entries in column 7 are obtained by subtracting successive values in column 6 whenever they (in 6) are of the same sign. Add the values in column 6 when they are of opposite sign.

Exercise

Normal distribution

1. Determine the area under the normal curve
   a. between $z = -1.2$ and $z = 2.4$
   b. between $z = 1.23$ and $z = 1.87$
   c. between $z = -2.35$ and $z = -0.5$
   d. to the left of $z = -1.90$
   e. to the left of $z = 1.0$
   f. to the right of $z = -2.40$
   g. to the left of $z = -3.0$ and to the right of $z = 2.0$.

Ans. (a) 0.8767 (b) 0.0786 (c) 0.2991 (d) 0.0287 (e) 0.8413 (f) 0.9918 (g) 0.0241

2. Find the value of z such that
   a. area between -0.23 and z is 0.5722
   b. area between 1.15 and z is 0.0730
   c. area between $-z$ and z is 0.9.

Ans. (a) $z = 2.08$ (b) $z = 0.1625$
(c) $z = -1.65$ to $+1.65$

3. **a.** Calculate the standard marks of two students whose marks are 93 and 62 in an examination given that the mean mark is 78 and s.d. is 10.
   b. If the standard marks of two students are -0.6 and 1.2, determine their respective marks.

Ans. (a) $z = 1.5, -1.6$ (b) $X = 72, 90$

4. Determine the probability that the amount of cosmic radiation X a pilot of jet plane will be exposed is more than 5.20 m rem if X is normally distributed with mean 4.35 m rem and s.d. 0.59 m rem.

Ans. $P(X > 5.20) = P(z > 1.44)$
$= 0.5 - 0.4251 = 0.0749$.

5. Suppose the life span X of certain motors is normally distributed with mean 10 years and s.d. 2 years. If the manufacturer is ready to replace only 3% of motors that fail, how many years of guarantee can he offer (Fig. 27.33).

Ans. $-1.88 = z = \frac{X-\overline{X}}{\sigma} = \frac{X-10}{2}$, $X = 6.24$ years

6. Determine the expected number of boys whose weight is
   a. between 65 and 70 kg
   b. greater than or equal to 72 kg

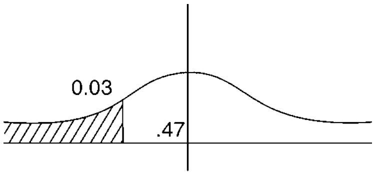


Fig. 27.33

if the weight X of 800 boys follows normal distribution with $\overline{X} = 66, \sigma = 5$.

Ans. **a.** $P(65 \le X \le 70) = P(-0.20 < z < 0.80)$
$= 0.0793 + 0.2881 = 0.3674$

Number $= 800(0.3674) = 294$

b. $P(X \ge 72) = P(z \ge 1.2) = 0.5 - 0.3849$
$= 0.1151$

Number of boys $= 800(0.1151) = 92$

7. Calculate the mean and s.d. of a normal distribution in which 31% are under 45 and 8% are over 64.

Ans. $\overline{X} = 50, \sigma = 10$

8. Assume that the average life span of computers produced by a company is 2040 hours with s.d. of 60 hours. Find the expected number of computers whose life span is
   a. more than 2150 hours
   b. less than 1950 hours
   c. more than 1920 hours and less than 2160 hours
   from a pool of 2000 computers assuming that the life span X is normally distributed.

Ans. **a.** $P(X > 2150) = P(z > 1.833) =$
$0.5 - 0.4664 = 0.0336$.

Expected number of computers whose life span is more than 2150 hours $= 2000(0.0336) = 67$.

b. $P(X < 1950) = P(z < -1.33)$
$= 0.5 - 0.40821 = 0.0918$

Expected number $= 2000(0.0918) = 184$

c. $P(1920 \le X \le 2160) = P(-2 \le z \le 2)$
$= 2(0.4772) = 0.9544$.

Expected number $= 2000 \times 0.9544 = 1909$.

9. If the top 15% of the students receives A grade and bottom 10% receives F grades in a mathematics examination, determine the
   a. minimum mark to get an A grade
   b. minimum mark to pass (not to get F grade). Assume that the marks are normally distributed with mean 76 and s.d. 15.

Ans. (a) 92 (b) 57

10. A university awards distinction, first class, second class, third class or pass class according as the student gets 80% or more; 60% or more; between 45% and 60%; between 30% and 45%; or 30% or more marks respectively. If 5% obtained distinction and 10% failed, determine the percentage of students getting second class. Assume that marks X are normally distributed.

Ans. 34% second class.

Hint: $P(X < 30) = 0.10$ failed; $P(X \ge 80) = 0.05$ distinction, $\frac{30-X}{\sigma} = -1.28$, $\frac{80-X}{\sigma} = 1.64$, $\overline{X} = 52, \sigma = 17.12$

$P(45 < X < 60) = P(-0.41 < z < 0.47) \le 0.1591 + 0.1808 = 0.3399$.

11. The amount of pollutant X released by an industry should lie between 30 and 35. Assume that X is normally distributed with mean $\overline{X} = 33$ and s.d. $\sigma = 3$. The industry gets a profit of Rs. 100 when $30 < X < 35$; Rs. 50 when $25 < X \le 30$ or $35 \le X < 40$ and incurs a fine of Rs. 60 otherwise. Determine the expected profit for the industry.

Ans. $100(0.5890) + 50(0.396) - 60(0.0137)$
$=$ Rs. 79.

EXERCISE

Fitting of normal distribution

1. Fit a normal curve to the following data:

Class	60–62	63–65	66–68	69–71	72–74
Frequency	5	18	42	27	8

Ans. $4.13 \approx 4$ $20.68 \approx 21$ $38.92 \approx 39$

$27.71 \approx 28 \quad 7.43 \approx 7$

Hint: $\overline{X} = 67.45, \sigma = 2.92, N = 100$.

2. Fit a normal distribution to the following frequency distribution

x:	2	4	6	8	10
f:	1	4	6	4	1

Ans. $0.97 \approx 1 \quad 3.9 \approx 4 \quad 6.1 \approx 6 \quad 3.9 \approx 4$
$0.97 \approx 1.0$

Hint: $\overline{X} = 6, \sigma = 2, N = 16$, x is taken as the mid value of the class, i.e., 2 is mid value of the class (1, 3), etc.

