CHAPTER 3

Random Variables and Probability Distributions

The topic of random variables is fundamental to probability and statistics. These concepts are natural extensions of the ideas of variables and relative frequency distributions.

This chapter introduces the basic concepts and definitions of random variables (both discrete and continuous) and their mean and standard deviation, and probability distributions. Permutations and combinations are defined briefly. Discrete probability distributions (hypergeometric, binomial and Poisson distributions) and their mean and standard deviations are derived. Continuous probability distribution, namely, the normal distribution, its properties, mean and variance and the standard normal distribution are presented. Approximating probability is also presented. They include Binomial approximation to the hypergeometric, Poisson approximate to the binomial and normal approximation to the binomial and rule for distribution are presented.

3.1 RANDOM VARIABLES

This section introduces the important concept of a probability distribution, which gives the probability for each value of a variable that is determined by chance. This section also introduces procedures for finding the mean and standard deviation for a probability distribution. A *random variable* is a variable (typically represented by x) that has a single numerical value, determined by chance, for each outcome of a procedure. In other words, a *random variable* is a quantitative variable whose value depends on chance. In Chapter 1, we defined a variable as a characteristic that varies from one member of a population to another. When one or more members are selected at random from the population, the variable, in that context, is called a *random variable*. Therefore, mathematically, a random variable is a function defined on the outcome of the sample space.

A *probability distribution* is a description that gives the probability for each value of the random variable. It is often expressed in the format of a graph, table, or formula. A random variable is classified a discrete or continuous, depending upon the range of its values. (The range of a random variable is the set of values it can assume.)

A *discrete random variable* has either a finite number of values or a countable number of values, where *countable* refers to the fact that there might be infinitely many values, but they can be associated with a counting process. Thus, we often say that a discrete random variable can assume at most a *countable infinite* number of values.

The number of eggs that a hen lays in a day, the salary of a new college graduate, the number of bacteria in a culture, the count of the number of students present in statistics class on a given day, the number of cavities that a child has, the price of gold on the exchange on a particular day, toss of a coin, the number of cars sold at a car dealership in a given month, color of a ball drawn from a collection of balls, etc., these are all examples of discrete random variables.

A *continuous random variable* has infinitely many values, and those values can be associated with measurements on a continuous scale without gaps or interruptions.

The maximum daily temperature, the amount of milk a cow produces in one day, the life of an electric bulb, the measure of voltage for a particular smoke detector battery, the distance by which a sharp shooter misses a target, the height of a person, the weight of a person, and the length of time a person has to wait at a bank counter, diameter of a manufactured shaft, time to failure of a machine component, repair time, the price of an automobile, the time taken to complete a medical examination, duration of a snow storm, etc. provide examples of continuous random variables.

3.1.1 Discrete Random Variables

Recall from Chapter 1 that the relative-frequency distribution or relative-frequency histogram of a discrete variable gives the possible values of the variable and the proportion of times each value occurs. We can extend the actions of relative-frequency distribution and relative-frequency histogram-concepts applying to variables of finite populations – to any discrete random variable. Here, we use the terms *probability distribution* and *probability histogram*.

The probability distribution of a discrete random variable, assuming a finite number of values, can be described by listing all the values that the random variable can assume, together with the corresponding probabilities. Such a listing is called the *probability function* of the random variable.

Probability histogram is a graph of the probability distribution that displays the possible values of a discrete random variable on the horizontal axis and the probabilities of those values on the vertical axis. The probability of each value is represented by a vertical bar whose height equals the probability.

In general, if a random variable X assumes the values $x_1, x_2, \ldots, x_k$, then we can represent the probability that X takes the values x_i by p_i. That is

$$P(X = x_i) = p\,(x_i)$$

The probability function can then be summarised in the form of a table, as shown in Table 3.1. By definition, a probability function gives the probabilities with which the values are assumed by the random variable. Hence, $0 \le p\,(x_i) \le 1$.

Table 3.1: The probability function of a random variable

Value x	Probability p(x)
x_1	$p(x_1)$
x_2	$p(x_2)$
$\vdots$	$\vdots$
x_k	$p(x_k)$
Sum	$\Sigma p(x_i) = 1$

Similarly, a random variable has to assume one of its possible values. Therefore, $\Sigma p(x_i) = 1$.

Properties of a Probability Function:

1. The probability that a random variable assumes a value x_i is always between 0 and 1.
   Hence, $0 \le p\,(x_i) \le 1$ (3.1)
2. The sum of all probabilities $p(x_i)$ is equal to 1.
   Hence, $\Sigma\, p\,(x_i) = 1$ (3.2)

Let X be a random variable. If the number of possible values of X is finite or countably infinite, X is called a *discrete random variable*. Each possible outcome is x_i. A number, $p(x_i) = p(X = x_i)$, gives the probability that the random variable X equals a value, x_i. The numbers, $p(x_i)$, for $i = 1, 2, \ldots$, must satisfy the following conditions:

1. $p(x_i) \ge 0$ for all i
2. $\sum_{i=1}^{\infty} p(x_i) = 1$ (3.3)
3. $0 \le p(x_i) \le 1$

The collection of pairs $[x_i, p(x_i)]$, $i = 1, 2, \ldots$ is called the *probability distribution* of X and $p(x_i)$ is called the *probability mass function* (or pmf) of X. The probability distribution for the variable X is characterised in Fig. 3.1. The *cumulative distribution function* of a discrete random variable is defined as

$$p(X = x_i) = \sum_{x_i \le X} p(x_i) \tag{3.4}$$

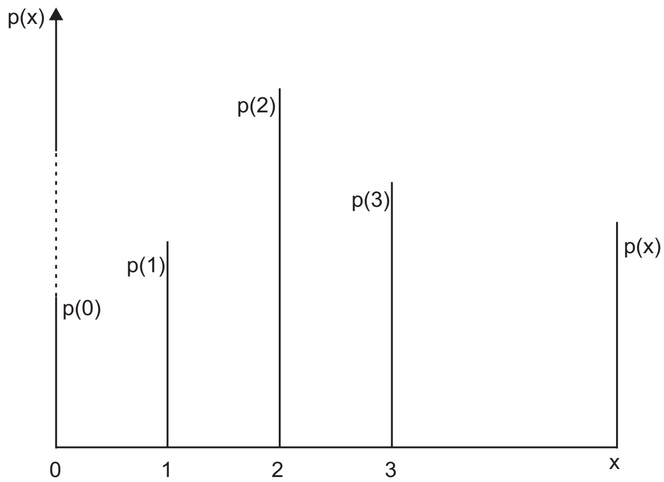


Fig. 3.1: Probability distribution of a discrete random variable x_i

The function is a step function that is constant over every interval not containing any of the points x_i, as shown in Fig. 3.2.

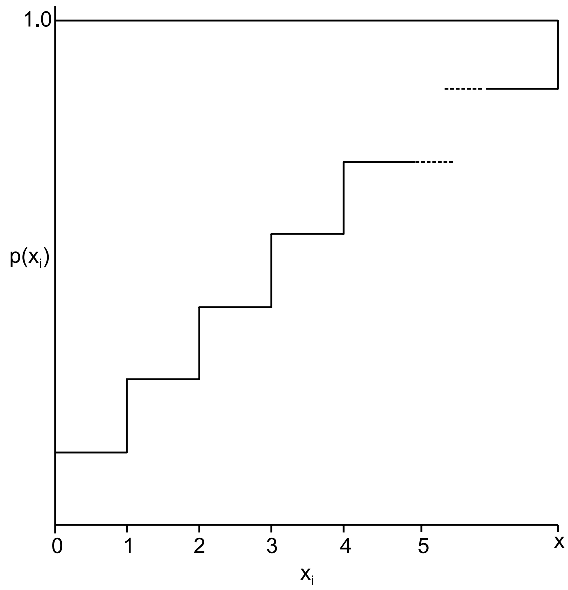


Fig. 3.2: Cumulative distribution function (CDF) of a discrete random variable x_i

3.1.2 Mean and Standard Deviation of a Discrete Random Variable

Consider a set of values, say $[x_i]$, $i = 1, 2, \ldots, n$, each of which occurs with frequency f_i for $i = 1, 2, \ldots, k$. The mean value is the sum of all the x's, hence $\sum_{i=1}^{k} x_i \cdot f_i$ divided by the total number of x's, $\sum_{i=1}^{k} f_i = n$, or

$$\frac{\sum_{i=1}^{k} x_i . f_i}{\sum_{i=1}^{k} f_i} = \sum_{i=1}^{k} \frac{x_i . f_i}{n} \tag{3.5}$$

But note that this could be written as

$$\frac{\sum_{i=1}^{k} x_i . f_i}{n} = \sum_{i=1}^{k} x_i . \frac{f_i}{n} \tag{3.6}$$

The relative frequencies f_i/n could be regarded as probabilities or values of $f(x)$. Hence,

If a random variable X has the discrete probability function $f(x)$, we define its *mean* or *expected value* or *expectation* as

$$E(X) = \mu = \Sigma x \cdot f(x) \text{ or } \Sigma x \cdot p(x) \tag{3.7}$$

where the summation extends over all the values of x and $p(x)$ are the probabilities.

The *variance* of a random variable X with probability distribution $f(x)$ is given by

$$\sigma^2 = \Sigma(x - \mu)^2 \cdot f(x) \tag{3.8}$$

where the summation extends over all the values of x.

The positive square root of σ^2, σ, is called the *standard deviation*.

We often denote the variance of X by $V(X)$.

It is often easier to use the following formula for σ^2:

$$\sigma^2 = \sum_x x^2 f(x) - \mu^2 \text{ or } \sum_x x^2 p(x) - \mu^2 \tag{3.9}$$

The variance of a random variable X is a measure of the dispersion or scatter in the possible values for X. The variance of X, denoted as σ^2 or $V(X)$ is

$$\sigma^2 = V(X) = \sum_x (x-\mu)^2 f(x) \tag{3.10}$$

$V(X)$ uses weight $f(x)$ as the multiplier of each possible squared deviation $(x - \mu)^2$.

Properties of summations and the definition of μ can be used to show that

$$V(X) = \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x)$$

$$= \sum_x x^2 f(x) - 2\mu^2 + \mu^2 = \sum_x x^2 f(x) - \mu^2 \tag{3.11}$$

Therefore, an alternative formula for $V(X)$ can be used.

Summarising, we have

The *mean* or *expected value* of the discrete random variable X, denoted as μ or $E(x)$ is

$$\mu = E(X) = \sum_x x f(x) \tag{3.12}$$

The *variance* of X, denoted as σ^2 or $V(X)$ is

$$\sigma^2 = V(X) = E(X-\mu)^2 = \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2 \tag{3.13}$$

The *standard deviation* of X is

$$\sigma = \sqrt{V(X)} \tag{3.13a}$$

Example E3.1

The number of defects in a gas turbine blade manufactured by a certain process varies from blade to blade. Overall, 50% of the blades produced have no defects, 40% have 1 defect, 8% have 2 defects and 2% have 3 defects. Let x be the number of defects in a randomly selected turbine blade manufactured by the process.

(*a*) find $p(x = 0)$, $p(x = 1)$, $p(x = 2)$ and $p(x = 3)$

(*b*) plot the probability mass function of the random variable x

(*c*) find $F(2)$ and $F(1.5)$

(*d*) plot the cumulative distribution function $F(x)$ of the random variable x that represent the number of defects in a randomly chosen turbine blade

(*e*) population mean of the sample random variable, x

SOLUTION:

(a) $P(x = 0) = 0.50$, $P(x = 1) = 0.40$, $P(x = 2) = 0.08$ and $P(x = 3) = 0.02$

(b) The plot of the probability mass function of the random variable, x is shown in Fig. E3.1(*a*).

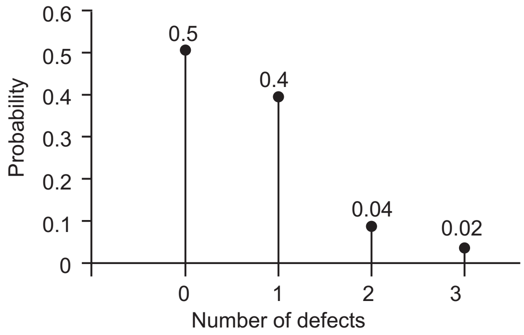


Fig. E3.1(a)

(c) Since $F(2) = P(x \leq 2)$, we need to find $P(x \leq 2)$. We can do this by summing the probabilities for the value of x that are less than or equal to 2, namely 0, 1, and 2. Hence

$F(2) = P(x \leq 2) = P(x = 0) + P(x = 1) + (x = 2) = 0.50 + 0.40 + 0.08 = 0.98$

Now $F(1.5) = P(x \leq 1.5)$. Hence, to compute $F(1.5)$, we must sum the probability for the values of x that are less than or equal to 1.5, which are 0 and 1. Therefore,

$F(1.5) = P(x \leq 1.5) = p(x = 0) + P(x = 1) = 0.50 + 0.40 = 0.90$

(d) First, we find $F(x)$ for each of the possible values of x, which are 0, 1, 2, and 3.

$F(0) = P(x \leq 0) = 0.50$

$F(1) = P(x \leq 1) = 0.50 + 0.40 = 0.90$

$F(2) = P(x \leq 2) = 0.50 + 0.40 + 0.08 = 0.98$

$F(3) = P(x \leq 3) = 0.50 + 0.40 + 0.08 + 0.02 = 1$

The plot of $F(x)$ is presented in Fig. E3.1(*b*).

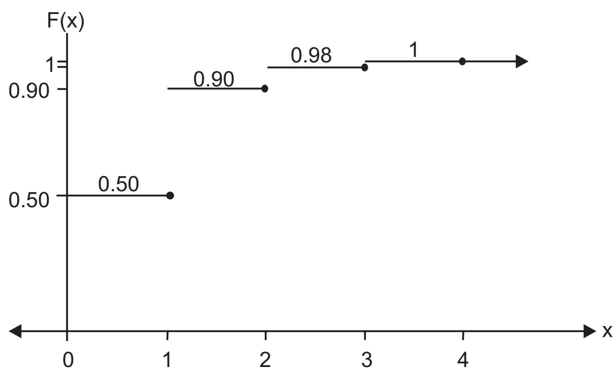


Fig. E3.1(b)

(*e*) The sample mean is the total number of defects divided by 100.

$$\text{mean} = \frac{0(50)+1(40)+2(8)+3(2)}{100} = 0.62$$

This can be written as

mean = 0(0.50) + 1(0.40) + 2(0.08) + 3(0.02) = 0.62

The above calculation for the mean shows that the mean of a perfect sample can be obtained by multiplying each possible value of *x* by its probability, and summing the products, which is the definition of the population mean of a discrete random variable.

Example E3.2

Table E3.2 gives the probability distribution of a discrete random variable *x*.

Table E3.2

x	0	1	2	3	4	5
P(x)	0.03	0.13	0.22	0.30	0.20	0.12

Find the following probabilities:

(*a*) $P(x = 1)$, $P(x \leq 1)$, $P(x \leq 3)$ and $P(0 \leq x \leq 2)$

(*b*) probability that *x* assumes a value less than 3

(*c*) probability that *x* assumes a value greater than 3

(*d*) probability that *x* assumes a value in the interval 2 and 4.

Solution:

(*a*) $P(x = 1) = 0.13$

$P(x \leq 1) = P(0) + P(1) = 0.03 + 0.13 = 0.16$

$P(x \leq 3) = P(3) + P(4) + P(5) = 0.30 + 0.20 + 0.12 = 0.62$

$P(0 \leq x \geq 2) = P(0) + P(1) + P(2) = 0.03 + 0.13 + 0.22 = 0.38$

(*b*) $P(x < 3) = P(0) + P(1) + P(2) = 0.03 + 0.13 + 0.22 = 0.38$

(*c*) $P(x > 3) = P(4) + P(5) = 0.20 + 0.12 = 0.32$

(*d*) $P(2 \leq x \leq 4) = P(2) + P(3) + P(4) = 0.22 + 0.30 + 0.20 = 0.72$

Example E3.3

Table E3.3 lists the probability distribution of *x*, where *x* is the number of car accidents that occur in a city during a week.

Table E3.3

x	0	1	2	3	4	5	6
P(x)	0.12	0.16	0.22	0.17	0.15	0.12	0.06

(*a*) draw a line graph for the probability distribution

(*b*) find the probability that the number of car accidents that will occur during a given week in the city is
(*i*) exactly 4 (*ii*) at least 3 (*iii*) less than 3 (*iv*) 3 to 5

SOLUTION:

(*a*) The line graph is shown in Fig. E3.3.

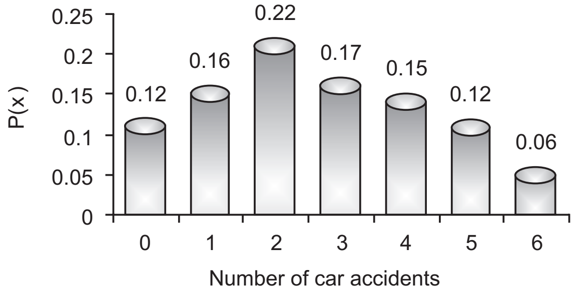


Fig. E3.3

(*b*) (*i*) $P(\text{exactly } 4) = P(4) = 0.15$

(*ii*) $P(\text{at least } 3) = P(x \geq 3) = P(3) + P(4) + P(5) + P(6) = 0.17 + 0.15 + 0.12 + 0.06 = 0.50$

(*iii*) $P(\text{less than } 3) = P(x < 3) = P(0) + P(1) + P(2) = 0.12 + 0.16 + 0.22 = 0.50$

(*iv*) $P(3 \text{ to } 5) = P(3) + P(4) + P(5) = 0.17 + 0.15 + 0.12 = 0.44$

Example E3.4

According to a survey, 75% of all elementary school teachers in a city in the year 2005 were women. Assume that this result holds true for the current population of all teachers in that city. Suppose 2 teachers are randomly selected from the population of all teachers in that city. Denoting x as the number of women in that sample, construct the probability distribution table of x. Draw a tree diagram for this problem.

SOLUTION:

Let M = teacher selected is a man

W = teacher selected is a woman

Then $P(W) = 0.75$ and $P(M) = 1 - P(W) = 1 - 0.75 = 0.25$

The tree diagram is shown in Fig. E3.4.

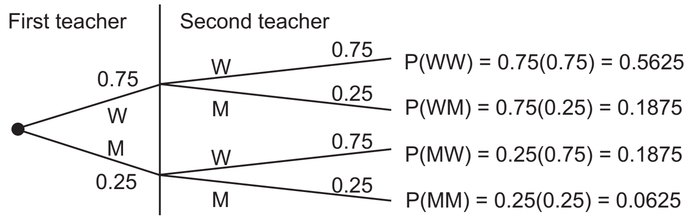


Fig. E3.4

Now, x = the number of women in a sample of 2 teachers. Table E3.4 lists the probability distribution of x, where $x = 0$ represents if neither teacher is a woman, $x = 1$ if the teacher is a woman and one is a man, and

$x = 2$ if the both teachers are woman. The probability in the table are obtained from the tree diagram in Fig. E3.4. The probability of $x = 1$ is obtained by adding the probabilities of *WM* and *MW*.

Table E3.4

Outcomes	x	P(x)
MM	0	0.0625
WM or MW	1	0.3750
WW	2	0.5625

Example E3.5

A machine design class has 20 students and 11 of them are female. Suppose 2 students are randomly selected from the class and x denotes the number of females in the sample.

(*a*) draw a tree diagram

(*b*) find the probability distribution of x.

Assume here that the draws are made without replacement from a small population and the probabilities of outcomes do not remain constant for each draw.

Solution:

Let A = first student selected is a male

B = first student selected is a female

C = second student selected is a male

D = second student selected is a female

(*a*) The tree diagram is shown in Fig. E3.5.

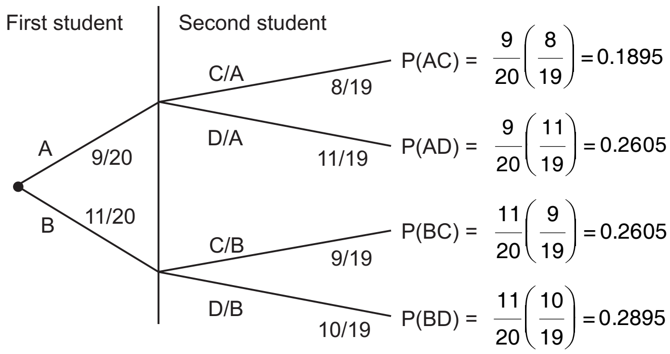


Fig. E3.5

(*b*) Table E3.5 lists the probability distribution of x, where x is the number of females in a sample of 2 students.

Table E3.5

Outcomes	x	P(x)
AC	0	0.1895
AD or BC	1	0.5210
BD	2	0.2895

Example E3.6

Table E3.6 lists the probability distribution of the number of breakdowns per week for a machine based on past data.

Table E3.6

Breakdowns per week	0	1	2	3
Probability	0.15	0.25	0.40	0.20

(a) present this probability distribution graphically

(b) find the probability that the number of breakdowns for this machine during a given week is

(i) exactly 2 *(ii)* 0 to 2 *(iii)* more than 1 *(iv)* at most 1.

SOLUTION:

Let x denotes the number of breakdowns for this machine during a given week. The Table E3.6 lists the probability distribution of x.

Table E3.6: Probability distribution of breakdowns

x	P(x)
0	0.15
1	0.25
2	0.40
3	0.20
	$\Sigma P(x) = 1.00$

(a) Figure E3.6 shows the line graph of the probability distribution.

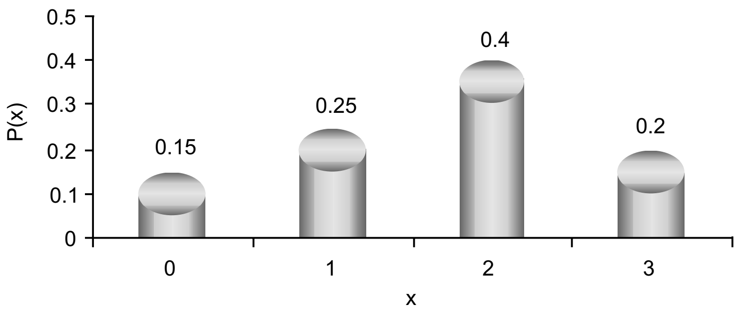


Fig. E3.6: Graphical representation of the probability distribution

(b) *(i)* The probability of exactly, two breakdowns is

$P(\text{exactly 2 breakdowns}) = P(x = 2) = 0.40$

(ii) The probability of 0 to 2 breakdowns is given by the sum of the probabilities of 0, 1 and 2 breakdowns

$P(\text{0 to 2 breakdowns}) = P(0 \le x \le 2)$

$= P(x = 0) + P(x = 1) + P(x = 2) = 0.15 + 0.25 + 0.40 = 0.80$

(iii) The probability of more than 1 breakdown is obtained by adding the probabilities of 2 and 3 breakdowns

$P(\text{more than 1 breakdown}) = P(x > 1)$

$= P(x = 2) + P(x = 3) = 0.40 + 0.20 = 0.60$

(iv) The probability of at most 1 breakdown is given by the sum of the probabilities of 0 and 1 breakdowns.

$P(\text{at most 1 breakdown}) = P(x \le 1)$

$$= P(x = 0) + P(x = 1) = 0.15 + 0.25 = 0.40$$

Example E3.7

Table E3.7 lists the probability distribution of x, where x is the number of defects contained in a randomly selected manufactured product. Find the mean and standard deviation of x.

Table E3.7

x	0	1	2	3	4
P(x)	0.41	0.12	0.05	0.03	0.01

Solution:

Refer to Table E3.7(a)

Table E3.7(a)

x	P(x)	x P(x)	x^2 P(x)
0	0.41	0	0
1	0.12	0.12	0.12
2	0.05	0.10	0.20
3	0.03	0.09	0.27
4	0.01	0.04	0.16
Total		0.35	0.75

Mean $\mu = \Sigma x P(x) = 0.35$ defects (from Table E3.7(a)).

Standard deviation $\sigma = \sqrt{x^2 P(x) - \mu^2} = \sqrt{0.75 - 0.35^2} = 0.79$ defects.

Example E3.8

Find the mean, the variance, and the standard deviation of x, where x denotes the number that shows up when a fair die is rolled.

Solution:

The probability function of x and the necessary calculations are summarised in Table E3.8.

Table E3.8

x	p(x)	x p(x)	x^2p(x)
1	1/6	1/6	1/6
2	1/6	2/6	4/6
3	1/6	3/6	9/6
4	1/6	4/6	16/6
5	1/6	5/6	25/6
6	1/6	6/6	36/6
Σ	1	21/6 = 3.5	91/6 = 15.1667

The mean given by

$$\mu = \Sigma x p(x) = \frac{21}{6} = 3.5$$

Variance is given by

$$\sigma^2 = \Sigma x^2 p(x) - \mu^2 = \frac{91}{6} - \left(\frac{21}{6}\right)^2 = 2.92$$

The standard deviation = $\sqrt{\sigma^2} = \sqrt{2.92} = 1.71$

Hence, mean = 3.5, variance = 2.92 and standard deviation = 1.71.

3.1.3 Continuous Random Variables

A random variable is *continuous* if its probabilities are given by areas under a curve. The curve is called a *probability density function* for the random variable. The probability density function is sometimes called the *probability distribution*.

If the range space of the random variable X is an interval or a collection of intervals, X is called a *continuous random variable*. If a and b are real numbers and $a < b$, then the probability that X lies in the interval $[a, b]$ is defined as

$$P[a \le X \le b] = \int_a^b f(x)\, dx \tag{3.14}$$

The function $f(x)$ is known as the *probability density function* (or pdf) of the random variable X. The pdf must satisfy the following conditions:

1. $f(x) \ge 0, \qquad -\infty < x < \infty$
2. $\int f(x)\, dx = 1$ (3.15)

The probability density function (pdf) is shown in Fig. 3.3. The shaded area in Fig. 3.3 represents the probability that X lies in the interval $[a, b]$.

From Eq. (3.14), for any specified value of x, $P(X = x) = 0$, since

$$\int_x^x f(t)\, dt = 0 \tag{3.16}$$

Hence, we can also write

$$P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b) \tag{3.17}$$

Since x is continuous

$$F(x) = \int_{-\infty}^{x} f(t)\, dt \tag{3.18}$$

The following conditions are satisfied

1. F is a non-decreasing function. If $a < b$, then $F(a) \le F(b)$.
2. $\lim_{x \to \infty} F(x) = 1$

3. $\lim_{x \to -\infty} F(x) = 0$ (3.19)

All probability questions about X can be expressed in terms of the cdf. For instance

$$P(a < X \le b) = F(b) - F(a) \text{ for } a < b \tag{3.20}$$

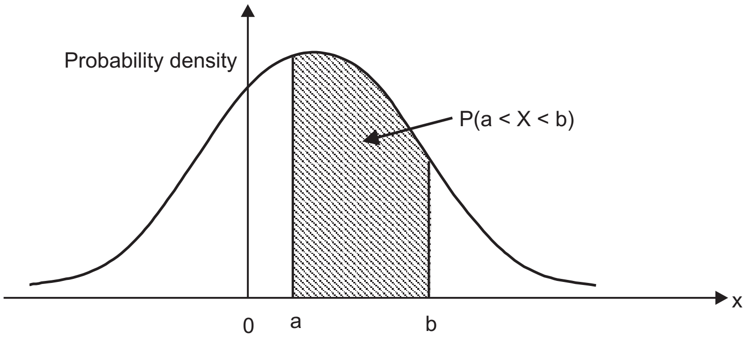


Fig. 3.3: The probability density function (pdf)
(The shaded area under the curve is P(a < X < b))

The cumulative distribution function of a continuous random variable X is $F(X) = P(X \le x)$, just like for a discrete random variable. For a continuous variable, the value of $F(x)$ is obtained by integrating the probability density function.

Since $F(x) = P(X \le x)$, we have

$$F(x) = \int_{-\infty}^{x} f(t)\, dt \tag{3.21}$$

where $f(t)$ is the probability density function.

Hence, the continuous distribution function (cdf) of X is the function

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt \tag{3.22}$$

3.1.4 Mean and Variance for Continuous Random Variables

The population mean and variance of a continuous random variable are defined in the same way as those of a discrete random variable, except that the probability density function is used instead of the probability mass function. Specifically, if X is a continuous random variable, its population mean is defined to be the center of mass of its probability density function and its population variance is the moment of inertia around a vertical axis through the population mean.

Let X be a continuous random variable with probability density function $f(x)$. Then, the mean is given by

$$\mu_x = \int_{-\infty}^{\infty} x\, f(x)\, dx \tag{3.23}$$

The *mean* of X is sometimes called the *expectation* or the *expected* value of X and denoted by $E(X)$ or by μ.

The variance of X is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x-\mu_x)^2 f(x)\, dx \tag{3.24}$$

An alternative formula for the variance is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu_X^2 \tag{3.25}$$

The variance of X is also denoted by $V(X)$ or by σ^2.

The standard deviation is the square root of the variance

$$\sigma_X = \sqrt{\sigma_X^2} \tag{3.26}$$

Alternate Formula for the Variance

Starting with the equation

$$\sigma_X^2 = \sum_x (x-\mu_x)^2 P(X = x) \tag{3.27}$$

Expanding the above equation, we obtain

$$\sigma_X^2 = \sum_x (x^2 - 2x\mu_X + \mu_X^2)\, P(X = x) \tag{3.28}$$

or

$$\sigma_X^2 = \sum_x [x^2 P(X = x) - 2x\mu_X P(X = x) + \mu_X^2 P(X = x)] \tag{3.29}$$

Summing the terms separately, we have

$$\sigma_X^2 = \sum_x x^2 P(X = x) - \sum_x 2x\mu_X P(X = x) + \sum_x \mu_X^2 P(X = x) \tag{3.30}$$

Noting that

$$\sum_x 2x\mu_X P(X = x) = 2\mu_X \sum_x xP(X = x) = 2\mu_X \mu_X = 2\mu_X^2 \tag{3.31}$$

and

$$\sum_x \mu_X^2 P(X = x) = \mu_X^2 \sum_x P(X = x) = \mu_X^2 (1) = \mu_X^2 \tag{3.32}$$

Substituting Eqs. (3.31) and (3.32) in Eq. (3.30), we get

$$\sigma_X^2 = \sum_x x^2 P(X = x) - 2\mu_X^2 + \mu_X^2 \tag{3.33}$$

or

$$\sigma_X^2 = \sum_x x^2 P(X = x) - \mu_X^2 \tag{3.34}$$

3.1.5 Expectation

A brief summary is given here:

If X is a random variable, the expected value of X, denoted by $E(X)$, for discrete and continuous variable is defined as

$$E(X) = \sum_{all\ i} x_i\, p(x_i) \qquad \text{if } X \text{ is discrete} \tag{3.35}$$

and
$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx \qquad \text{if } X \text{ is continuous} \tag{3.36}$$

the expected value $E(X)$ of a random variable X is also referred to as the *mean* of X, μ_X (or μ), or the *first moment* of X. The quantity, $E(X^n)$, $n \geq 1$, is called the n^{th} *moment* of X and is defined as

$$E(X'') = \sum_{all\ i} x_i^n\, p(x_i) \qquad \text{if } X \text{ is discrete} \tag{3.37}$$

and
$$E(X'') = \int_{-\infty}^{\infty} x^n f(x)\, dx \qquad \text{if } X \text{ is continuous} \tag{3.38}$$

The variance of a random variable, X, denoted by $V(X)$, Var(X), or σ^2, is defined by

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

A useful identity in computing $V(X)$ is given by

$$V(X) = E(X^2) - [E(X)]^2 \tag{3.39}$$

The mean $E(X)$ is a measure of the central tendency of a random variable. The variance, $V(X)$, is a measure of the spread or dispersion of the possible values of X around the mean $E(X)$. The standard deviation, σ, is the square root of the variance.

Example E3.9

The life of an electronic component is given by X, a continuous random variable assuming all values in the range $x \geq 0$. The pdf of the lifetime, in years is given as

$$f(x) = \begin{cases} \dfrac{1}{4} e^{-\frac{x}{4}}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

(*a*) determine the probability that the life of this component is between 2 and 4 years.
(*b*) determine cdf of the components
(*c*) determine the probability that this electronic component will last for 4 years
(*d*) find the mean and variance of the life of the component.