3. Fit a normal curve to the following observed data:

Class	9.3–9.7	9.8–10.2	10.3–10.7	10.8–11.2
f	2	5	12	17

Class	11.3–11.7	11.8–12.2	12.3–12.7	12.8–13.2
f	14	6	3	1

Ans. $1.704 \approx 2 \quad 5.562 \approx 6 \quad 11.7420 \approx 12$
$15.624 \approx 16 \quad 13.942 \approx 14 \quad 7.62 \approx 8$
$2.712 \approx 3 \quad 0.168 \approx 1$

Hint: $\overline{X} = 11.09$, s.d. $= 0.733, N = 60$.

4. Fit a normal distribution to the following data:

Class	150–158	159–167	168–176	177–185
f	9	24	51	66

Class	186–194	195–203	204–212	213–221	222–230
f	72	48	21	6	3

Ans. 9.0 25.4 51.5 71.2 67.8 44.6 20.2
6.3 1.4

Hint: $\overline{X} = 184.3, \sigma = 14.54, N = 300$.

27.10 NORMAL APPROXIMATION TO BINOMIAL DISTRIBUTION

For large n, the calculation of binomial probabilities is very cumbersome. In such cases they are computed by approximation procedures. For $n \to \infty$ and $p \to 0$ B.D. can be approximated by Poisson distribution with $\lambda = np$.
For $n \to \infty$ and $p \not\to 0$, i.e., *p not* close to 0 or 1, B.D. can be approximated by normal distribution.

Theorem: *Let X be a binomial random variable with mean* $\overline{X} = np$ *and s.d.* $= \sqrt{npq}$ *then the limiting form of the distribution of*

$$z = \frac{X - np}{\sqrt{npq}}$$

as $n \to \infty$ *is standard normal distribution* $N(z; 0, 1)$.

Normal approximation to B.D. will be fairly good even when

a. *n is small and p is close to* $\frac{1}{2}$
b. *both np and nq are* ≥ 5.

WORKED OUT EXAMPLES

Normal approximation to B.D.

Example 1: If 10% of the truck drivers on road are drunk determine the probability that out of 400 drivers randomly checked

a. at most 32
b. more than 49
c. at least 35 but less than 47 drivers are drunk on the road.

Solution: Here p = probability of a driver drunk $= \frac{10}{100} = 0.1$ and $n = 400$ = no. of trials.
Let X = number of truck drivers drunk.
Here X is a binomial random variable with B.D. $= b(x; 400, 0.1)$. This B.D. can be approximated by normal distribution with A.M. $= \overline{X} = np = 400 \times \frac{10}{100} = 40$ and s.d. $= \sigma = \sqrt{npq} = \sqrt{400 \times \frac{10}{100} \times \frac{90}{100}} = 6$.

a. For $X = 32, z = \frac{X - \overline{X}}{\sigma} = \frac{31.5 - 40}{6} = -\frac{8.5}{6} = -1.416$ since X is treated as continuous variable, values upto and more that 31.5 will be rounded up to 32 (Fig. 27.34).

$$P(X < 32) = P(z \leq -1.416) = 0.5 - 0.4222$$
$$= 0.0778$$

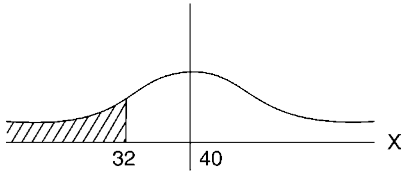


Fig. 27.34

b. Since 49.5 and above are rounded to 50, for $X > 49$, $z = \frac{49.5-40}{6} = 1.58$ (Fig. 27.35)

$$P(X > 49) = P(z \geq 1.58) = 0.5 - 0.4429$$
$$= 0.0571$$

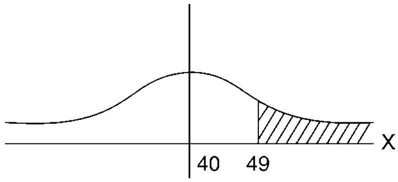


Fig. 27.35

c. $X \geq 35$ includes values of X upto 34.5 and $x < 47$ includes values of X upto 46.5 (Fig. 27.36).

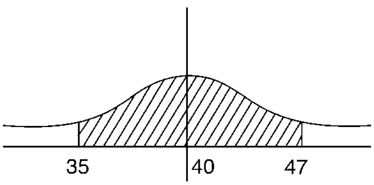


Fig. 27.36

$$z_1 = \frac{34.5 - 40}{6} = -0.916,$$
$$z_2 = \frac{46.5 - 40}{6} = 1.083$$
$$P(35 \leq X < 47) = P(-0.916 \leq z \leq 1.083)$$
$$= 0.3212 + 0.3599 = 0.681.$$

Example 2: A pair of dice is rolled 180 times. Determine the probability that a total of 7 occurs

a. at least 25 times
b. between 33 and 41 times inclusive
c. exactly 30 times.

Solution: Sum 7 occurs in a single throw of a pair of dice as follows: (1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3).

p = probability of 7 occurring $= \frac{6}{36} = \frac{1}{6}$, $n = 180$; X = number of occurrences of a sum of 7 in a pair of dice = a binomial random variable.

Treating X as a continuous R.V., B.D. $= b\left(x; 180, \frac{1}{6}\right)$ can be approximated by normal distribution with A.M. $= \overline{X} = np = 180 \times \frac{1}{6} = 30$, σ = s.d. $= \sqrt{npq} = \sqrt{180 \cdot \frac{1}{6} \cdot \frac{5}{6}} = 5$

a. For $X \geq 25$, X includes up to 24.5. Thus

$$z = \frac{24.5 - 30}{5} = -1.1$$
$$P(X \geq 25) = P(z \geq -1.1) = 0.5 + 0.3643$$
$$= 0.8643 \text{ (Fig. 27.37)}$$