Solution:

(*a*) $$P(2 \leq X \leq 4) = \frac{1}{4}\int_2^4 e^{-\frac{x}{4}} dx = -e^{-1} + e^{-\frac{1}{2}} = 0.2386$$

The probability that the laser will last between 2 and 4 years is obtained also from

$P(2 \leq X \leq 4) = F(4) - F(2) = (1 - e^{-4/4}) - (1 - e^{-2/4}) = e^{-1/2} - e^{-1} = 0.6065 - 0.3679 = 0.2386$

(b) The cdf of this component is given by

$$F(x) = \frac{1}{4}\int_0^x e^{-\frac{x}{4}} dx = 1 - e^{-\frac{x}{4}}$$

(c) The probability that the component will last for 4 years or less is given by

$$P(0 \le X \le 4) = F(4) - F(0) = F(4) = 1 - e^{4/4} = 0.6321$$

(d) The mean and variance of the life of the electronic component are determined as follows:

$$E(X) = \frac{1}{4}\int_0^\infty xe^{-\frac{x}{4}} dx = xe^{-\frac{x}{4}}\Big|_0^\infty + \int_0^\infty e^{-\frac{x}{4}} dx = 0 + \frac{1}{1/4} e^{-\frac{x}{4}}\Big|_0^\infty = 4 \text{ years}$$

To compute $V(X)$, first compute $E(X^2)$ as follows:

$$E(X^2) = \frac{1}{4}\int_0^\infty x^2 e^{-\frac{x}{4}} dx = x^2 e^{-\frac{x}{4}}\Big|_0^\infty + 8\int_0^\infty xe^{-\frac{x}{4}} dx = 32$$

Hence, variance is given by

$$V(X) = 32 - 4^2 = 16 \text{ years}$$

Example E3.10

Given $f(x) = e^{-x}$ for $0 < x$. Find the following probabilities:

(a) $P(1 < X)$

(b) $P(1 < X < 2.5)$

(c) $P(X = 3)$

(d) $P(X < 4)$

(e) $P(3 \le X)$

Solution:

(a) $$P(1 < X) = \int_1^\infty e^{-x} dx = (-e^{-x})\Big|_1^\infty = e^{-1} = 0.3679$$

(b) $$P(1 < X < 2.5) = \int_1^{2.5} e^{-x} dx = (-e^{-x})\Big|_1^{2.5} = e^{-1} - e^{-2.5} = 0.2858$$

(c) $$P(X = 3) = \int_3^3 e^{-x} dx = 0$$

(d) $$P(X < 4) = \int_0^4 e^{-x} dx = (-e^{-x})\Big|_0^4 = 1 - e^{-4} = 0.9817$$

(e) $P(3 \le X) = \int_3^{\infty} e^{-x} dx = (-e^{-x})\Big|_3^{\infty} = e^{-3} = 0.0498$

Example E3.11

The probability density of the continuous random variable X is given by

$$f(x) = \left.\begin{matrix} 1/5 \\ 0 \end{matrix}\right\}; \quad \begin{matrix} \text{for } 2 < x < 7 \\ \text{elsewhere} \end{matrix}$$

Find (a) $P(3 < X < 5)$

(b) area under the curve.

Solution:

(a) $P(3 < X < 5) = \int_3^5 \frac{1}{5}\, dx = \frac{1}{5}\int_3^5 1 dx = \frac{1}{5}(5-3) = \frac{2}{5}$

(b) Area under the curve $= \int_{-\infty}^{\infty} f(x)\, dx = \int_2^7 \frac{1}{5}\, dx = \frac{1}{5}x\Big|_2^7 = \frac{1}{5}(7-2) = 1$

Example E3.12

Given $f(x) = 1.5x^2$ for $-1 < x < 1$. Find the mean and variance of x.

Solution:

$$E(X) = \int_{-1}^{1} x(1.5x^2)\, dx = \int_{-1}^{1} 1.5x^3 dx = 1.5\frac{x^4}{4}\Big|_{-1}^{1} = 0$$

$$V(X) = \int_{-1}^{1} 1.5x^2(x-0)^2\, dx = 1.5\int_{-1}^{1} x^4 dx = 1.5\frac{x^5}{5}\Big|_{-1}^{1} = 0.6$$

Example E3.13

The pdf of the random variable X is given by

$$f(x) = \left.\begin{matrix} \dfrac{a}{\sqrt{x}} \\ 0 \end{matrix}\right\}; \quad \begin{matrix} \text{for } 0 < x < 4 \\ \text{elsewhere} \end{matrix}$$

Find (a) the value of a

(b) $P\left(X < \frac{1}{4}\right)$ and $P(X > 1)$.

SOLUTION:

(a) Area $= 1 = \int_0^4 \frac{a}{\sqrt{x}}\,dx = a\int_0^4 x^{-1/2}dx = a\left.\frac{x^{1/2}}{1/2}\right|_0^4 = 2a{\cdot}2 = 4a$

Hence, $a = 1/4$.

(b) $P\left(X < \frac{1}{4}\right) = \int_0^{1/4} \frac{1}{4\sqrt{x}}\,dx = \frac{1}{4}\int_0^{1/4} x^{-1/2}dx = \left.\frac{1}{4}\frac{\sqrt{x}}{1/2}\right|_0^{1/4} = \frac{1}{2}.\frac{1}{2} = \frac{1}{4}$

$P(X > 1) = 1 - \int_0^1 \frac{1}{4\sqrt{x}}\,dx = 1 - \left.\frac{1}{2}\sqrt{x}\right|_0^1 = \frac{1}{2}$

Example E3.14

The thickness of a conductive coating in micrometers has a density function of $f(x) = 600\,x^{-2}$ for 100 μm < x < 120 μm.

(a) find the mean and variance of the coating thickness

(b) if the cost of coating is $0.25 per micrometer of thickness on each component, find the average cost of the coating per component.

SOLUTION:

(a) $E(X) = \int_{100}^{120} x\frac{600}{x^2}\,dx = 600\ln x\Big|_{100}^{120} = 109.39$

$$V(X) = \int_{100}^{120} (x - 109.39)^2 \frac{600}{x^2}\,dx = 600\int_{100}^{120} 1 - \frac{2(109.39)}{x} + \frac{(109.39)^2}{x^2}\,dx$$

$$= 600(x - 218.75\ln x - 109.39^2 x^{-1})\Big|_{100}^{120} = 33.19$$

(b) Average cost per component = $0.25(109.39) = $27.35.

Example E3.15

The distribution function of the random variable X is given by

$$F(x) = \begin{cases} 1-(1+x)e^{-x} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

Find (a) $P(X \le 2)$

(b) $P(1 < X < 3)$

(c) $P(X > 4)$

SOLUTION:

(a) $P(X \le 2) = F(2) = 1 - 3e^{-2} = 1 - 3(0.1353) = 1 - 0.4074 = 0.5926$

(b) $P(1 < X < 3) = F(3) - F(1) = 1 - 4e^{-3} - 1 + 2e^{-1} = 2e^{-1} - 4e^{-3} = 2(0.3679) - 4(0.0498) = 0.5366$

(c) $P(X > 4) = 1 - F(4) = 5e^{-4} = 5(0.0183) = 0.0915$

Example E3.16

The pdf of a random variable X is given by

$$f(x) = \begin{cases} \frac{4}{3}(1-x^3) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Sketch the pdf and find the values of the following:

(*a*) $P\left(X < \frac{1}{2}\right)$

(*b*) $P\left(\frac{1}{4} < X < \frac{3}{4}\right)$

(*c*) $P\left(X > \frac{1}{3}\right)$

SOLUTION:

The plot is shown in Fig. E3.16.

(*a*) $P\left(X < \frac{1}{2}\right) = \int_0^{1/2} 4(1-x^3)\frac{dx}{3} = 0.6458$

(*b*) $P\left(\frac{1}{4} < X < \frac{3}{4}\right) = \int_{1/4}^{3/4} 4(1-x^3)\frac{dx}{3} = 0.5625$

(*c*) $P\left(X > \frac{1}{3}\right) = \int_{1/3}^{1} 4(1-x^3)\frac{dx}{3} = 0.5597$

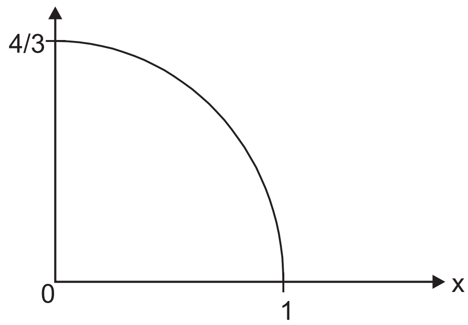


Fig. E3.16 pdf of the function f(x)

Example E3.17

The pdf of a random variable X is given by

$$f(x) = \begin{cases} ax^2 & \text{for } 0 \le x \le 2 \\ 0 & \text{elsewhere} \end{cases}$$

(*a*) find the value of the constant a and sketch the pdf

(*b*) find $P(X > 3/2)$.

SOLUTION:

(*a*) We must have

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_1^2 a\,x^2\,dx = \frac{7}{3}a = 1$$

Hence, $a = 3/7$. The pdf is shown in Fig. E3.17.

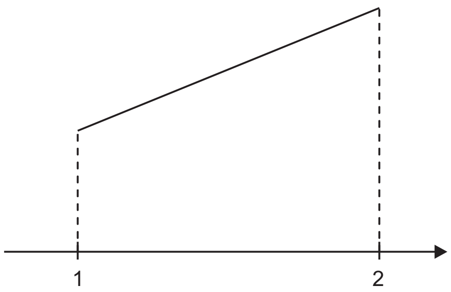


Fig. E3.17

(b) $P(X > 3/2) = \int_{3/2}^{2} f(x)\,dx = \int_{3/2}^{2} \frac{3}{7}x^2 dx = 37/56$

3.2 PERMUTATIONS AND COMBINATIONS

Factorial Notation: The product of the positive integers from 1 to n inclusive is denoted by $n!$ (read as "n factorial"). The value of the factorial of a number is obtained by multiplying all integers from that number to 1.

The symbol $n!$, read as "n factorial" represents the product of all integers from n to 1. In other words,

$$n! = n(n-1)(n-2)(n-3) \cdots 3{\cdot}2{\cdot}1 \tag{3.40}$$

It is also convenient to define

$$0! = 1,\ 1! = 1 \text{ and } n! = n{\cdot}((n-1)!) \tag{3.41}$$

Example E3.18

Evaluate the following:

(a) $2!, 3!, 4!, 5!, 6!$ and $7!$

(b) $(7-3)!, (14-12)!, (7-2)!, (13-4)!$, and $(6-6)!$

(c) $\frac{8!}{6!}, \frac{12!}{9!}$ and $\frac{12!}{9!3!}$

SOLUTION:

(a) $2! = 2{\cdot}1 = 2$

$3! = 3{\cdot}2{\cdot}1 = 6$

$4! = 4{\cdot}3{\cdot}2{\cdot}1 = 24$

$5! = 5{\cdot}4{\cdot}3{\cdot}2{\cdot}1 = 120$

$6! = 6{\cdot}5{\cdot}4{\cdot}3{\cdot}2{\cdot}1 = 720$

$7! = 7{\cdot}6{\cdot}5{\cdot}4{\cdot}3{\cdot}2{\cdot}1 = 5040$

(b) $(7-3)! = 4! = 4{\cdot}3{\cdot}2{\cdot}1 = 24$

$(14-12)! = 2! = 2.1 = 2$

$(7-2)! = 5! = 5{\cdot}4{\cdot}3{\cdot}2{\cdot}1 = 120$

$(13-4)! = 9! = 9{\cdot}8{\cdot}7{\cdot}6{\cdot}5{\cdot}4{\cdot}3{\cdot}.2{\cdot}1 = 362{,}880$

$(6-6)! = 0! = 1$

(c) $\dfrac{8!}{6!} = \dfrac{8{\cdot}7{\cdot}6!}{6!} = 8.7 = 56$

$$\frac{12!}{9!} = \frac{12{\cdot}11{\cdot}10{\cdot}9!}{9!} = 12{\cdot}11{\cdot}10 = 1320$$

$$\frac{12!}{9!3!} = \frac{12{\cdot}11{\cdot}10{\cdot}9!}{9!3!} = \frac{12{\cdot}11{\cdot}10}{3.2} = 220$$

Example E3.19

Find the value of 9! by using the table in Appendix-*A*.

SOLUTION:

We locate 9 in the column labeled n. Then we read the value in the column for $n!$ entered next to 9. Hence,

$9! = 362,\ 880$

3.2.1 Permutations

Any arrangement of a set of n objects in a given order is called a *permutation* of the objects (taken all at a time). Any arrangement of any $r \le n$ of those objects in a given order is called an *r-permutation* or *a permutation of the n objects taken r at a time*.

Ordered arrangement of a set of objects are called *permutations*. When the order in which they are arranged is disregarded, then the arrangements are known as *combinations*. The number of permutations of n different items is the number of different arrangements in which these items can be placed. A permutation of n different objects taken r at a time is an arrangement of r out of the n objects with attention given to the order of arrangement. The number of permutations of n objects taken r at a time is denoted by nP_r, $P(n, r)$ P_r^n, $(n)_r$ or $P_{n,r}$ and is written as

$$nP_r = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!} \tag{3.42}$$

Derivation of the Formula nP$_r$

The first element in an r-permutation of n objects can be chosen in n different ways. Following this way, the second element in the permutation can be chosen in $n-1$ ways, and the 3rd element in the permutation can be chosen in $n-2$ ways etc. Finally, the r^{th} (last) element in the r-permutation can be chosen in $n-(r-1) = n-r+1$ ways. Hence, from the fundamental principle of counting, we have

$$nP_r = P(n, r) = n(n-1)\,(n-2)\cdots(n-r+1) \tag{3.43}$$

or
$$n(n-1)\,(n-2)\cdots(n-r+1) = \frac{n(n-1)(n-2)\cdots(n-r+1).(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!} \tag{3.44}$$

in which $n!$ is the symbol for factorial $n = n(n-1)\cdots 3 \times 2 \times 1$, when $r = n$, $n\,P_n = n!$ (note by definition $0! = 1$). Tables of the values of factorial n are given in Appendix-*A*.

The permutation of n items taken all at a time, when the n items consists of r_1 alike, r_2 alike, … r_k alike, so that by definition $r_1 + r_2 + \cdots + r_k = n$.

The number of permutation is then

$$P = \frac{n!}{r_1!r_2!\cdots r_k!} \tag{3.45}$$

Also, we have $nP_n = n(n-1)\cdots 3{\cdot}2{\cdot}1 = n!$

Example E3.20

How many five-digit numbers can be formed from the numbers 1 to 8?

Solution:

Here $\quad n = 8$ and $r = 5$.

$$nP_r = \frac{n!}{(n-r)!} = \frac{8!}{(8-5)!} = \frac{8\times7\times6\times5\times4\times3\times2\times1}{3\times2\times1} = 1680 \text{ numbers.}$$

Example E3.21

How many numbers of permutations of letters are there in the word *quality*?

Solution:

Here $\quad n = 7$ (7 letters in word *quality*)

$r_1 = 1(q)$

$r_2 = 1(u)$

$r_3 = 1(a)$

$r_4 = 1(l)$

$r_5 = 1(i)$

$r_6 = 1(t)$

$r_7 = 1(y)$

Therefore,

$$P = \frac{n!}{r_1!r_2!\cdots r_{k!}} = \frac{7!}{1!1!1!\cdots1!} = 7\times6\times5\times4\times3\times2\times1 = 5040$$

Example E3.22

(*a*) Evaluate $7P_3$, $6P_4$, $16P_1$, $3P_3$ and $6P_3$

(*b*) five persons arrive at the bank teller window at the same time. In how many different ways can these people line up?

(*c*) find the number of ways of arranging the letters in the word *subject*

(*d*) find the number of ways if each arrangement in part *(c)* starts with the letter *j*.

SOLUTION:

(*a*) $7P_3 = 7{\cdot}6{\cdot}5 = 210$

$6P_4 = 6{\cdot}5{\cdot}4{\cdot}3 = 360$

$16P_1 = 16$

$3P_3 = 3{\cdot}2{\cdot}1 = 6$

$6P_3 = 6{\cdot}5{\cdot}4 = 120$

(*b*) These paper can line up in different ways

$5! = 5{\cdot}4{\cdot}3{\cdot}2{\cdot}1 = 120$ ways.

(*c*) There are 7 distinct letters in the word *subject*. Hence, the number of arrangements is 7!. That is

$7! = 7{\cdot}6{\cdot}5{\cdot}4{\cdot}3{\cdot}2{\cdot}1 = 5040$

(*d*) Since the first letter is fixed as *j*, the total number of possibilities are found by arranging the remaining six letters in all possible ways. This can be done by 6! ways.

$6! = 6{\cdot}5{\cdot}4{\cdot}3{\cdot}2{\cdot}1 = 720$ ways.

3.2.2 Combinations

As noted earlier, the order in which objects are arranged matters in *permutations*. However, order does not matter in *combinations*.

The number of combinations of n different objects taken r at a time is a selection of r out of the n objects with no attention given to the order of arrangement. The number of combinations of n objects taken r at a time is denoted by nC_r $C(n, r)$, $C_{n,r}$ or $\binom{n}{r}$ is $r!$ times smaller than the number of permutations.

Therefore,

$$nC_r = \binom{n}{r} = \frac{nP_r}{r!} = \frac{n!}{(n-r)!r!} \tag{3.46}$$

It should be noted from symmetry that the identity is valid

$$nC_r = nC_{n-r} \tag{3.47}$$

Also $\binom{n}{r} = \frac{nP_r}{r!}$

The arrangement for dividing the number of permutations by $r!$ to get the number of combinations is that each combination give rise to $r!$ possibilities when arranged in all possible ways, thereby giving all the permutations.

Tables of binomial coefficients $\left\{\frac{n}{r}\right\}$ are given in Appendix-*B*.