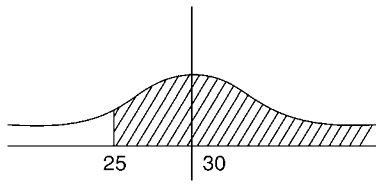


Fig. 27.37

b. $P(33 \leq X \leq 41) = P(0.5 \leq z \leq 2.3)$

$= 0.4893 - 0.1915 = 0.2978$ (Fig. 27.38)

since $z_1 = \frac{32.5-30}{5} = 0.5$, $z_2 = \frac{41.5-30}{5} = 2.3$

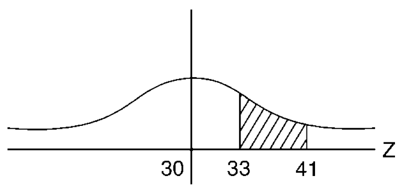


Fig. 27.38

c. $P(X = 30) = P(29.5 \leq X \leq 30.5)$

$= P(-0.1 \leq z \leq 0.1)$

$= 2(0.0398) = 0.0796$ (Fig. 27.39)

since $z_1 = \frac{29.5-30}{5} = -0.1$, $z_2 = \frac{30.5-30}{5} = 0.1.$

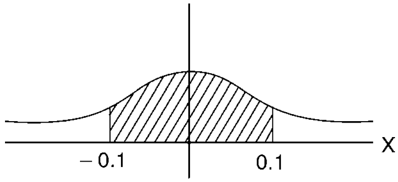


Fig. 27.39

EXERCISE

Normal approximation of B.D.

1. Determine the probability that by guess-work a student can correctly answer 25 to 30 questions in a multiple-choice quiz consisting of 80 questions. Assume that in each question with four choices, only one choice is correct and student has no knowledge.

Ans. $P(25 \le X \le 30) = P(1.16 \le z \le 2.71) =$ $0.9960 - 0.8770 = 0.1196$.

Hint: $\bar{X} = np = (80)\left(\frac{1}{4}\right) = 20$,

$\sigma = \sqrt{npq} = \sqrt{80 \cdot \frac{1}{4} \cdot \frac{3}{4}} = 3.873$

$z_1 = \frac{24.5-20}{3.873} = 1.16$, $z_2 = \frac{30.5-20}{3.573} = 2.71$.

2. Find the probability that out of 100 patients
   a. between 84 and 95 inclusive
   b. fewer than 86,
   will survive a heart-operation given that the chances of survival is 0.9.

Ans. **a.** $P(84 \le X \le 95) = P(-2.166 \le z \le 1.833)$ $= 0.4850 + 0.4664 = 0.9514$

Hint: $\bar{X} = np = (100)(0.9) = 90$,
$\sigma = \sqrt{npq} = \sqrt{(100)(0.9)(0.1)} = 3$

$z_1 = \frac{83.5-90}{3} = -2.1666$, $z_2 = \frac{95.5-90}{3}$ $= 1.8333$

b. $P(X < 86) = P(z_1 \le -1.5)$ $= 0.5 - 0.4332 = 0.0668$

Hint: $z_1 = \frac{85.5-90}{3} = -1.5$.

3. Find the probability P that the number of heads occurring, when a fair coin is tossed 12 times, is between 4 and 7 inclusive by (a) B.D. (b) normal approximation to B.D.

Ans. **a.** $P = \sum_{x=4}^{7} b\left(x; 12, \frac{1}{2}\right) = \sum_{x=4}^{7} x_{C_4}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{12-x}$

$= \frac{495+4096+924+792}{4096} = 0.7332$

b. $P = P(3.5 \le X \le 7.5) =$ $P(-1.45 \le z \le -0.87) =$

$= 0.4265 + 0.3078 = 0.7343$

Hint: $\bar{X} = np = 12 \cdot \frac{1}{2} = 6$,
$\sigma = \sqrt{npq} = \sqrt{12\frac{1}{2} \cdot \frac{1}{2}} = 1.73$.

4. The probability that a patient needs an ICU is 0.05 in a hospital with 600 patients. How many ICU's should be available so that the probability of none of the patients of the hospital are turned away due to lack of ICU's is more than 0.90.

Ans. $P(x < x_1) = P(0 < z < z_1) > 0.90$,
$z_1 = \frac{x_1-30}{5.3}$

so $z_1 = 1.28$ or $\frac{x_1-30}{5.3} = z_1 > 1.28$ so $x_1 >$ $36.784 \approx 37$

Hint: $\bar{X} = np = (600)(0.05) = 30$,
$\sigma = \sqrt{npq} = \sqrt{(600)(0.5)(0.95)} = 5.3$.

27.11 ERROR FUNCTION

Error function of x (also known as error integral), denoted by $erf\,x$, is defined as

$$\boxed{erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt} \qquad (1)$$

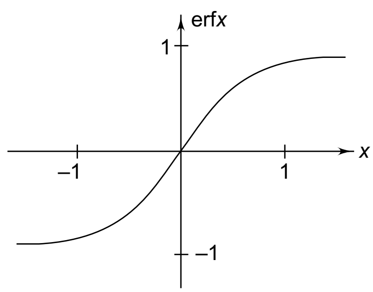


Fig. 27.40

It occurs in probability theory, thermodynamics, heat conduction problems. $Erf\,x$ is known as a 'special function', since (1) can not be evaluated in terms of 'elementary functions' by the usual methods of calculus.

Properties

1. $erf(0) = 0$
2. $erf(\infty) = \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-t^2} dt = \frac{2}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2} = 1$
3. It is defined for all x, $-\infty < x < \infty$, monotonically increasing in the interval $(0, \infty)$; passes through origin. Asymptotic to $y = \pm 1$.
4. It is an odd function since

$$erf(-x) = \frac{2}{\sqrt{\pi}} \int_0^{-x} e^{-t^2} dt = \frac{2}{\sqrt{\pi}} \int_0^x e^{-v^2}(-dv)$$

where $v = -t$

$$= -\frac{2}{\sqrt{\pi}} \int_0^x e^{-v^2} dv = -erf(x)$$

5. $erf(-\infty) = -erf(\infty) = -1$
6. $erf(x) + erf(-x) = erf(x) - erf(x) = 0$
7. Complementary error function of x, denoted by

$$erfc(x) = 1 - erf(x) = erf(\infty) - erf(x)$$
$$= \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-t^2} dt - \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$
$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$$

8. $erfc(x) + erfc(-x) = [1 - erf(x)] + [1 - erf(-x)] = 2 - erf(x) + erf(x) = 2$
9. Probability integral (Normal distribution function) of mathematical statistics is defined as

$$\boxed{\phi(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\theta} e^{-\omega^2/2} d\omega}$$

put $t = \omega/\sqrt{2}$ in the error function (1); then

$$erf(\theta) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{2}\theta} e^{-\omega^2/2} d\omega$$

Thus the error function and probability integral are related by

$$\boxed{erf(\theta) = 2\phi(\sqrt{2}\theta) - 1} \text{ or}$$

$$\boxed{\phi(x) = \tfrac{1}{2} + \tfrac{1}{2} erf\left(\tfrac{x}{\sqrt{2}}\right)}$$

10. Approximate formula (due to C. Hastings Jr.)

$$\boxed{erf(x) \approx 1 - (a_1 p + a_2 p^2 + a_3 p^3) e^{-x^2}}$$

where $p = \frac{1}{1+0.47047x}$, $a_1 = 0.3480242$, $a_2 = -0.0958798$, $a_3 = 0.7478556$, accurate upto ± 0.000025.

Note The factor $\frac{2}{\sqrt{\pi}}$ is included in the definition of error function to normalize it so that $erf(\infty) = 1$ since $\int_0^\infty e^{-t^2} dt = \frac{\sqrt{\pi}}{2}$.

11. $P(\mu - K\sigma \le X \le \mu + K\sigma) = erf\left\{\frac{K}{\sqrt{2}}\right\}$.

Let X denote the measured quantity in a certain experiment. Then the measurement error is indicated by the probability of an event such as $\mu - K\sigma \le X \le \mu + K\sigma$. Thus

$$P(\mu - K\sigma \le X \le \mu + K\sigma)$$
$$= N(\mu + K\sigma) - N(\mu - K\sigma)$$
$$= \Phi(K) - \Phi(-K)$$
$$= 2\Phi(K) - 1$$
$$= erf\left\{\frac{K}{\sqrt{2}}\right\} \text{ using above result 9.}$$

For example, for $K = 3$, we get

$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) = 0.977.$$

Thus, on the average in only 0.3 per cent of the trials, the Gaussian random variable deviates from its mean by more than ± 3 standard deviations.

WORKED OUT EXAMPLES

Example 1: Expand $erf(x)$ in ascending powers of x.

Solution: By definition

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt = \frac{2}{\sqrt{\pi}} \int_0^x \sum_{n=0}^{\infty} \frac{(-t^2)^n}{n!} dt$$

since $e^t = \sum_{n=0}^{\infty} \frac{t^n}{n!}$. Now carrying term by term integration, we have

$$erf(x) = \frac{2}{\sqrt{\pi}}\left[\int_0^x \left(1-\frac{t^2}{1!}+\frac{t^4}{2!}-\frac{t^6}{3!}+\frac{t^8}{4!}+\cdots\right)dt\right]$$

$$= \frac{2}{\sqrt{\pi}}\left[t-\frac{t^3}{3}+\frac{t^5}{5\cdot 2!}-\frac{t^7}{7\cdot 3!}+\frac{t^9}{9\cdot 4!}+\cdots\right]\Bigg|_{t=0}^{x}$$

$$erf(x) = \frac{2}{\sqrt{\pi}}\left[x-\frac{x^3}{3}+\frac{x^5}{10}-\frac{x^7}{42}+\frac{x^9}{216}-\cdots\right]$$

Example 2: Find $\frac{d}{dx}[erf(ax)]$.

Solution: From the above result replacing x by ax we have

$$erf(ax) = \frac{2}{\sqrt{\pi}}\left[ax-\frac{a^3x^3}{3}+\frac{a^5x^5}{10}-\frac{a^7x^7}{42} + \frac{a^9x^9}{216}-\cdots\right]$$

Differentiating both sides w.r.t. 'x' we get

$$\frac{d}{dx}[erf(ax)] = \frac{2}{\sqrt{\pi}}\left[a-a^3\cdot x^2+a^5\frac{x^4}{2!}-a^7\frac{x^6}{3!} + a^9\frac{x^8}{4!}-\cdots\right]$$

$$= \frac{2a}{\sqrt{\pi}}\left[1-a^2x^2+\frac{(a^2x^2)^2}{2!}-\frac{(a^2x^2)^3}{3!} + \frac{(a^2x^2)^4}{4!}-\cdots\right]$$

$$= \frac{2a}{\sqrt{\pi}}e^{-a^2x^2}$$

Example 3: Compute $erf(0.5)$ correct to three decimal places.

Solution: Putting $x = 0.5$ in example 1, above, we have

$$erf(0.5) = \frac{2}{\sqrt{\pi}}\left[\frac{1}{2}-\left(\frac{1}{2}\right)^3\frac{1}{3}+\left(\frac{1}{2}\right)^5\frac{1}{10} - \left(\frac{1}{2}\right)^7\frac{1}{42}+\left(\frac{1}{2}\right)^9\frac{1}{216}-\cdots\right]$$

$$= \frac{0.922544642}{1.77245384} = 0.52049$$

Example 4: Show that

$$\int_0^\infty e^{-x^2-2ax}dx = \frac{\sqrt{\pi}}{2}e^{a^2}\left[1-erf(a)\right]$$

Solution: $\int_0^\infty e^{-x^2-2ax}dx = \int_0^\infty e^{-(x^2+2ax+a^2)}\cdot e^{a^2}\,dx = e^{a^2}\int_0^\infty e^{-(x+a)^2}dx$. Put $x+a=t$, so t varies from a to ∞ and $dx = dt$

$$= e^{a^2}\int_a^\infty e^{-t^2}dt = e^{a^2}\cdot\frac{\sqrt{\pi}}{2}erfc(a)$$

$$= e^{a^2}\cdot\frac{\sqrt{\pi}}{2}[1-erf(a)]$$

EXERCISE

1. Prove that $\frac{d}{dx}[erfc(ax)] = -\frac{2a}{\sqrt{\pi}}e^{-a^2x^2}$.

   Hint: $\frac{d}{dx}[erfc(ax)] = \frac{d}{dx}[1-erf(ax)] = -\frac{d}{dx}erf(ax)$ use W.E. 2., above

2. Compute $erf(0.3)$ correct to three decimal places.

Ans. 0.3248

   Hint: Put $x = 0.3248$ in W.E. 1.