Note that in combinations, n is always greater than or equal to r. If n is smaller than r, then we cannot select r distinct elements from n.

Example E3.23

(a) Evaluate the following:

$$\binom{8}{4}, \binom{6}{6}, \binom{9}{3}, \binom{4}{0}, \binom{5}{3}, \binom{15}{2} \text{ and } \binom{20}{8}$$

(b) Four cards are chosen in succession from a deck of 52 cards. Find the number of ways this can be done *(i)* with replacement, *(ii)* without replacement.

(c) find the number m of groups of 3 that can be formed from 8 students in a class.

Solution:

(a)

$$\binom{8}{4} = \frac{8 \cdot 7 \cdot 6 \cdot 5}{4!} = \frac{8 \cdot 7 \cdot 6 \cdot 5}{4 \cdot 3 \cdot 2 \cdot 1} = 70$$

$$\binom{6}{6} = \frac{6!}{6!(6-6)!} = \frac{1}{0!} = \frac{1}{1} = 1$$

$$\binom{9}{3} = \frac{9!}{(9-3)!3!} = \frac{9!}{6!3!} = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 3 \cdot 2 \cdot 1} = 84$$

$$\binom{4}{0} = \frac{4!}{0!(4-0)!} = \frac{4!}{1 \cdot 4!} = 1$$

$$\binom{5}{3} = \frac{5!}{(5-3)!3!} = \frac{5!}{2!3!} = \frac{5.4.3.2.1}{2.1.3.2.1} = 10$$

$$\binom{15}{2} = \frac{15!}{(15-2)!2!} = \frac{15 \cdot 14 \cdot 13 \cdots 2 \cdot 1}{13 \cdot 12 \cdots 2 \cdot 1 \cdot 2 \cdot 1} = 105$$

$$\binom{20}{8} = \frac{20!}{(20-8)!8!} = 125,970$$

(b) *(i)* Since each card is replaced before the next card is chosen, each card can be chosen by 52 ways. Hence, there are

$(52)(52)(52)(52) = 52^4 = 7,311,616$

different ordered samples of size $n = 4$ with replacement.

(ii) Since there is no replacement, the first card can be chosen in 52 ways, the second card in 51 ways, 3rd card in 50 ways and the last card in 49 ways. Thus there are

$P(52, 4) = 52(51)(50)(49) = 6,497,400$

different ordered samples of size $r = 4$ without replacement.

(c)

$$m = C(8,3) = \binom{8}{3} = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} = 56$$

Since each group is essentially a combination of 8 students taken 3 at a time.

Example E3.24

The president of a university has 12 departments under his administration. The president plans to visit 3 of these departments during the next week. If he randomly selects 3 departments from these 12, how many total selections are possible? Verify the answer using the table in Appendix-*B*.

SOLUTION:

The total possible selections for selecting 3 departments from 12 are

$$\binom{12}{3} = \frac{12!}{(12-3)!3!} = 220$$

From the table in Appendix-*B*, we read for $n = 12$ and $r = 3$, the answer is 220.

Example E3.25

Out of 5 men and 7 women, a committee consisting of 2 men and 3 women is to be formed. In how many ways can this be done if

(*a*) any men and any women can be included
(*b*) one particular women must be included on the committee
(*c*) two particular men cannot be included on the committee.

SOLUTION:

(*a*) 2 men out of 5 can be selected in $5C_2$ ways.
3 women out of 7 can be selected in $7C_3$ ways.
Hence, the number of possible selections = $(5C_2)\ (7C_3) = (10)\ (35) = 350$

(*b*) 2 men out of 5 can be selected $5C_2$ ways.
2 additional women out of 6 can be selected in $6\ C_2$ ways.
Hence, the number of possible selections = $(5C_2)\ (6C_3) = (10)\ (15) = 150$

(*c*) 2 men out of 3 can be selected $3C_2$ ways.
3 women out of 7 can be selected in $7C_2$ ways.
Hence, the number of possible selections = $(3C_2)\ (7C_3) = (3)\ (35) = 105$

3.3 DISCRETE DISTRIBUTIONS

Discrete random variables are used to describe random phenomena in which only integer values can occur.

3.3.1 Hypergeometric Distribution

In this section we consider dependent Bernoulli random variables. A common source of dependent Bernoulli random variables is sampling without replacement from a finite population. Suppose that a finite population consists of a known number of successes and failures. If we sample a fixed number of units from that population, the number of successes in our sample will have a distribution that is a member of the family of hypergeometric distributions.

Definition of the Hypergeometric Distribution

A set of N objects contains K objects classified as successes and $N - K$ objects classified as failures. A sample of size n objects is selected randomly (without replacement) from the N objects, where $K \le N$ and $n \le N$.

Let the random variable X denotes the number of successes in the sample. Then X is a *hypergeometric random variable* and

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \qquad x = \max\{0, n + K - N\} \text{ to } \min\{K, n\} \tag{3.48}$$

The expression $\{K, n\}$ is used in the definition of the range of X because the maximum number of successes that can occur in the sample is the smaller of the sample size, n, and the number of successes available, K.

Also, if $n + K > N$, at least $n + K - N$ successes must occur in the sample.

It should be noted here that in Eq. (3.48)

$$\binom{a}{b} = \frac{a!}{b!(a-b)!} \tag{3.49}$$

is the number of a parts taken b at a time. The hypergeometric distribution is the appropriate possibility model for sampling without replacement. Other distributions are often employed when the ratio of n/N becomes small, say 0.10 or less.

The mean and variance of a hypergoemetric random variable can be determined from the trials that comprise the experiment. However, the trials are not independent, and so the calculations are more difficult than for a binomial distribution. The results are stated as follows:

Mean and Variance of a Hypergeometric Distribution

If X is a hypergoemetric random variable with parameters N, K, and n, then the mean

$$\mu = E(X) = np \tag{3.49a}$$

and the variance

$$\sigma^2 = V(X) = np(1-p)\left(\frac{N-n}{N-1}\right) \tag{3.49b}$$

where $$p = \frac{K}{N}\cdot \tag{3.49c}$$

Here, p is interpreted as the proportion of successes in the set of N objects. For a hypergeometric random variable, $E(X)$ is similar to the mean of a binomial random variable. Also, $V(X)$ differs from the result for a binomial random variable only by the term shown below:

The term in the variance of a hypergeometric random variable $\frac{N-n}{N-1}$ is called the *finite population correlation factor*.

Example E3.26

Given that X has a hypergeometric distribution with $N = 100$, $n = 4$ and $K = 20$. Determine the following:

(*a*) $P(X = 1)$

(*b*) $P(X = 6)$

(*c*) $P(X = 4)$

(*d*) the mean and variance of X.

Solution:

Here, $K = 20$, $X = 1$, $N = 100$ and $n = 4$.

(*a*) $$P(X=1) = \frac{\binom{K}{n}\binom{N-K}{n-X}}{\binom{N}{n}} = \frac{\binom{20}{4}\binom{80}{3}}{\binom{100}{4}} = \frac{20(82160)}{3921225} = 0.4191$$

(*b*) $P(X = 6) = 0$, since the sample size is only 4.

(*c*) $$P(X=4) = \frac{\binom{K}{n}\binom{N-K}{n-X}}{\binom{N}{n}} = \frac{\binom{20}{4}\binom{80}{0}}{\binom{100}{4}} = \frac{4845(1)}{3921225} = 0.001236$$

(*d*) $$\text{Mean} = E(X) = np = n\left(\frac{K}{N}\right) = 4\left(\frac{20}{100}\right) = 0.8$$

$$\text{Variance} = V(X) = np(1-p)\left(\frac{N-n}{N-1}\right) = 40(0.2)(0.8)\left(\frac{96}{99}\right) = 0.6206$$

Example E3.27

Suppose a box contains five red balls and ten blue balls. If seven balls are selected at random without replacement, find the probability that at least 3 red balls will be obtained.

Solution:

Let X denotes the number of red balls that are obtained. Then, X has a hypergeometric distribution, $D = 5$, $A + B = 15$, $n = 7$, $B = 10$ and $A = 5$. The maximum value of X is $\min\{n, A\} = 5$. Hence,

$$P(X \geq 3) = \sum_{x=3}^{5} \frac{\binom{5}{x}\binom{10}{7-x}}{\binom{15}{7}} = \frac{2745}{6435} = 0.4266$$

Example E3.28

A lot of 75 gaskets contains five in which the variability in thickness around the circumference of the gasket is unacceptable. A sample of 10 gaskets is selected at random, without replacement. What is the probability that

(a) none of the unacceptable gaskets is in the sample
(b) at least one unacceptable gasket is in the sample
(c) exactly one unacceptable gasket is in the sample
(d) the mean number of unacceptable gaskets in the sample.

Solution:

Let X denotes the number of unacceptable gaskets in the sample of 10.

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

(a) $$P(X=0) = \frac{\binom{5}{0}\binom{70}{10}}{\binom{75}{10}} = \frac{\dfrac{70!}{10!60!}}{\dfrac{75!}{10!65!}} = \frac{65\times64\times63\times62\times61}{75\times74\times73\times72\times71} = 0.4786$$

(b) $P(X \geq 1) = 1 - P(X=0) = 1 - 0.4786 = 0.5214$

(c) $$P(X=0) = \frac{\binom{5}{1}\binom{70}{9}}{\binom{75}{10}} = \frac{\dfrac{5!70!}{9!6!}}{\dfrac{75!}{10!65!}} = \frac{5\times65\times64\times63\times62\times10}{75\times74\times73\times72\times71} = 0.3923$$

(d) $$E(X) = 10\left(\frac{5}{75}\right) = \frac{2}{3}$$

Example E3.29

Of 50 manufactured steel rods in a production process by a company, 12 have defects. If 10 steel rods are selected at random for inspection,

(a) find the probability that exactly 3 of the 10 have defects
(b) find the mean and variance of X.

Solution:

(a) Let X represents the number of sampled steel rods that have defects. Then, $K = 12$, $N = 50$ and $n = 10$.

$$P(X=3)=\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}=\frac{\binom{12}{3}\binom{38}{7}}{\binom{50}{10}}=\frac{(220)(12,620,256)}{10,272,278,170}=0.2703$$

(b) Mean = $\mu_X = np = \dfrac{(10)(12)}{50} = 2.4$

where $p = \dfrac{12}{50}$

$$\text{Variance} = \sigma_X^2 = np(1-p)\left(\frac{N-n}{N-1}\right) = 10\left(\frac{12}{50}\right)\left(1-\frac{12}{50}\right)\left(\frac{50-10}{50-1}\right) = 1.4890$$

Example E3.30

A large bin contains 80 balls of which 32 are red balls and 48 are blue balls. Suppose 15 balls are picked at random. Find the probability of getting 4 red balls, the mean number of red balls, and the standard deviation of the number of red balls if the sample is picked

(a) with replacement

(b) without replacement.

SOLUTION:

We will identify success with "picking a red ball" and let X = number of red balls in the sample.

(a) If the sampling is with replacement, we have a binomial distribution with 15 trials and probability of success $p = \dfrac{32}{80} = 0.4$. Hence,

$P(X = 4) = \text{Binomial}(4;\ 15,\ 0.4) = 0.127$ (from binomial distribution in Appendix-C)

$\mu = np = 15(0.4)$

$\sigma = \sqrt{np(1-p)} = \sqrt{15(0.4)(0.6)} = 1.9$

(b) If the sampling is without replacement, we have a hypergeometric distribution and

$$P(X=4)=\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}=\frac{\binom{32}{4}\binom{48}{11}}{\binom{80}{15}}=0.122$$

Note that this answer is not much different from the one obtained in part *(a)* using the binomial formula. The reason is that N (= 80), compared to n (= 15), is fairly large.

Mean, $\mu = np = 15(0.4) = 6$

Standard deviation

$$\sigma = \sqrt{np(1-p)\left(\frac{N-n}{n-1}\right)} = \sqrt{15\left(\frac{32}{80}\right)\left(1-\frac{32}{80}\right)\left(\frac{80-15}{80-1}\right)} = 1.72$$

3.3.2 The Binomial Probability Distribution

The binomial probability distribution is one of the most widely used discrete probability distribution. It is used to find the probability that an outcome will occur x times in n performances of an experiment. The binomial distribution is applied to experiments that satisfy the four conditions of a *binomial experiment*. Each repetition of a binomial experiment is called a *trial* or a *Bernoulli trial* (after Jacob Bernoulli).

A trial with only two possible outcomes is used so frequently as a building block of a random experiment that is called a *Bernoulli trial*. It is usually assumed that the trials that constitute the random experiment are *independent*. This implies that the outcome from one trial has no effect on the outcome to be obtained from any other trial. In addition, it is often reasonable to assume that the *probability of a success in each trial is constant*.

3.3.3 The Binomial Experiment

An experiment that satisfies the following four conditions is called a *binomial experiment*.

1. There are n identical trials. In other words, the given experiment is repeated n times. All these repetitions are performed under identical conditions.
2. Each trial has two and only two outcomes. These outcomes are usually called a *success* and a *failure*.
3. The probability of success is denoted by p and that of failure by q, and $p + q = 1$. The probabilities p and q remain constant for each trial.
4. The trials are independent. In other words, the outcome of one trial does not affect the outcome of another trial.

One of the two outcomes of a trial is called a *success* and the other a *failure*.

The random variable X that equals the number of trials that result in a success has a *binomial random variable* with parameters $0 < p < 1$ and $n = 1, 2, \ldots$. The probability mass function (pmf) of X is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n \tag{3.50}$$

3.3.4 The Binomial Formula

3.3.4.1 Binomial Theorem

The n^{th} power of $(p + q)$ can be expressed in terms of binomial coefficients as

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

$$= \sum_{r=0}^{n} \binom{n}{r} p^{n-r} q^r = \sum_{r=0}^{n} {}_nC_r\, p^{n-r} q^r \tag{3.50a}$$

Since $(p + q) = (q + p)$, we can also write

$$(p+q)^n = \sum_{r=0}^{n} {}_nC_r\, p^r q^{n-r} \tag{3.51}$$

Consider n trials in each of which the probability of success is p. Then the probability of failure is $1 - p = q$. To find the probability of r successes, we observe that:

the probability of 1 success in 1 try is p,

the probability of 2 successes in 2 tries is $p \times p$ or p^2,

the probability of 3 successes in 3 tries is $p \times p \times p$ or p^3,

$\vdots$

the probability of r successes in r tries is p^r and

the probability of subsequent $(n - r)$ failures in $(n - r)$ tries is $(1 - p)^{n-r} = q^{n-r}$.

The probability of r successes followed by $(n - r)$ failures is $p^r(1 - p)^{n-r}$. Here, we have considered only one particular group or combination of r events; i.e., we have started with r successes and finished with $(n - r)$ failures; every other possible ordering of r successes and $(n - r)$ failures will also have the same probability.

The number of possible orderings or the number of selections for r successes and $(n - r)$ failures in n trials is $n!/[r!(n - r)!]$.

Therefore, the probability P_r of an event succeeding r times is

$$P_r = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \tag{3.52}$$

or

$$P_r = {}_nC_r\, p^r q^{n-r} \tag{3.53}$$

This term is similar to the r^{th} term of the binomial expansion $(q + p)^n$ [see Eq. (3.51)], which can be written

$$(q+p)^n = \sum_{r=0}^{n} {}_nC_r\, p^r q^{n-r} \tag{3.54}$$

The successive terms of the expansion give the probability P_r of an event succeeding r times in n trials for values of r varying in steps of one of 0 to n.

3.3.4.2 Cumulative Terms for Binomial Distribution

The probability of an event succeeding at least r times in n trials.

The probability P_r of an event succeeding exactly r times in n trials is given by Eq. (3.53)

$$P_r = {}_nC_r p^r q^{n-r}$$

The probability of an event succeeding at least r' times in n trials is given by

$$\sum_{r=r'}^{r=n} P_r = \sum_{r=r'}^{r=n} {}_nC_r\, p^r q^{n-r} \tag{3.55}$$

The values of the summation of Eq. (3.55) are given in Appendix-*D* for p ranging from 0.05 to 0.50; with n between 2 and 20, and r' between 1 and 20.

For $p > 0.5$, we can utilise the fact that the probability is

$$P_r = 1 - \sum_{r=n-r'+1}^{r=n} {}_nC_r q^r p^{n-r} \tag{3.56}$$

3.3.4.3 Mean and Standard Deviation of Binomial Distribution

If p is the proportion of successes in the population, then the mean number of successes in n trials is

$$\mu = np \tag{3.57}$$

This is clear, as the mean number of successes in n trials is equal to the probability of success in one trial times the number of trials.

The standard deviation for a binomial frequency distribution is

$$\sigma = \sqrt{npq} \text{ or } \sqrt{np(1-p)} \tag{3.58}$$

q is not independent but is equal to $(1 - p)$. The binomial distribution can be expressed in terms of two parameters, n and p.