3. Prove that
   (a) $\int_a^b e^{-t^2}dt = \frac{\sqrt{\pi}}{2}[erf(b)-erf(a)]$
   (b) $\int_{-b}^b e^{-t^2}dt = \sqrt{\pi}erf(b)$
   Hint: (a) $\int_a^b = \int_a^0 + \int_0^b = \int_0^b - \int_0^a$
   (b) $\int_{-b}^b = \frac{\sqrt{\pi}}{2}[erf(b)-erf(-b)] = \frac{\sqrt{\pi}}{2}[erf(b)+erf(b)]$

4. Show that

$$\int_0^t erfc(ax)dx = t\cdot erfc(at) - \frac{e^{-a^2t^2}}{a\sqrt{\pi}}+\frac{1}{a\sqrt{\pi}}$$

   Hint: Integrating by parts

$$x\cdot erfc(ax)|_0^t - \int_0^t x\cdot d(erfc(ax))$$

   Use result in exercise example 1, then

$$= t\cdot erf(at) + \frac{a}{\sqrt{\pi}}\cdot\frac{e^{-a^2x^2}}{a^2}\Bigg|_t^0$$

27.12 THE EXPONENTIAL DISTRIBUTION

Many scientific experiments involve the measurement of the duration of time X between an initial point of time and the occurrence of some phenomenon of interest. For example X is the life time of a light bulb which is turned on and left until it burns out. The continuous random variable X having the probability density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geqslant 0 \\ 0 & \text{if } x < 0 \end{cases}$$

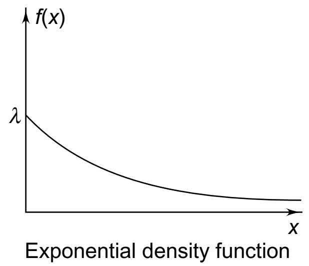


Fig. 27.41

is said to have an *exponential distribution.* Here the only parameter of the distribution is λ which is greater than zero. This distribution, also known as the *negative exponential distribution*, is a special case of the gamma distribution (with $r = 1$). Examples of random variables modeled as exponential are

a. (inter-arrival) time between two successive job arrivals
b. duration of telephone calls
c. life time (or time to failure) of a component or a product
d. service time at a server in a queue
e. time required for repair of a component

The exponential distribution occurs most often in applications of **Reliability Theory** and **Queuing Theory** because of the memoryless property and relation to the (discrete) **Poisson Disribution**. Exponential distribution can be obtained from the Poisson distribution by considering the inter-arrival times rather than the number of arrivals.

Mean and Variance

For any $r \geq 0$,

$$E(X^r) = \int_0^\infty x^r f(x)dx = \int_0^\infty x^r \lambda e^{-\lambda x} dx$$

put $\lambda x = t,\ x = \frac{t}{\lambda},\ dx = \frac{1}{\lambda}dt$. Then

$$E(X^r) = \int_0^\infty \left(\frac{t}{\lambda}\right)^r \cdot \lambda \cdot e^{-t} \cdot \frac{1}{\lambda}dt = \frac{1}{\lambda^r}\int_0^\infty e^{-t}t^r dt$$

$$E(X^r) = \frac{\Gamma(r+1)}{\lambda^r}$$

In particular with $r = 0$,

$$\int_0^\infty f(x)dx = \Gamma(1) = 1$$

(i.e., $f(x)$ is a probability density function).
With $r = 1$, mean $= \mu = E(X) = \frac{\Gamma(2)}{\lambda} = \frac{1}{\lambda}$

with $r = 2$, variance $= \sigma^2 = E(X^2) - \{E(X)\}^2 = \frac{\Gamma(3)}{\lambda^2} - \frac{1}{\lambda^2}$

$$\sigma^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Note: Both the mean and standard deviation of the exponential distribution are equal to $\frac{1}{\lambda}$.

Cumulative Distribution Function

$$F(x) = \int_0^x f(t)dt = \int_0^x \lambda e^{-\lambda t}dt = \left.\frac{\lambda e^{-\lambda t}}{-\lambda}\right|_{t=0}^{x}$$

$$F(x) = 1 - e^{-\lambda x} \quad \text{for } x \geqslant 0,$$

and $F(x) = 0$ when $x < 0$

$F(x)$ gives the probability that the "system" will "die" before x units of time have passed.

Probability Calculations

For any $a \geqslant 0$,

$$P(X \geq a) = P(X > a) = 1 - F(a) = e^{-\lambda a}$$

$$\begin{aligned} P(a \leq X \leq b) &= P(a \leq X < b) = P(a < X < b) \\ &= P(a < X \leq b) = F(b) - F(a) \\ &= e^{-\lambda a} - e^{-\lambda b} \end{aligned}$$

In table (A22) in appendix, the values of e^{-t} are tabulated for $t = 0.00(0.01)7.99$.

Corollary 1: $P\left(X > \frac{1}{\lambda}\right) = e^{-\lambda\frac{1}{\lambda}} = e^{-1} = 0.368 < \frac{1}{2}$

Survival Function

It gives the probability that the "system" survives more than x units of time and is given by

$$P(X > x) = 1 - F(x) = \begin{cases} 1 & \text{if } x < 0 \\ e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$$

Memoryless or Markov Property

Among all distributions of non-negative continuous variables, only the exponential distributions have "no memory" (like the discrete geometric distribution) which results in analytical tractibility.

For any $s > 0,\ t > 0$

$$P(X > s+t | X > s) = \frac{P(\{X > s+t\} \cap \{X > s\})}{P(X > s)}$$

$$= \frac{P(X > s+t)}{P(X > s)} \quad (1)$$

When $X > s+t$ then X is also greater than s i.e., $X > s$. Since $\{X > s+t\} \cap \{X > s\} = \{X > s+t\}$

Thus the event $X > s$ in the numerator is redundant because both events can occur iff $X > s+t$.

Now

$$P(X > s+t | X > s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}$$

$$P(X > s+t | X > s) = P(X > t) \quad (2)$$

This memoryless property asserts that the conditional probability of additional waiting time is the *same* as the unconditional probability of the original waiting time. Thus the distribution of additional lifetime is exactly the same as the original distribution of lifetime, so at each point of time the component shows no effect of wear. In other words the distribution of "remaining" lifetime is independent of current age. In this sense, the exponential distribution has "*no memory*" of the past.

Combining (1) and (2) we have

$$P(X > s+t) = P(X > s) \cdot P(X > s+t | X > s)$$
$$= P(X > s) \cdot P(X > t)$$

which yields the famous functional equations known as *Cauchy equation*.

$$h(s+t) = h(s)h(t), \quad s > 0, t > 0$$

Here $h(s) = P\{X > s\},\ s > 0$.