Equations (3.57) and (3.58) will now be derived using the definition of expectation for the mean and variance. We have

$$E(r) = \mu = \sum_{r=0}^{n} rP_r = \sum_{r=0}^{n} r\frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$

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

or

$$\mu = np \tag{3.59}$$

For the variance σ^2

$$E(r-\mu)^2 = \sigma^2 = \sum_{r=0}^{n}(r-\mu)^2 P_r = \sum_{r=0}^{n} r^2 P_r - 2\mu\sum_{r=0}^{n} rP_r + \mu^2\sum_{r=0}^{n} P_r$$

$$= \sum_{r=0}^{n} r^2 P_r - 2\mu(\mu) + \mu^2$$

Since $$\sum_{r=0}^{n} P_r = 1$$

and $$\sum_{r=0}^{n} rP_r = \mu = np$$

Thus $$\sigma^2 \sum_{r=0}^{n} r^2 P_r - (np)^2 = \sum_{r=0}^{n}[r(r-1)+r]P_r - (np)^2$$

$$= \sum_{r=2}^{n} r(r-1)\frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} + \sum_{r=0}^{n} rP_r - (np)^2$$

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

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

or $$\sigma^2 = n(n-1)p^2 + np - (np)^2 = np(1-p) = npq \tag{3.60}$$

Hence $$\sigma = \sqrt{npq} \tag{3.61}$$

Summarizing, if X is a binomial random variable with parameters p and n, then the mean,

$$\mu = E(X) = np \tag{3.61a}$$

and the variance,

$$\sigma^2 = V(X) = np(1-p) \tag{3.61b}$$

A binomial distribution with $n = 20$ and $p = 0.10$ is shown in Fig. 3.4(a).

Cumulative binomial distribution are given in Appendix-D.

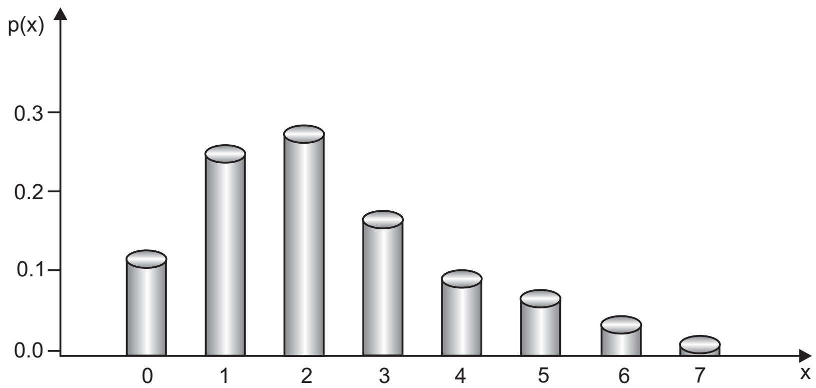


Fig. 3.4(a): Binomial pmf for n = 20, p = 0.10

Example E3.31

Let X denotes the number of mechanical components that are defective in a testing process and assume that X is a binomial random variable with $p = 0.001$. If 1000 of these components are tested, find the following:

(a) $P(X = 1)$
(b) $P(X \geq 1)$
(c) $P(X \leq 2)$
(d) mean and variance of X

SOLUTION:

We have from Eq. (3.50)

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

(a) $$P(X = 1) = \binom{1000}{1}(0.001)^1(0.999)^{999} = 0.368$$

(*b*) $P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{1000}{1}(0.001)^1(0.999)^{999} = 0.632$

(*c*) $$P(X \leq 2) = \binom{1000}{0}(0.001)^0(0.999)^{1000} + \binom{1000}{1}(0.001)^1(0.999)^{999}$$
$$+\binom{1000}{2}(0.001)^2(0.999)^{998} = 0.920$$

(*d*) $E(X) = 1000(0.001) = 1$
$V(X) = 1000(0.001)\,(0.999) = 0.999$

Example E3.32

A professional basket player makes 80% of the free throws he tries. Assuming this percentage will hold true for future attempts, find the probability that in the next eight tries the number of free throws he will make is

(*a*) exactly 8
(*b*) exactly 5.

Solution:

Here $n = 8$, $p = 0.80$, $q = 1 - p = 1 - 0.80 = 0.20$

(*a*) $P(\text{exactly } 8) = P(8) = \binom{n}{x} p^x(1-p)^{n-x} = \binom{8}{8}(0.80)^8(0.20)^0 = 1(0.167772)(1) = 0.1678$

(*b*) $P(\text{exactly } 5) = P(5) = \binom{n}{x} p^x(1-p)^{n-x} = \binom{8}{5}(0.80)^5(0.20)^3 = (56)(0.32768)(0.008) = 0.1468$

Example E3.33

A soft drink company conducted a taste survey before marketing a new soft drink. The results of the survey showed that 80% of the people like this soft drink. On a certain day, 8 customers bought it.

(*a*) let x denotes the number of customers in this sample of 8 who will like this soft drink. Using the binomial probabilities table, find the probability distribution of x and draw a graph of the probability distribution. Find the mean and standard deviation.

(*b*) Using the binomial distribution of part *(a)*, find the probability that exactly three of the eight customers will like the soft drink.

Solution:

(*a*) Here $n = 8$, $p = 0.80$. From Appendix-*C*, for $n = 8$, and $p = 0.88$ we have x, $P(x)$ tabulated as in Table E3.33.

Table E3.33

x	P(x)
0	0.0000
1	0.0001
2	0.0011
3	0.0092
4	0.0459
5	0.1468
6	0.2936
7	0.3355
8	0.1678

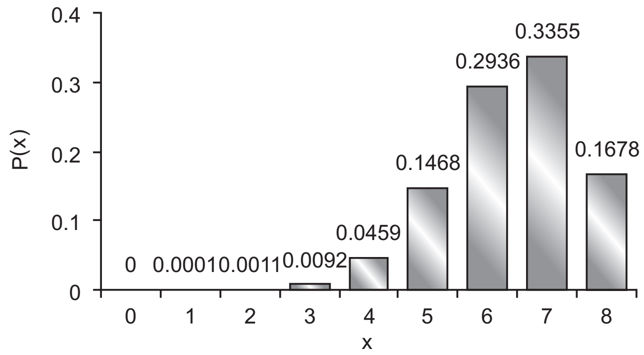


Fig. E3.33

The mean and standard deviation of x are (from Eqs. (3.61a) and (3.61b)

Mean

$$\mu = np = 8(0.80) = 6.4$$

Standard deviation

$$\sigma = \sqrt{npq} = \sqrt{8(0.80)(0.20)} = 1.1317$$

(b) P(exactly 3 customers like the soft drink) = $P(3) = 0.0092$

Example E3.34

A certain mechanical system contains 10 components. Assuming that the probability of each individual component will fail is 0.2 and the components fail independently of each other. Given that at least one of the components has failed, what is the probability that at least 2 of the components have failed?

Solution:

The number X of components that fail will have a binomial distribution with parameters $n = 10$ and $p = 0.2$. Hence,

$$P(X \geq 2 \mid X \geq 1|) = \frac{P(X \geq 2)}{P(X \geq 1)} = \frac{1 - P(X = 0) - P(X = 1)}{1 - P(X = 0)}$$

From Appendix-*C*, for $n = 10$, $p = 0.2$, we have

$$P(X \geq 2 \mid X \geq 1|) = \frac{1 - 0.1074 - 0.2684}{1 - 0.1074} = \frac{0.6242}{0.8926} = 0.6993$$

Example E3.35

(a) If a fair coin is (probability of heads equals 1/2) is tossed independently 10 times. Use the table in Appendix-*C* of the binomial distribution to find the probability that strictly more heads are obtained than tails.

(b) Suppose that the probability that a certain experiment will be successful is 0.3, and let X denotes the number of successes that are obtained in 15 independent performances of the experiment. Use the table in Appendix-*C* for the binomial distribution to determine the value of $P(6 \leq X \leq 9)$.

Solution:

(a) Let X be the number of heads obtained. More heads than tails are obtained if $X \in \{6, 7, 8, 9, 10\}$. The probability of this event is the sum of the numbers in the binomial table in Appendix-*C*. Corresponding to

$$P = 0.5$$

and $$n = 10$$

for $$x = 6, 7, 8, 9, 10 = 0.2051 + 0.1172 + 0.0439 + 0.0098 + 0.0010 = 0.37695.$$

By the symmetry of this binomial distribution, we can also compute the sum as

$$(1 - P(X = 5))/2 = (1 - 0.2461)/2 = 0.37695.$$

(b) From the table in Appendix-*C* for the binomial distribution with parameters $n = 15$ and $p = 0.3$, that

$$P(6 \leq X \leq 9) = P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9)$$
$$= 0.1472 + 0.0811 + 0.0348 + 0.0116 = 0.2747$$

Example E3.36

According to a particular survey, the probability is 0.40 that a traffic fatality involves an intoxicated or alcohol-impaired driver or non-occupant. In 8 traffic fatalities, find the probability that the number, A, which involve an intoxicated or alcohol-impaired driver or non-occupant is

(a) exactly 3; at least 3; at most 3

(b) between 2 and 4, inclusive

(c) find and interpret the mean of the random variable A

(d) obtain the standard deviation of A.

Solution:

Here, $n = 8$ and $p = 0.40$. Thus $q = 1 - p = 1 - 0.40 = 0.60$

From Eq. (3.50), we have

$$f(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

Refer to the table in Appendix-*C* for $n = 8$ and $p = 0.40$

(*a*) $P(\text{exactly } 3) = \binom{8}{3}(0.4)^3(0.6)^5 = 0.0241$

$P(3 \text{ or more}) = 1 - P(0 \text{ or } 1 \text{ or } 2) = 1 - \binom{8}{0}(0.4)^0(0.6)^8 - \binom{8}{1}(0.4)^1(0.6)^7 - \binom{8}{2}(0.4)^2(0.6)^6$

$= 1 - 0.0168 - 0.0896 - 0.2090 = 0.6846$

$P(\text{at most } 3) = P(0) + P(1) + P(2) + P(3)$

$= \binom{8}{0}(0.4)^0(0.6)^8 - \binom{8}{1}(0.4)^1(0.6)^7 - \binom{8}{2}(0.4)^2(0.6)^6 + \binom{8}{3}(0.4)^3(0.6)^5$

$= 0.0168 + 0.0896 + 0.2090 + 0.2787 = 0.5941$

(*b*) $P(2 \le A \le 4) = P(2) + P(3) + P(4) = \binom{8}{2}(0.4)^2(0.6)^6 - \binom{8}{3}(0.4)^3(0.6)^5 - \binom{8}{4}(0.4)^4(0.6)^4$

$= 0.2092 + 0.2787 + 0.2322 = 0.7201 \approx 0.72$

(*c*) The mean of A is $\mu = np = 8(0.4) = 3.2$. On average, of 8 traffic fatalities, 3.2 will involve an intoxicated or alcohol-impaired driver or non-occupant.

(*d*) Standard deviation $= \sqrt{\sigma^2} = \sqrt{np(1-p)} = \sqrt{8(0.4)(0.6)} = \sqrt{19.2} = 4.3418$

Example E3.37

According to a particular survey, 14.9% of those who have received a doctor's degree in engineering are blacks. Suppose that 6 people who have received their doctor's degree in engineering are randomly selected. Find the probability that

(*a*) exactly 2 are black
(*b*) exactly 4 are black
(*c*) at least 2 are black
(*d*) find the probability distribution of the number of blacks in a sample of 6 persons who have received their doctor's degree in engineering
(*e*) why is the probability distribution obtained in part *(d)* only approximately correct? What is the exact distribution called?

Solution:

Here, $n = 6$ and $p = 0.149$ and $q = 1 - p = 1 - 0.149 = 0.851$

From Eq. (3.50), we have

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

(*a*) $P(2) = \binom{6}{2}(0.149)^2(0.851)^4 = 0.1747$

(*b*) $P(4) = \binom{6}{4} (0.149)^4(0.851)^2 = 0.0054$

(*c*) $P(\text{at least } 2) = 1 - P(0 \text{ or } 1) = 1 - \binom{6}{0} (0.149)^0(0.851)^6 - \binom{6}{1} (0.149)^1(0.851)^5$

$$= 1 - 0.3798 - 0.3990 = 0.2212$$

(*d*) $P(X = x) = \binom{6}{x} (0.149)^x(0.851)^{6-x} \quad \text{for } x = 0, 1, 2, \ldots, 6.$

Applying this formula for each value of *x* gives the results in the Table E3.37.

Table E3.37

x	P(X = x)
0	0.3798
1	0.3990
2	0.1747
3	0.0408
4	0.0054
5	0.0004
6	0.0000

(*e*) The sampling was actually done without replacement, so the trials are not independent and the success probability changes very slightly from trial to trial. The exact probability distribution is called a hypogeometric distribution.

Example E3.38

Suppose the probability is 0.67 that the favorite in a horse race will finish in the money (first, second, or third place). In the next 5 races, what is the probability that the favorite finishes in the money

(*a*) exactly 2 times
(*b*) exactly 4 times
(*c*) at least 4 times
(*d*) between 2 and 4 times, inclusive
(*e*) find the probability distribution of the random variable, *X*, the number of times the favorite finishes in the money in the next 5 races
(*f*) identify the probability distribution of *X* as right skewed, symmetric, or left skewed without checking its probability distribution or its histogram
(*g*) Draw a probability histogram for *X*
(*h*) find the mean and standard deviation of the random variable *X*.

SOLUTION:

Refer to the table in Appendix-*C*.

Here $n = 5$, $p = 0.67$ and $q = 1 - p = 1 - 0.67 = 0.33$

From Eq. (3.50) we have

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$P(0) = \binom{5}{0} (0.67)^0 (0.33)^5 = 0.004$$

$$P(1) = \binom{5}{1} (0.67)^1 (0.33)^4 = 0.040$$

$$P(2) = \binom{5}{2} (0.67)^2 (0.33)^3 = 0.161$$

$$P(3) = \binom{5}{3} (0.67)^3 (0.33)^2 = 0.328$$

$$P(4) = \binom{5}{4} (0.67)^4 (0.33)^1 = 0.332$$

$$P(5) = \binom{5}{5} (0.67)^5 (0.33)^0 = 0.135$$

(*a*) $P(2) = 0.161$

(*b*) $P(4) = 0.332$

(*c*) $P(X \geq 4) = P(4) + P(5) = 0.332 + 0.135 = 0.467$

(*d*) $P(2 \leq X \leq 4) = P(2) + P(3) + P(4) = 0.161 + 0.328 + 0.332 = 0.821$

(*e*)

x	P(X = x)
0	0.004
1	0.040
2	0.161
3	0.328
4	0.332
5	0.135

(*f*) Left-skewed

(*g*) See Fig. E3.38.

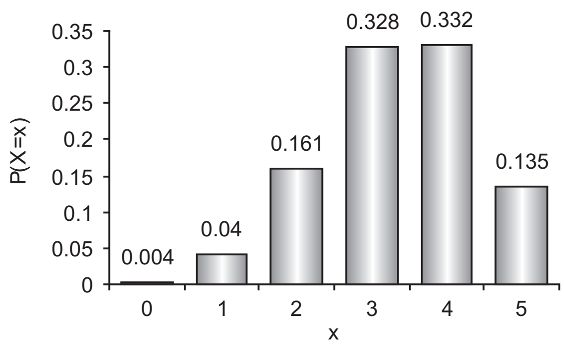


Fig. E3.38

x	P(X = x)	xP(X = x)	x^2	x^2P(X = x)
0	0.004	0.000	0	0.000
1	0.040	0.040	1	0.040
2	0.161	0.322	4	0.644
3	0.328	0.984	9	2.952
4	0.332	1.328	16	5.312
5	0.135	0.675	25	3.375
Σ		3.349		12.323

(h) Mean = μ = 3.349 (from the above table)

Variance,

$$\sigma^2 = 12.323 - 3.349^2 = 1.107$$

Hence, standard deviation = $\sqrt{1.107}$ = 1.052

Alternate calculation:

$$\mu = np = 5(0.67) = 3.35$$

$$\sigma^2 = np(1 - p) = 5(0.67)(0.33) = 1.1055$$

or $$\sigma = \sqrt{1.1055} = 1.051$$

3.3.5 Poisson Distribution

The Poisson distribution, named after the French mathematician Simeon D. Poisson is another important probability distribution of a discrete random variable that has many applications. The Poisson distribution is applied to experiments with random and independent occurrences.

Given an interval of real numbers, assume events occur at random throughout the interval. If the interval can be partitioned into subintervals of small enough length such that

1. the probability of more than one event in a subinterval is zero
2. the probability of one event in a subinterval is the same for all subintervals and proportional to the length of the subinterval, and
3. the event in each subinterval is independent of other subintervals, the random experiment is called a *Poisson process*.

Independence of occurrences means that one occurrence (or non-occurrence) of an event does not influence the successive occurrence or nonoccurrences of that event. The occurrences are always considered with respect to an interval. The interval may be a time interval, a space interval, or a volume interval. The actual number of occurrences within an interval is random and independent. If the average number of occurrences for a given interval is known, then by using the Poisson probability distribution we can determine the probability of a certain number of occurrences, x, in that interval. Note that the number of actual occurrences in an interval is denoted by x.

Conditions to apply Poisson probability distribution

The following three conditions must be satisfied to apply the Poisson probability distribution:

1. x is a discrete random variable
2. The occurrences are random
3. The occurrences are independent

The following are a few examples of discrete random variables for which the occurrences are random and independent.

Some of the phenomena that follow the Poisson distribution are:

1. counts of flaws in castings
2. the number of vehicles on a highway
3. the number of customers visiting a bank
4. the number of accidents that occur on a given highway during a period of time
5. the number of telephone calls
6. counts of power outages
7. counts of atomic particles emitted from a specimen.

The random variable X that equals the number of events in the interval is a *Poisson random variable* with parameter $0 < \lambda$, and the probability mass function (pmf) of X is

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}; \qquad x = 0, 1, 2, \ldots \tag{3.62}$$

The Poisson distribution is made up of a series of terms:

$$e^{-1}, e^{-1}\lambda, \frac{e^{-\lambda}\lambda^2}{2!}, \frac{e^{-\lambda}\lambda^3}{3!}, \ldots.$$

representing respectively, the probabilities of occurrence of 0, 1, 2, 3, 4, etc. events, where e is the base of the natural logarithm and λ is the mean frequency of occurrence. The sum of the probabilities is one (1) because

$$\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \tag{3.62a}$$

and the summation on the right-hand side of the equation (3.62*a*) is recognized to be Taylor's expansion of e^x evaluated at λ. Hence, the summation equals e^{λ} and the right-hand side equals $e^{-\lambda}e^{\lambda} = 1$. Poisson distribution for $\lambda = 3$ is shown in Fig. 3.4(*b*).

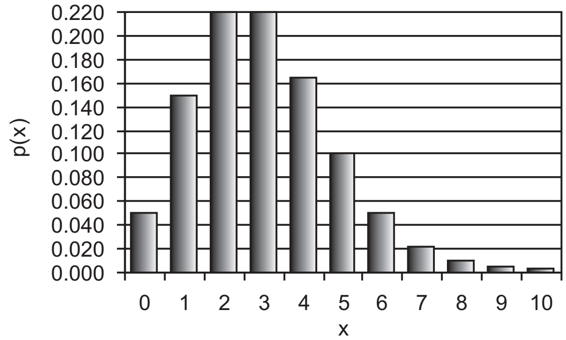


Fig. 3.4(b)

3.3.5.1 Derivation from Binomial Distribution

The Poisson distribution can also be deduced from the binomial distribution, provided that n is large ($\to\infty$), p is very small ($\to 0$), and np is finite. It was shown earlier that in n trials the probability of an event succeeding r times is

$$P_r = \frac{n!}{r!(n-r)!} p^r q^{n-r} \tag{3.63}$$

When n is large compared with r,

$$\frac{n!}{(n-r)!} = n(n-1)(n-2)...(n-r+1) \simeq n^r$$

Therefore, the probability of r successes becomes

$$P_r = \frac{n^r}{r!} p^r q^{n-r} \tag{3.64}$$

Now if p is very small and r is not large,

$$q^r = (1-p)r \simeq 1$$

and $$q^{n-r} \simeq q^n = (1-p)^n$$

Hence $$P_r = \frac{(np)^r}{r!}(1-p)^n = \frac{(np)^r}{r!}\left[1 - np + \frac{n(n-1)(-p)^2}{2!} + \frac{n(n-1)(n-2)(-p)^3}{3!} + \cdots\right]$$

$$\simeq \frac{(np)^r}{r!}\left[1 - np + \frac{(np)^2}{2!} - \frac{(np)^3}{3!} + \cdots\right]$$

Thus $$P_r = \frac{(np)^r}{r!} e^{-np} \tag{3.65}$$

This, then, is the probability of r successes in n trials.

3.3.5.2 Mean and Standard Deviation

The mean number of occurrences of an event per unit of time (or space) is

$$\mu = np \tag{3.66}$$

and the standard deviation of the numbers of events is

$$\sigma = \sqrt{np} \tag{3.67}$$

Thus the mean and variance are equal to one another:

$$\mu = \sigma^2 = np \tag{3.68}$$

Equations (3.66) and (3.68) will now be derived.

$$\text{mean} = E(r) = \sum_{r=0}^{\infty} rP_r = \sum_{r=0}^{\infty} \frac{re^{-\lambda}\lambda^r}{r!} = 0 + \lambda e^{-\lambda} + 2\frac{\lambda^2 e^{-\lambda}}{2!} + 3\frac{\lambda^3 e^{-\lambda}}{3!} + \cdots$$

$$= \lambda e^{-\lambda}\left[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right] = \lambda e^{-\lambda} e^{\lambda} = \lambda \tag{3.69}$$

The variance σ^2 is given by

$$\sigma^2 = E(r-\lambda)^2 = \sum_{r=0}^{\infty}(r-\lambda)^2 P_r$$

$$\sigma^2 = \sum_{r=0}^{\infty} r^2 P_r - 2\lambda \sum_{r=0}^{\infty} rP_r + \lambda^2 \sum_{r=0}^{\infty} P_r$$

Now
$$-2\lambda \sum_{r=0}^{\infty} rP_r = -2\lambda(\lambda) = -2\lambda^2$$

$$\lambda^2 \sum_{r=0}^{\infty} P_r = \lambda^2(1) = \lambda^2$$

and
$$\sum_{r=0}^{\infty} r^2 P_r = \sum_{r=0}^{\infty}[r(r-1)+r]P_r = \sum_{r=0}^{\infty} r(r-1)P_r + \lambda \;= \left[0+0+2\frac{\lambda^2 e^{-\lambda}}{2!} + 6\frac{\lambda^3 e^{-\lambda}}{3!} + \cdots\right] + \lambda$$

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

Thus, collecting the terms, we have

$$\sigma^2 = \lambda^2 + \lambda - 2\lambda^2 + \lambda^2$$

or
$$\sigma^2 = \lambda \tag{3.70}$$

Hence, if X is a Poisson random variable with parameter λ, then

the mean $\qquad \mu = E(X) = \lambda$, and

the variance $\qquad \sigma^2 = V(X) = \lambda^2$

Note that the Poisson distribution contains only one parameter, np, the mean occurrence of an event, and we do not know the value of n. In the binomial distribution we know the number of times an event occurs and the number of times an event does not occur.

Cumulative Poisson probabilities are given in Appendix-*D*. A plot of the Poisson distribution for $\lambda = 3$ is shown in Fig. 3.5. This Poisson distribution has a very long tail to the right (the distribution is skewed).

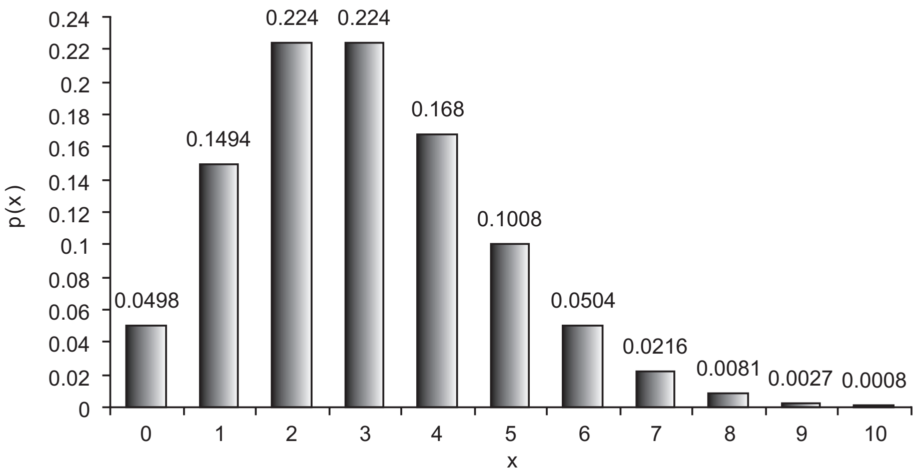


Fig. 3.5: Poisson distribution for $\lambda = 3$

Example E3.39

A survey found that 1.5% of occupied housing units have 7 or more people living within. Use the Poisson distribution to determine the approximate probability that, of 200 randomly selected occupied housing units, there are

(a) none with 7 or more persons

(b) 3 or more with 7 or more persons

Solution:

We use $\lambda = np = 200(0.15) = 3.0$. We also note that $n \geq 100$ and $np \leq 10$. Hence, apply Eq. (3.62),

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

(a) $P(X = 0) = \dfrac{e^{-3.0}(3.0)^0}{0!} = 0.0498$

(b) $P(\text{at least } 3) = 1 - P(X = 0 \text{ or } 1 \text{ or } 2)$

$$= 1 - \left[e^3 \frac{3^0}{0!} + e^{-3}\frac{3^1}{1!} + e^{-3}\frac{3^2}{2!}\right] = 1 - [0.0498 + 0.1494 + 0.2240] = 0.5768$$

Example E3.40

Assume the number of errors along a magnetic recording surface is a Poisson random variable with a mean of one error every 10^5 bits. A sector of data consists of 5000 eight-bit bytes. Find

(a) the probability of more than one error in a sector

(b) the mean number of sectors until an error is found.

Solution:

Let x denotes the number of errors in a sector. Then, X is a Poisson random variable with $\lambda = 0.4$.

(a) $P(X > 1) = 1 - P(X < 1) = 1 - e^{-0.4} - e^{-0.4}(0.4) = 1 - 0.6703 - 0.2681 = 0.06157$

(b) Let Y denotes the number of sectors until an error is found. Then, Y is a geometric random variable and

$P = P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-0.4} = 1 - 0.6703 = 0.3297$

$$E(Y) = \frac{1}{p} = \frac{1}{0.3297} = 3.03306$$

Example E3.41

The probability that an individual recovers from an illness in a 2-week time period without medical treatment is 0.1. If 20 independent individuals suffering from this illness are treated with a medicine and 4 recover in a 2-week period. If the medicine has no effect, what is the probability that 4 or more people recover in a 2-week time period?

SOLUTION:

Let x denotes the number of individuals that recover in 2-week time period. Assume the individuals are independent, then, X is binomial random variable with $n = 20$ and $p = 0.1$.

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - [P(X = 0) + P(x = 1) + P(x = 2) + P(X = 3)]$$

$$= 1 - \left[\binom{20}{0}(0.1)^0(0.9)^{20} + \binom{20}{1}(0.1)^1(0.9)^{19} + \binom{20}{2}(0.1)^2(0.9)^{18} + \binom{20}{3}(0.1)^3(0.9)^{17}\right]$$

$$= 1 - [0.1216 + 0.2702 + 0.2852 + 0.1901] = 0.1330$$

Example E3.42

A survey found that the traffic flowing through an intersection with an average of 3 cars per 30 seconds. Assume the traffic flow can be modeled as a Poisson distribution.

(*a*) find the probability of no cars through the intersection within 30 seconds

(*b*) find the probability of 3 or more cars through the intersection within 30 seconds

(*c*) find the minimum number of cars through the intersection so that the probability of this number or fewer cars in 30 seconds is at least 90%

(*d*) if the variance of the number of cars through the intersection per minute is 20, is the Poisson distribution appropriate?

SOLUTION:

(*a*) $\lambda = 3$ cars/30 seconds

we have $f(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

$$P(X = 0) = \frac{e^{-3}3^0}{0!} = 0.0498$$

(*b*) $P(X \geq 3) = 1 - P(X < 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$

$$= 1 - \left[\frac{e^{-3}3^0}{0!} + \frac{e^{-3}3^1}{1!} + \frac{e^{-3}3^2}{2!}\right] = 1 - [0.0498 + 0.1494 + 0.2240] = 0.5768$$

(*c*) $P(X \leq x) \geq 0.9$ or $x = 5$

(*d*) $\sigma^2 = \lambda = 6$. This is not appropriate.

Example E3.43

A student's campus newspaper in a particular university contains an average of 1.2 typographical errors per page.

(*a*) find the probability that a randomly selected page of this newspaper will contain exactly 4 typographical errors using the Poisson formula

(*b*) find the probability that the number of typographical errors on a randomly selected page will be

  (*i*) more than 3

  (*ii*) less than 4

Use the Poisson probabilities table in Appendix-*D*.

Solution:

Let x be the number of typographical errors on a randomly selected page of this newspaper. Since it contains an average of 1.2 typographical errors per page, $\lambda = 1.2$. We have Eq. (3.62)

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

(a) $P(\text{exactly } 4) = P(x = 4) = \dfrac{\lambda^x e^{-\lambda}}{x!} = \dfrac{(1.2)^4 e^{-1.2}}{4!} = 0.0260$

(b) (i) $P(\text{more than } 3) = P(x > 3) = P(4) + P(5) + P(6) + P(7) = 0.0260 + 0.0062 + 0.0012 + 0.0002 = 0.0336$

(ii) $P(\text{less than } 4) = P(x < 4) = P(0) + P(1) + P(2) + P(3) = 0.3012 + 0.3614 + 0.2169 + 0.0867 = 0.9662$

Example E3.44

In a statistics course final examination, 15% of the students fail.

(a) find the probability that in a random sample of 100 students in that statistics class who took the final examination exactly 20 will fail. Use the Poisson formula.

(b) find the probability that the number of students who fail this statistics final examination in a randomly selected 100 students is

(i) at most 9

(ii) 10 to 16

(iii) at least 20

Use the Poisson probabilities table in Appendix-*D*.

Solution:

Let x be the number of students in a random sample of 100 who fail the final examination in statistics. Since, an average, 15% of the students fail the examination, $\lambda = 0.15(100) = 15$.

(a) $P(\text{exactly 20 fail}) = P(x = 20) = \dfrac{e^{-\lambda}\lambda^x}{x!} = \dfrac{(e^{-15})(15)^{20}}{20!} = 0.0418$

(b) (i) $P(\text{at most 9 fail}) = P(x \le 9) = P(0) + P(1) + P(2) + P(3) + \cdots + P(9)$

$= 0.0000 + 0.0000 + 0.0000 + 0.0002 + 0.0006 + 0.0019 + 0.0048 + 0.0104 + 0.0194 + 0.0324$

$= 0.0697$

(ii) $P(\text{10 to 16 fail}) = P(10 \le x \le 16) = P(10) + P(11) + P(12) + P(13) + P(14) + P(15) + P(16)$

$= 0.0486 + 0.0663 + 0.0829 + 0.0956 + 0.1024 + 0.1024 + 0.0960 = 0.5942$

(iii) $P(\text{at least } 20) = P(x \ge 20) = P(20) + P(21) + P(22) + \cdots + P(39)$

$= 0.0418 + 0.0299 + 0.0204 + 0.0133 + 0.0083 + 0.0050 + 0.0029 + 0.0016 + 0.0009 + 0.0004$
$+ 0.0002 + 0.0001 + 0.0001 + 0.0000 + \cdots + 0.0000 = 0.1249$

Example E3.45

A hardware store in a big city receives an average of 9.8 telephone calls per hour.

(a) find the probability that exactly 6 telephone calls will be received at this store during a certain hour. Use the Poisson formula.

(b) find the probability that the number of telephone calls received at this store during a certain hour will be

(i) less than 8

(ii) more than 12

(iii) 5 to 8

Use the Poisson distribution table in Appendix-*D*.

SOLUTION:

Let x be the number of telephone calls received at this hardware store during a certain hour. Since the average number of telephone calls per hour is 9.8, $\lambda = 9.8$.

(a) $$P(\text{exactly } 6) = P(x = 6) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{(e^{-9.8})(9.8)^6}{6!} = 0.0677$$

(b) *(i)* $P(\text{less than } 8) = P(x < 8) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5) + P(6) + P(7)$
$= 0.0001 + 0.0005 + 0.0027 + 0.0087 + 0.0213 + 0.0418 + 0.0682 + 0.0955 = 0.2388$

(ii) $P(\text{more than } 12) = P(x > 12) = P(13) + P(14) + P(15) + \cdots$
$= 0.0685 + 0.0479 + 0.0313 + 0.0192 + 0.01111 + 0.0060 + 0.0031 + 0.0015 + 0.0007 + 0.0003$
$+ 0.0001 + 0.0001 = 0.1898$

(iii) $P(5 \text{ to } 8) = P(5 \le x \le 8) = P(5) + P(6) + P(7) + P(8) = 0.0418 + 0.0682 + 0.0955 + 0.1170 = 0.3225$

Example E3.46

An average of 0.7 accidents occur per day in a large city.