Example: Suppose when a person arrives, one telephone booth has just been occupied (engaged) while another telephone booth has been occupied since (say 110 minutes) long. Then the probability distribution of the length of waiting time (to use the phone) will be the same for either phone booths. Therefore it does not matter which phone booth the person descides to wait!

Worked Out Examples

Example 1: Let the mileage (in thousands of miles) of a particular tyre be a random variable X having the probability density

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

Find the probability that one of these tyres will last (1) at most 10,000 miles (b) anywhere from 16,000 to 24,000 miles (c) at least 30,000 miles. (d) Find the mean (e) Find the variance of the given probability density function.

Solution: (a) Probability that a tyre will last almost 10,000 miles

$$= P(X \leq 10) = \int_0^{10} f(x)dx$$

$$= \int_0^{10} \frac{1}{20}e^{-x/20}dx$$

$$= \frac{1}{20} \cdot e^{-x/20} \cdot \left(\frac{-20}{1}\right)\Bigg|_0^{10}$$

$$= 1 - e^{-\frac{1}{2}} = 0.3934$$

(b) $$P(16 \leq X \leq 24) = \int_{16}^{24} f(x)dx$$

$$= \int_{16}^{24} \frac{1}{20}e^{-x/20}dx$$

$$= -e^{-\frac{x}{20}}\Big|_{16}^{24} = e^{-\frac{4}{5}} - e^{-\frac{6}{5}}$$
$$= 0.148$$

(c) $$P(X \geq 30) = \int_{30}^{\infty} f(x)dx$$
$$= \int_{30}^{\infty} \frac{1}{20} e^{-x/20} dx = -e^{x/20}\Big|_{30}^{\infty} = e^{-\frac{3}{2}}$$
$$= 0.2231$$

(d) $$\text{Mean } = \mu = \int_{-\infty}^{\infty} x \cdot f(x)dx$$
$$= \int_{0}^{\infty} x \cdot \frac{1}{20} e^{-\frac{x}{20}} dx$$
$$= -\int_{0}^{\infty} x \cdot d\left(e^{-\frac{x}{20}}\right)$$
$$= -xe^{-\frac{x}{20}} - 20e^{-\frac{x}{20}}\Big|_{0}^{\infty} = 0 - (-20)$$
$$\mu = 20 = \frac{1}{\lambda}$$

(e) $$\text{Variance } = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$
$$= \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

Consider

$$\int_{-\infty}^{\infty} x^2 f(x)dx = \int_{0}^{\infty} x^2 \frac{1}{20} e^{-\frac{x}{20}} dx$$
$$= -x^2 e^{-\frac{x}{20}}\Big|_{0}^{\infty} + 2 \cdot 20 \cdot \int_{0}^{\infty} \frac{1}{20} \cdot xe^{-x/20} dx$$
$$= 0 + 2 \cdot 20 \cdot \mu = 2.20.20 = 2.20^2$$

Then $$\sigma^2 = \int_{0}^{\infty} x^2 f(x)dx - \mu^2 = 2.20^2 - 20^2$$
$$= 20^2 = \frac{1}{\lambda^2}$$

Example 2: The length of time for one person to be served at a cafeteria is a random variable X having an exponential distribution with a mean of 4 minutes. Find the probability that a person is served in less than 3 minutes on at least 4 of the next 6 days.

Solution: The probability that a person is served at a cafeteria in less than 3 minutes is

$$P(T < 3) = 1 - P(T \geq 3)$$

Since the mean $\mu = \frac{1}{\lambda} = 4$ or $\lambda = \frac{1}{4}$, the exponential distribution is $\frac{1}{4}e^{-\frac{x}{4}}$. Now

$$P(T < 3) = 1 - P(T \geq 3) = 1 - \int_{3}^{\infty} \frac{1}{4} e^{-\frac{t}{4}} dt$$

$$P(T < 3) = 1 - \frac{1}{4} e^{-\frac{t}{4}} \cdot \left(\frac{-4}{1}\right)\Big|_{3}^{\infty} = 1 - e^{-\frac{3}{4}}$$

Let X represent the number of days on which a person is served in less than 3 minutes. Then using the binomial distribution, the probability that a person is served in less than 3 minutes on at least 4 of the next 6 days is

$$P(X \geq 4) = \sum_{x=4}^{6} {}^6C_x (1 - e^{-3/4})^x (e^{-3/4})^{6-x} = 0.3968$$

EXERCISE

1. Let T be the time (in years) to failure of certain components of a system. The random variable T has exponential distribution with mean time to failure $\beta = 5$. If 5 of these components are in different systems, find the probability that at least 2 are still functioning at the end of 8 years.

Ans. 0.2627

Hint: $P(T > 8) = \frac{1}{5} \int_8^{\infty} e^{-t/5} dt$ $= e^{-8/5} \simeq 0.2$, $P(X \geq 2) = \sum_{x=2}^{\infty} b(x; 5, 0.2) =$ $1 - \sum_{x=0}^{1} b(x, 5, 0.2) = 1 - 0.7373$

2. If a random variable X has the exponential distribution with mean $\mu = \frac{1}{\lambda} = \frac{1}{2}$ calculate the probabilities that (a) X will lie between 1 and 3 (b) X is greater than 0.5 (c) X is at most 4.

Ans. (a) 0.133 (b) 0.368 (c) 0.98168

Hint: $PDF \ \ f(x) = 2e^{-2x}$ (a) $\int_1^3 2e^{-2x} dx = e^{-2} - e^{-6}$
(b) $\int_{0.5}^{\infty} 2e^{-2x} dx = e^{-1}$ (c) $\int_0^4 2e^{-2x} dx = 1 - e^{-4}$

3. The life (in years) of a certain electrical switch has an exponential distribution with an average life of $\frac{1}{\lambda} = 2$. If 100 of these switches are installed in

different systems, find the probability that at most 30 fail during the first year.

Hint: $P(T > 1) = \int_1^\infty \frac{1}{2} e^{-\frac{t}{2}} \ dt = +e^{-\frac{1}{2}} = 0.606$

Ans. $P(X \le 30) = \sum_{x=0}^{30} b(x, 2, 0.606) = \sum_{x=0}^{30} {}^{100}C_x (0.606)^x \ (0.39346)^{100-x}$

4. Suppose the life length X (in hours) of a fuse has exponential distribution with mean $\frac{1}{\lambda}$. Fuses are manufactured by two different processes. Process I yields an expected life length of 100 hours and process II yields an expected life length of 150 hours. Cost of production of a fuse by process I is Rs. C while by the Process II it is Rs 2C. A fine of Rs K is levied if a fuse lasts less than 200 hours. Determine which process should be preferred?