(a) find the probability that no accidents will occur in that city on a given day

(b) write the probability distribution of x, where x denotes the number of accidents that will occur in that city on a given day

(c) find the mean, variance and standard deviation of the probability distribution developed in part *(b)*.

SOLUTION:

x = number of accidents on a given day in that city.

Since the average number of accidents per day is 0.7, $\lambda = 0.7$.

(a) $$P(\text{no accidents}) = P(x = 0) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-0.7}(0.7)^0}{0!} = 0.4966$$

(b) From the table in Appendix-*D*, we have

x	P(x)
0	0.4966
1	0.3476
2	0.1217
3	0.0284
4	0.0050
5	0.0007
6	0.0001
7	0.0000

(*c*) The mean, variance and standard deviation are:

mean $\mu = \lambda = 1.7$

variance $\sigma^2 = \lambda = 0.7$ and

standard deviation, $\sigma = \sqrt{\lambda} = \sqrt{0.7} = 0.8366$

Example E3.47

The number of male mates of a queen bee was found to have a Poisson distribution with parameter $\lambda = 2.7$. Find the probability that the number, *N*, of male mates of a queen bee is

(*a*) exactly 2

(*b*) at most 2

(*c*) between 1 and 3, inclusive

(*d*) on average, how many male mates does a queen bee have?

(*e*) develop a table of probabilities for the random variable, *N*. Compute the probabilities until they are zero to 4 decimal places

(*f*) draw a histogram of the probabilities in part (*c*).

Solution:

(*a*) $P(N = 2) = e^{-2.7}\dfrac{2.7^2}{2!} = 0.2450$

(*b*) $P(N \le 2) = P(0 \text{ or } 1 \text{ or } 2) = e^{-2.7}\dfrac{2.7^0}{0!} + e^{-2.7}\dfrac{2.7^1}{1!} + e^{-2.7}\dfrac{2.7^2}{2!} = 0.0672 + 0.1815 + 0.2450 = 0.4937$

(*c*) $P(1 \le N \le 3) = P(1 \text{ or } 2 \text{ or } 3) = e^{-2.7}\dfrac{2.7^1}{1!} + e^{-2.7}\dfrac{2.7^2}{2!} + e^{-2.7}\dfrac{2.7^3}{3!} = 0.1815 + 0.2450 + 0.2205 = 0.6470$

(*d*) On average, the number of mates of a queen bee is

$\mu = \lambda = 2.7$

(*e*) Refer to Appendix-*D*. The results are shown in Table E3.47.

Table E3.47

N	P(N = n)
0	0.0672
1	0.1815
2	0.2450
3	0.2205
4	0.1488
5	0.0804
6	0.0362
7	0.0139
8	0.0047
9	0.0014
10	0.0004
11	0.0001
12	0.0000

(*f*) See Fig. E3.47

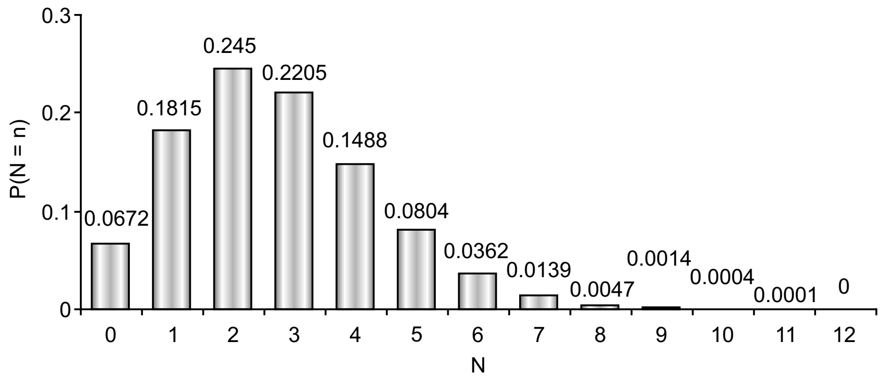


Fig. E3.47

3.4 CONTINUOUS PROBABILITY DISTRIBUTIONS

3.4.1 The Normal Distribution

In everyday life, people deal with and use a wide variety of variables. Some of these variables — such as heights of people, scores in final examinations, aptitude test scores, TOEFL (Test of English as a Foreign Language), and GRE (Graduate Record Examination) share an important characteristic: their distributions have roughly the shape of a normal curve, that is, a special type of bell-shaped curve like one shown in Fig. 3.6.

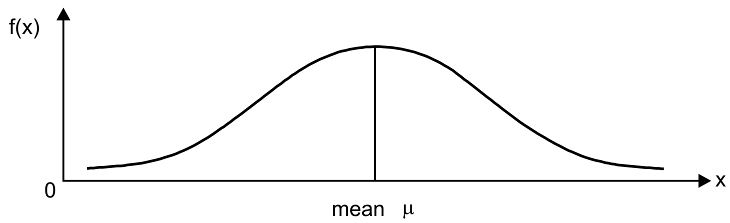


Fig. 3.6 Normal curve

The normal distribution is in many respects the cornerstone of statistics. A random variable X is said to have a normal distribution with mean $\mu(-\infty < \mu < \infty)$ and variance $\sigma^2 > 0$ if it has the density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)\sigma]^2} \qquad -\infty < X < \infty \tag{3.71}$$

The distribution is illustrated graphically in Fig. 3.6. The normal distribution is used so extensively that the shorthand notation $X \sim N(\mu, \sigma^2)$ is often used to indicate that the random variable X is normally distributed with mean μ and variance σ^2.

A variable is said to be a normally distributed variable or to have a *normal distribution* if its distribution has the shape of a normal curve.

If a variable of a population is normally distributed and is the only variable under consideration, the general practice is to assume that the population is *normally distributed* or that it is a *normally distributed population*. In reality, a distribution is unlikely to have exactly the shape of a normal distribution curve. If a variable's distribution is shaped roughly like a normal curve, we can consider that the variable is an *approximately normally distributed variable* or that it has *approximately a normal distribution*.

A normal distribution (and hence a normal curve) is completely determined by the mean and standard deviation. Hence, two normally distributed variables having the same mean and standard deviation must have the same distribution. We identify a normal curve by stating the corresponding mean and standard deviation and calling those to *parameters* of the normal curve.

A normal distribution is symmetric about and centred at the mean of the variable, and its spread depends on the standard deviation of the variable. The larger the standard deviation, the flatter and more spread out is the distribution.

In summary, the normal curve associated with a normal distribution is bell shaped, centered at μ, and close to the horizontal axis outside the range from $\mu - 3\sigma$ to $\mu + 3\sigma$ as shown in Figs. 3.6 and 3.6(*a*).

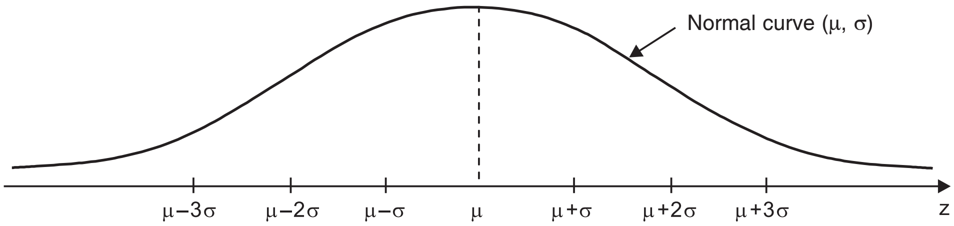


Fig. 3.6(a)

Random variables with different means and variances can be modeled by normal probability functions with appropriate choices of the centre and width of the curve. The value of $E(X) = \mu$ determines the centre of the probability density function and the value of $V(X) = \sigma^2$ determines the width. Figure 3.6(*c*) illustrates several normal probability density functions with selected values of *m* and σ^2. Each one has the characteristic symmetric bell-shaped curve, but the centers and dispersions differ.

Definition of Normal Distribution

A random variable *X* with probability density function

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

is a *normal random variable* with parameters μ, where $-\infty < x < \infty$, and $\sigma > 0$.

Also $\quad E(X) = \mu$ and $V(X) = \sigma^2$

and the notation $N(\mu, \sigma^2)$ is used to denote the distribution. The mean and variance of *X* are equal to μ and σ^2, respectively.

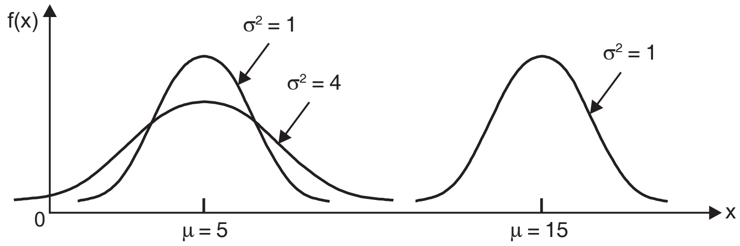


Fig. 3.6(b): Normal probability density functions for selected values of the parameters μ and σ²

3.4.1.1 Properties of the Normal Distribution

The normal distribution has several important properties:

1. $\int_{\infty}^{\infty} f(x)\,dx = 1$
2. $f(x) \geq 0$ for all x
3. $\lim_{x\to\infty} f(x) = 0$ and $\lim_{x\to\infty} f(x) = 0$
4. $f[(x + \mu)] = f[-(x - \mu)]$. The density is symmetric about μ.
5. The maximum value of $f(x)$ occurs at $x = \mu$.
6. The points of inflection of $f(x)$ are at $x = \mu \pm \sigma$.

Also, $P(a \leq x \leq b) = \int_a^b f(x)\,dx$ = area under $f(x)$ from a to b for any a and b.

Property 1 may be demonstrated as follows. Let $y = (x - \mu)/\sigma$ in Eq. (3.71) and denote the integral as I. That is,

$$I = \frac{1}{\sqrt{2\pi}}\int_{\infty}^{\infty} e^{-(1/2)y^2}\,dy \frac{1}{\sqrt{2\pi}}\int_{\infty}^{\infty} e^{-(1/2)x^2}\,dx \tag{3.72}$$

on changing to polar coordinates with the transformation of variables $y = r \sin\theta$ and $z = r\cos\theta$, the integral becomes

$$I^2 = \frac{1}{2\pi}\int_0^{\infty}\int_0^{2\pi} re^{-(1/2)r^2}\,d\theta\,dr = \int_0^{\infty} re^{-(1/2)r^2}\,dr = 1 \tag{3.73}$$

3.4.1.2 Mean and Variance of the Normal Distribution

The mean of the normal distribution may be determined easily. Since

$$E(X) = \int_{-\infty}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma^2]}dx \tag{3.74}$$

and if we let $z = (x - \mu)/\sigma$, we obtain

$$E(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}(\mu + \sigma z)e^{-z^2/2}dz \tag{3.75}$$

$$= \mu = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}dz + \sigma\frac{1}{\sqrt{2\pi}}ze^{-z^2/2}dz \tag{3.76}$$

Since the integrand of the first integral is that of a normal density with $\mu = 0$ and $\sigma^2 = 1$, the value of the first integral is one. The second integral has value zero, that is,

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}ze^{-z^2/2}dz = \frac{1}{\sqrt{2\pi}}ze^{-z^2/2}\Big|_{-\infty}^{\infty} = 0$$

and thus $$E(X) = \mu[1] = \sigma[0] = \mu \tag{3.77}$$

To find the variance we must evaluate

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

And letting $z = (x-\mu)/\sigma$, we obtain

$$V(X) = \int_{-\infty}^{\infty} \sigma^2 z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \sigma^2 \left[\int_{-\infty}^{\infty} \frac{z^2}{\sqrt{2\pi}} e^{-z^2/2} dz \right]$$

$$= \sigma^2 \left[\frac{-ze^{-z^2/2}}{\sqrt{2\pi}} \Bigg|_{-\infty}^{\infty} \right] + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \sigma^2 (0+1)$$

so that $$V(X) = \sigma^2 \tag{3.78}$$

In summary the mean and variance of the normal density given in Eq. (3.78) are μ and σ^2, respectively.

3.4.1.3 The Cumulative Normal Distribution

The distribution function F is

$$F(x) = P(X \le x) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma]^2} du \tag{3.79}$$

It is impossible to evaluate this integral without resorting to numerical methods, and even then the evaluation would have to be accomplished for each pair (μ, σ^2). However, a simple transformation of variables, $z = (X-\mu)/\sigma$, allows the evaluation to be independent of μ and σ.

That is,

$$F(X) = P(X \le x) = P\left(Z \le \frac{x-\mu}{\sigma}\right) = \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \int_{-\infty}^{(x-\mu)/\sigma} \varphi(z)\, dz = \phi\left(\frac{x-\mu}{\sigma}\right) \tag{3.80}$$

3.4.1.4 The Standard Normal Distribution

The probability distribution in Eq. (3.80) above,

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < x < \infty$$

is a normal distribution with mean 0 and variance 1; that is, $Z \sim N(0, 1)$ and we say that Z has a *standard normal distribution*. A graph of the probability density function is shown in Fig. 3.7.

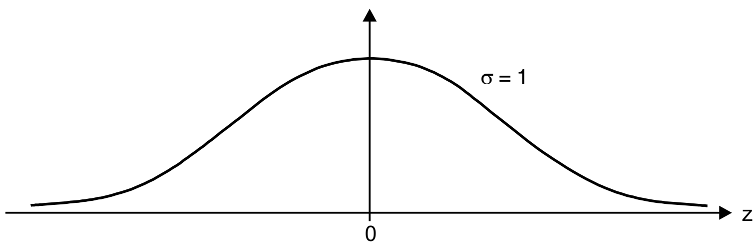


Fig. 3.7: The standard normal distribution

The corresponding distribution function is ϕ, where

$$\varphi(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} du \tag{3.81}$$

where $$z = \frac{x-\mu}{\sigma}$$

and this function has been well tabulated.

A few useful results for a normal distribution are summarised below and shown in Fig. 3.7(a). For random variable,

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$$

In other words, any normally distributed variable has the following properties:

1. 68.27% of all possible observations lie within one standard deviation to either side of the mean, that is, between $\mu - \sigma$ and $\mu + \sigma$.
2. 95.45% of all possible observations lie within two standard deviation to either side of the mean, that is, between $\mu - 2\sigma$ and $\mu + 2\sigma$.
3. 99.73% of all possible observations lie within three standard deviation to either side of the mean, that is, between $\mu - 3\sigma$ and $\mu + 3\sigma$.

These properties are shown in Fig. 3.7(b).

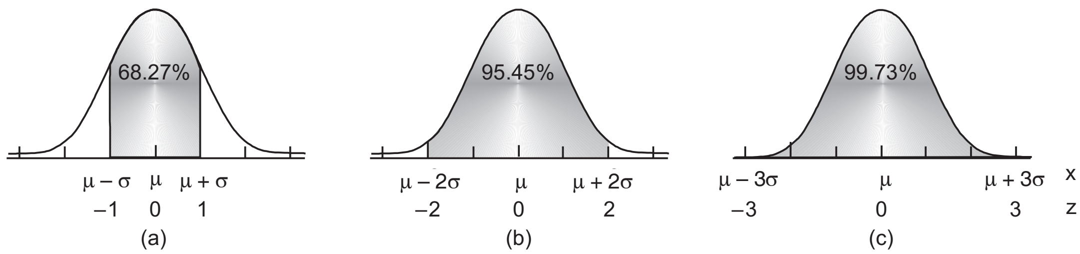


Fig. 3.7(a)

In addition, from the symmetry of $f(x)$, $P(X > \mu) = P(X < \mu) = 0.5$. The above model assigns some probability to each interval of the real line since $f(x)$ is positive for all x. The probability density function decreases as x moves further from μ. As a result, the probability that a measurement falls far from μ is small, and at some distance from μ the probability of an interval can be approximated as zero.

For a normally distributed variable, the percentage of all possible observations that (i.e., within any specified range) equals the corresponding area under its associated normal curve, expressed as a percentage. This result holds approximately for a variable that is approximately normally distributed.

A normally distributed variable having mean 0 and standard deviation 1 is said to have the *standard normal distribution*. Its associated normal curve is called the *standard normal curve*, which is shown in Fig. 3.7(c).

The area under a normal probability density function beyond 3σ from the mean is quite small. Since more than 0.9973 of the probability of a normal distribution is within the interval $(\mu - 3\sigma, \mu + 3\sigma)$, 6σ is often referred to as the *width* of a normal distribution. Advanced integration methods can be used to show that the area under the normal probability density function from $-\infty < x < \infty$ is 1.

3.4.1.5 Problem-Solving Procedure

If X is a normal random variable with $E(X) = \mu$ and $V(X) = \sigma^2$, the random variable

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

is a normal random variable with $E(Z) = 0$ and $V(Z) = 1$. That is, Z is a standard normal random variable.

Suppose X is a normal random variable with mean μ and variable σ^2. Then

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P(Z \leq z)$$

where Z is a *standard normal variable*, and $z = \left(\frac{x - \mu}{\sigma}\right)$ is the z-value obtained by standardising X.