Ans. Prefer Process I if $C > 0.13K$

Hint: $c_1 = c$ if $X > 200$
$= c + k$ if $X \le 200$

$$E(c_1) = c \cdot P(X > 200) + (c + k)P(X \le 200)$$

$$= c \cdot e^{-\frac{1}{100} \cdot 200} + (c + k)(1 - e^{-\frac{1}{100} \cdot 200})$$

$$= k(1 - e^{-2}) + c$$

$E(c_2) = k(1 - e^{-4/3}) + 2c, \qquad E(c_2) - E(c_1) = c - 0.13k$

5. Suppose N_t be a discrete random variable denoting the number of arrivals in time interval $(0, t]$. Let X be the time of the next arrival, so X is the elapsed time between the occurrences of two successive events. Assuming that N_t is Poisson distributed with parameter λt, show that X is exponentially distributed.

Here λ is the expected numbers of events occurring in one unit of time.

Ans. $P(X > t) = P(N_t = 0) = \dfrac{e^{-\lambda t}(\lambda t)^0}{0} = e^{-\lambda t}$

6. If the average rate of job submission is $\lambda = 0.1$ jobs/second, find the probability that an interval of 10 seconds elapses without job submission.

Ans. $P(X \ge 10) = \int_{10}^\infty 0.1 e^{0.1t} dt = e^{-1} = 0.368$

Hint: Assume that the number of arrivals/unit time is poisson distributed and the inter arrival time X is exponentially distributed with parameter λ.

7. Let the mileage (in thousands of miles) of a certain radial tyre is a random variable with exponential distribution with mean 40,000 miles. Determine the probability that the tyre will last (a) at least 20,000 km (b) at most 30,000 km.

Ans. (a) $P(X \ge 20,000) = e^{-0.5} = 0.6065$
(b) $P(X \le 30,0001 = 1 - e^{-0.75} = 0.5270$

8. The amount of time (in hours) required to repair a T.V. is exponentially distributed with mean $\frac{1}{2}$. Find the (a) probability that the repair time exceeds 2 hours (b) the conditional probability that repair takes at least 10 hours given that already 9 hours have been spent repairing the TV.

Ans. (a) $P(X > 2) = e^{-1} = 0.3679$
(b) $P(X \ge 10 | X > 9) = P(X > 1) = e^{-0.5} = 0.6065$
(because of the memoryless property).

9. The duration of time X in seconds between presses of the white rat on a bar, which are periodically conditioned, has an exponential distribution with parameter $\lambda = 0.20$. Find the probability that the duration is more than one second but less than 3 seconds (b) more than 3 seconds.

Ans. (a) $P(1 \le X \le 3) = e^{-0.2(1)} - e^{-(0.2)3} = 0.819 - 0.549 = 0.270$
(b) $P(X > 3) = e^{-0.2(3)} = 0.549$

10. The time X (seconds) that it takes a certain on-line computer terminal (the elapsed time between the end of user's inquiry and the beginning of the system's response to that inquiry) has an exponential distribution with expected time 20 seconds. Compute the probabilities (a) $P(X \le 30)$ (b) $P(X \ge 20)$ (c) $P(20 \le X \le 30)$ (d) For what value of t is $P(X \le t) = 0.5$ (i.e., t is the fiftieth percentile of the distribution)

Ans. (a) 0.777 (b) 0.368 (c) 0.145 (d) 13.863

27.13 THE GAMMA DISTRIBUTION

Consider a system consisting of one original component and $(r-1)$ spare components such that when the original component fails, one of the $(r-1)$ spare components is used. If this component fails, one of the $(r-2)$ spare components is used. System fails only when the original component and all the $(r-1)$ spare components fail. Assume that the lifetimes $X_1, X_2, \ldots, X_r$ of the r duplicates of the essential components have infinite lifetimes (except for the original component). Suppose each of the random variables $X_1, X_2, \ldots, X_r$ have the same exponential distribution with parameter λ and are probabilistically independent. Then the lifetime (time until failure) of the entire system is the sum $Y = \sum_{i=1}^{r} X_i$ having the *gamma distribution* with density function

$$f(y) = \begin{cases} \frac{\lambda^r y^{r-1} e^{-\lambda y}}{\Gamma(r)}, & \text{if } y \geq 0 \\ 0, & \text{if } y < 0 \end{cases} \tag{1}$$

(1) is a skewed distribution.

The two parameters of (1) are the positive numbers λ and r (although r need *not* be an integer). If r is a positive integer, then gamma distribution (1) is known as *Erlang distribution*. Introducing $V = \lambda y$, (1) reduces to

$$f(v) = \frac{1}{\lambda} f\left(\frac{v}{\lambda}\right) = \frac{1}{\lambda}\left\{\lambda^r \left(\frac{v}{\lambda}\right)^{r-1} \cdot \frac{e^{-v}}{\Gamma(r)}\right\}$$

$$= \begin{cases} \frac{v^{r-1}e^{-v}}{\Gamma(r)} & \text{if } v \geq 0 \\ 0 & \text{if } v < 0 \end{cases} \tag{2}$$

The probability density function of the random variable V given by (2) is known as the "*standard gamma function*" with parameter r (and is independent of λ). When $r = 1$, the density function (2) reduces to the density function of exponential distribution with the parameter $\lambda = 1$. For large r (say $r \geq 50$) (2) resembles a normal distribution with mean and variance approximately equal to r. The gamma distribution with parameter $\lambda = \frac{1}{2}$ and $r = \frac{\nu}{2}$ (where ν is a positive integer) reduces to the chi-squared distribution with ν degrees of freedom credited to Karl Pearson (1857–1936) and F.R. Helmert (1843–1917).

The *incomplete gamma function* defined by

$$F_V(t) = \int_0^t \frac{v^{r-1}e^{-v}}{\Gamma(r)} dv = I_r(t), \qquad t \geq 0 \tag{3}$$

is tabulated in the tables of the appendix A23 to A28 for $r = 1(1)5$, $t = 0.2(0.2)8.0(0.5)15.0$ and for $r = 6(1)10$, $t = 1.0(0.2)8.0(0.5)17.0$.

Now

$$P(Y \geq a) = P(Y > a) = 1 - F(a) = 1 - I_r(\lambda a)$$

and

$$P(a \leq Y \leq b) = F(b) - F(a) = I_r(\lambda b) - I_r(\lambda a)$$

Moments of the Gamma Distribution

For any $k \geq 0$,

$$E(Y^k) = E\left(\frac{V^k}{\lambda^k}\right) = \frac{1}{\lambda^k} E(V^k)$$

$$= \frac{1}{\lambda^k}\int_0^\infty v^k \cdot \left(\frac{v^{r-1}e^{-v}}{\Gamma(r)}\right) dv$$

$$= \frac{1}{\lambda^k \Gamma(r)} \int_0^\infty e^{-v} v^{k+r-1} dv$$

Then $$E(Y^k) = \frac{\Gamma(r+k)}{\lambda^k \Gamma(r)} \tag{4}$$

For $k = 0$ in (4) we have $\int_0^\infty f(y)dy = 1$ so $f(y)$ is a probability density function.

For $k = 1$, in (4) we get the mean $= \mu = E(Y) = \frac{\Gamma(r+1)}{\lambda\Gamma(r)} = \frac{r\Gamma(r)}{\lambda\Gamma(r)}$.

So $\mu = \frac{r}{\lambda}$. (5)

With $k = 2$ in (4), we get

variance $= \sigma^2 = E(Y^2) - \{E(Y)\}^2$

$$= \frac{\Gamma(r+2)}{\lambda^2\Gamma(r)} - \frac{r^2}{\lambda^2}$$

$$\sigma^2 = \frac{r(r+1)\Gamma(r)}{\lambda^2\Gamma(r)} - \frac{r^2}{\lambda^2} = \frac{r}{\lambda^2} \tag{6}$$

Thus the parameter r and λ are determined from (5) and (6) as

$$\lambda = \frac{\mu}{\sigma^2}, \; r = \frac{\mu^2}{\sigma^2} \tag{7}$$

Relation Between Exponential, Gamma and Poisson Distributions

Suppose the lifetimes of a batch of components each have exponential distribution with parameter λ. Staring at time $t = 0$, the first component is used until its extinction (until it "dies" or "fails"). Replace the component by another instantaneously and wait unit this new component also fails. Continuing this Process, stop at a given time t. Then the number of failed components L is a random variable having the Poisson distribution with $\lambda^* = \lambda t$. Also the lifetime of the entire process $Y = \sum_{i=1}^{k} X_i$ follows gamma distribution with parameters λ and k.

Let $X_1, X_2, \ldots, X_k$ be the lifetimes of the first k components that have failed. Assume that each lifetime X_i has an exponential distribution with parameter λ and are probabilistically independent. Recall that, then $Y = \sum_{i=1}^{k} X_i$ has a gamma distribution with parameters λ and k. If $X_1 + X_2 + \ldots + X_k$ (total lifetime of the process) $\leq t$ then in this case at least k components have all failed, (one after the other) before the time is up.

Now probability of the event that at least k components have failed at the time of termination of the experiment is given by

$$P(L \geq k) = P\{Y \leq t\} = I_k(\lambda t)$$

$$= \int_0^{\lambda t} \frac{v^{k-1} e^{-v}}{\Gamma(k)} dv$$

Integrating by parts, we have

$$= \frac{v^{k-1}}{\Gamma(k)} \cdot \left(\frac{e^{-v}}{-1} \right) \Bigg|_{v=0}^{\lambda t} + \int_0^{\lambda t} \frac{(k-1) v^{k-2} \cdot e^{-v}}{\Gamma(k)} dv$$

$$= -\frac{(\lambda t)^{k-1} e^{-\lambda t}}{(k-1)!} + \int_0^{\lambda t} \frac{v^{k-2} \cdot e^{-v}}{\Gamma(k-1)} dv.$$

Integrating by parts $(k-1)$ more times,

$$P(L \geq k) = -\sum_{i=1}^{k-1} \frac{(\lambda t)^i e^{-\lambda t}}{i!} + \int_0^{\lambda t} \frac{e^{-v}}{\Gamma(1)} dv$$

But $\Gamma(1) = 1$ and $\int_0^{\lambda t} e^{-v} dv = \frac{e^{-v}}{-1}\Big|_0^{\lambda t} = 1 - e^{-\lambda t}$.

Thus $P(L \geq k) = 1 - \sum_{i=0}^{k-1} \frac{(\lambda t)^i e^{-\lambda t}}{i!}$

But

$$P(L = k) = P(L \geq k) - P(L \geq k+1)$$

$$= \left[1 - \sum_{i=0}^{k-1} \frac{(\lambda t)^i e^{-\lambda t}}{i!} \right] - \left[1 - \sum_{i=0}^{k} \frac{(\lambda t)^i e^{-\lambda t}}{i!} \right]$$

$$P(L = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!} = \frac{\lambda^{*k} \cdot e^{-\lambda^*}}{k!}, \ \lambda^* = \lambda t$$

Thus the probability distribution of the discrete random variable L is a Poisson (discrete) distribution with parameter $\lambda^* = \lambda t$.

The cumulative distribution of the gamma distribution of Y can be calculated in terms of tabulated cumulative distribution of the Poisson distribution from

$$F(t) = P(L \geq k) = P(Y \leq t) = 1 - \sum_{i=0}^{k-1} \frac{(\lambda t)^i e^{-\lambda t}}{i!}$$

WORKED OUT EXAMPLES

Example 1: The daily consumption of electric power (in millions if kW-hours) in a certain city is a random variable X having the probability density

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

Find the probability that the power supply is inadequate on any given day if the daily capacity of the power plant is 12 million kW hours.

Solution: Observe that this is gamma distribution with $r = 2$ and $\lambda = \frac{1}{3}$. The power supply is inadequate when $X > 12$. Now $P(X > 12) = 1 - F(12) = 1 - I_2\left(\frac{1}{3} 12\right) = 1 - I_2(4)$. Using table A23 to A28, we get

$$P(X > 12) = 1 - 0.90892 = 0.09108$$

Example 2: The lifetime X (in months) of a computer has a gamma distribution with mean 24 months and standard deviation 12 months. Find the probability that the computer will

(a) last between 12 and 24 months.