The probability is obtained by using the table in Appendix-*E* with $z = (x - \mu)/\sigma$.

Basic Properties of the Standard Normal Curve

1. The total area under the standard normal curve is 1.
2. The standard normal curve extends indefinitely in both directions, approaching, but never touching, the horizontal axis as it does so.
3. The standard normal curve is symmetric about 0. That is, the part of the curve to the left of the vertical line at the centre in Fig. 3.7(*b*) is the mirror image of the curve to the right of it.
4. Almost all the area under the standard normal curve lies between –3 and 3.

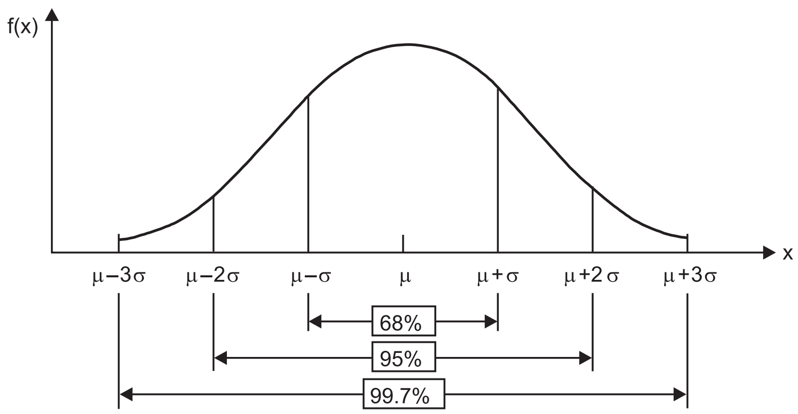


Fig. 3.7(b): Probabilities associated with a normal distribution

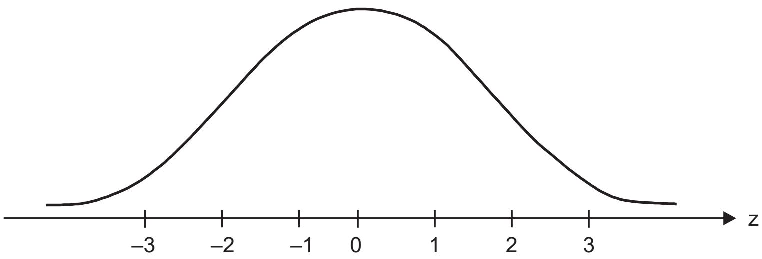


Fig. 3.7(c): Standard normal distribution

A normal random variable with $\mu = 0$ and $\sigma^2 = 1$ is called a *standard normal variable* and is denoted as Z. The cumulative distribution function of a standard normal random variable is denoted as $\phi(z) = P(Z \le z)$.

The tables in Appendix-*E* provides cumulative probabilities for a standard normal random variable.

The procedure in solving problems involving calculating the cumulative normal probabilities is simple. For example, suppose that $X \sim N(50, 4)$, and we wish to find the probability that x is less than or equal to 54; that is $P(x \le 54) = F(54)$. Since that standard normal random variable is

$$z = \frac{x - \mu}{\sigma}$$

we can *standardize* the point of interest $x = 104$ to obtain

$$z = \frac{x - \mu}{\sigma} = \frac{54 - 50}{2} = 2$$

Now the probability that the *standard* normal random variable z is less than or equal to 2 is equal to the probability that the *original* normal random variable x is less than or equal to 54. Expressed mathematically,

$$F(x) = \phi\left(\frac{x - \mu}{\sigma}\right) = \phi(z)$$

$$F(54) = \phi(2)$$

The table in Appendix-*E* contains cumulative standard normal probabilities for various values of z. From this table, we can read

$$\phi(2) = 0.9772$$

Note that in the relationship $z = (x - \mu)/\sigma$, the variable z measures the departure of x from the mean μ in standard deviation (σ) units. In our example $F(54) = f(2)$ indicates that 54 is *two* standard deviations ($\sigma = 2$) above the mean. In general, $x = \mu + \sigma\, z$. In solving problems, we sometimes need to use the symmetry property of φ in addition to the tables.

In order to find the percentages of all possible observations of a normally distributed variable that lie within any specified range, we express the range in terms of z-scores and then determine the corresponding area under the standard normal curve. The stepwise procedure to determine a percentage or probability for a normally distributed variable is presented below:

1. Sketch the normal curve associated with the variable as shown in Fig. 3.7(*d*).
2. Shade the region of interest and mark its delimiting x-value(s).

3. Compute the z-scores for the delimiting x-value(s) found in step 2.
4. Use the table in Appendix-E to find the area under the standard normal curve delimited by the z-score(s) found in step 3.

The probability distributions presented in this chapter are summarised in Table 3.1.

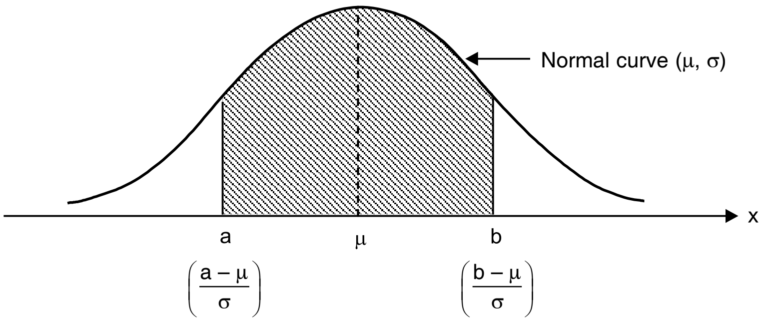


Fig. 3.7(d)

Table 3.1: Summary of probability distributions

Name	Probability Density Function	Mean	Variance
Discrete			
Hypergeometric	$\dfrac{\binom{K}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$ $x = \max(0, n - N + K), 1, \ldots$ $\min(K, n), K \leq N, n \leq N$	np, where $p = \dfrac{K}{N}$	$np(1-p)\left(\dfrac{N-m}{N-1}\right)$
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n, 0 \leq p \leq 1$	np	$np(1-p)$
Poisson	$\dfrac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 2, \ldots, \lambda > 0$	λ	λ
Continuous			
Normal	$\dfrac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ $-\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$	μ	σ^2

Example E3.48

(a) Determine the area under the standard normal curve that lies to the left of 1.24, as shown in Fig. E3.48(a).

(b) Determine the area under the standard normal curve that lies to the right of 0.77, as shown in Fig. E3.48(*b*).

(c) Determine the area under the standard normal curve that lies between –0.69 and 1.83, as shown in Fig. E3.48(*c*).

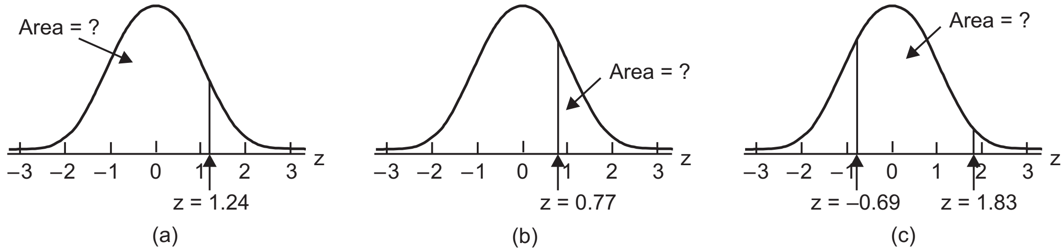


Fig. E3.48

SOLUTION:

(a) Refer to the table in Appendix-*E*. First, we go down the left hand column, labeled *z*, to "1.2". Then, going across that row to the column labeled "0.04", we reach 0.892512. This number is the area under the standard normal curve that lies to the left of 1.24, as shown in Fig. E3.48(*d*).

(b) Because the total area under the standard normal curve is 1, the area to the right of 0.77 equals 1 minus the area of the left of 0.77. We find this by first giving down the *z*-column to "0.7". Then, going across that row to the column, labeled "0.07", we reach 0.779350, which is the area under the standard normal curve that lies to the left of 0.77. Thus, the area under the standard normal curve that lies to the right of 0.77 is 1 – 0.779350 = 0.220650 as shown in Fig. E3.48(*e*).

(c) The area under the standard normal curve that lies between –0.69 and 1.83 equals the area to the left of 1.83 minus the area to the left of –0.69. The table in Appendix-*E* shows that these latter two areas are 0.966375 and 0.245097, respectively. Hence the area we seek is 0.966375 – 0.245097 = 0.721278, as shown in Fig. E3.48(*f*).

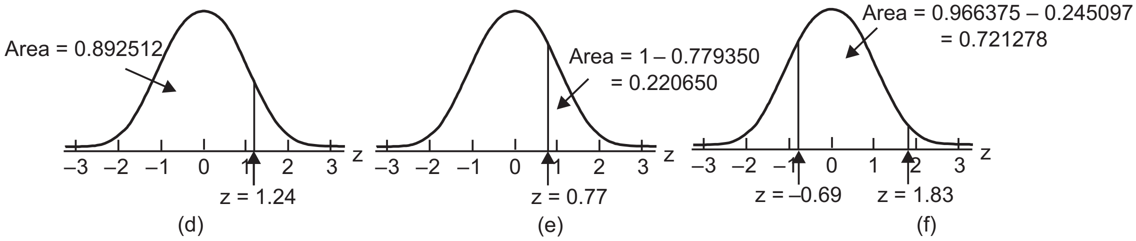


Fig. E3.48

Example E3.49

Determine the *z*-score having an area of 0.06 to its left under the standard normal curve, as shown in Fig. E3.49.

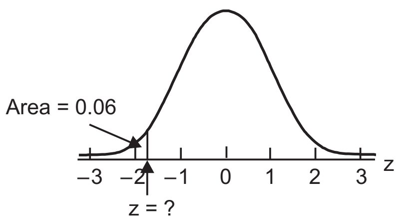


Fig. E3.49

SOLUTION:

See Fig. E3.48(*a*).

Search the body of the table in Appendix-*E* for the area 0.06. There is no such area, so use the area closest to 0.06, which is 0.060571. The z-score corresponding to that area is –1.55. Hence, the z-score having area 0.06 to its left under the standard normal curve is roughly –1.55, as shown in Fig. E3.49(*a*).

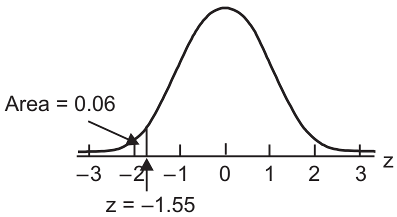


Fig. E3.49(a)

Example E3.50

Use the table in Appendix-*E* to find

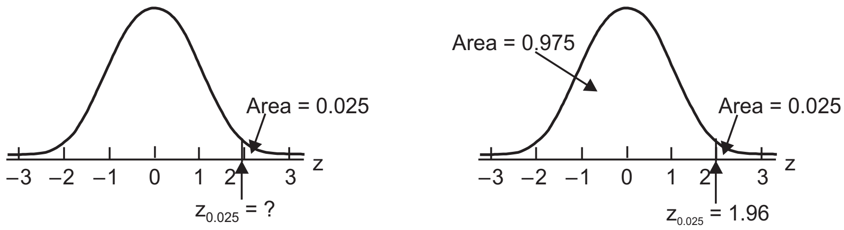
(*a*) $z_{0.025}$

(*b*) $z_{0.05}$

SOLUTION:

(*a*) $z_{0.025}$ is the z-score that has an area of 0.025 to its right under the standard normal curve, as shown in Fig. E3.50(*a*). Because the area to its right is 0.025, the area to its left is 1 – 0.025 = 0.975, as shown in Fig. E3.50(*b*). Table in Appendix-*E* contains an entry for the area 0.975002, its corresponding z-score is 1.96. Thus, $z_{0.025}$ = 1.96, as shown in Fig. E3.50(*b*).

(*b*) $z_{0.05}$ is the z-score that has an area of 0.05 to its right under the standard normal curve, as shown in Fig. E3.50(*c*). Because the area to its right is 0.05, the area to its left is 1 – 0.05 = 0.95, as shown in Fig. E3.50(*d*). Table in Appendix-*E* does not contain any entry for the area 0.95 and has two area entries equally closest to 0.95 – namely, 0.949497 and 0.950529. The z-scores corresponding to those two areas are 1.64 and 1.65 respectively. So our approximation of $z_{0.05}$ is the mean of 1.64 and 1.65, that is, $z_{0.05}$ = 1.645, as shown in Fig. E3.50(*d*).



(a, c)

(b, d)

Fig. E3.50

Example E3.51

Find the two z-scores that divide the area under the standard normal curve into a middle 0.95 area and two outside 0.025 areas, as shown in Fig. E3.51(a).

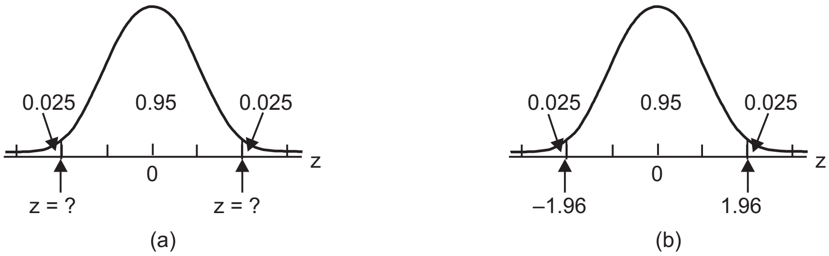


Fig. E3.51

SOLUTION:

The area of the shaded region on the right in Fig. E3.51(a) is 0.025. The corresponding z-score, $z_{0.025}$, is 1.96. Because the standard normal curve is symmetric about 0, the z-score on the left is –1.96. Therefore the two required z-scores are ±1.96, as shown in Fig. E3.51(b).

Example E3.52

In a long distance run of 10 km in a city the times of finishers are normally distributed with mean 71 minutes and standard deviation 9 minutes.

(a) determine the percentage of finishers with times between 60 and 80 minutes
(b) find the percentage of finishers with times less than 85 minutes
(c) obtain and interpret the 40th percentile for the finishing times
(d) find and interpret the 8th decile for the finishing times.

SOLUTION:

(a) For finishers with times of 60 and 80 minutes, the z-values are

$$z = \frac{x-\mu}{\sigma} = \frac{60-71}{9} = -1.22 \text{ and } z = \frac{x-\mu}{\sigma} = \frac{80-71}{9} = 1.00$$

The area to the left of z = –1.225 is 0.111233 and the area to the left of z = 1.00 is 0.841345. Hence, the area between z = –1.22 and z = 1.00 is 0.841345 – 0.111233 = 0.730112. Thus, the percentage of finishers with times between 60 minutes and 80 minutes in that city 10 km run is 73.01%.

(b) For finishing time of 85 minutes, the z-value is

$$z = \frac{x-\mu}{\sigma} = \frac{85-71}{9} = 1.56$$

The area of the left of z = 1.56 is 0.9406. Thus, the percentage of finishers with times less than 85 minutes is 94.06%.

(c) Using the table in Appendix-E, we find that an area of 0.40 lies to the left of z = –0.25. We convert this z-value to an x-value using $x = \mu + z\sigma$. Hence 40% of the finishers had times less than 40th percentile, x = 71 + (–0.25)(9) = 68.75 minutes.

(d) The 8^{th} decile is the same as the 80^{th} percentile. Using the table in Appendix-*E*, we find that an area of 0.80 lies to the left of $z = 0.84$. We convert the z-value to an x-value using $x = \mu + z\sigma$. Thus, 80% of the finishing times were less than $71 + (0.84)9 = 78.56$ minutes.

Example E3.53

If X has a normal distribution for which the mean is 1 and standard deviation is 2, find the value of each of the following:

(a) $P(X \le 3)$

(b) $P(X > 1.5)$

(c) $P(X = 1)$

(d) $P(2 < X < 5)$

(e) $P(X \ge 0)$

(f) $P(-1 < X < 0.5)$

(g) $P(|X| \le 2)$

(h) $P(1 \le -2X + 3 \le 8)$

SOLUTION:

Given $\mu = 1$ and standard deviation = 2. Refer to the table in Appendix-*E* for cumulative standard normal standard distribution.

$$z = \frac{x - \mu}{\sigma} = \frac{x - 1}{2}$$

then z has a standard normal distribution.

(a) $P(X \le 3) = P(Z \le 1) = \Phi(1) = 0.841345$

(b) $P(X > 1.5) = P(Z > 0.25) = 1 - \Phi(0.25) = 0.4013$

(c) $P(X = 1) = 0$, because X has a continuous distribution.

(d) $P(2 < X < 5) = P(0.5 < Z < 2) = \Phi(2) - \Phi(0.5) = 0.2858$

(e) $P(X \ge 0) = P(Z \ge -0.5) = P(Z \le 0.5) = \Phi(0.5) = 0.6915$

(f) $P(-1 < X < 0.5) = P(-1 < Z < -0.25) = P(0.25 < Z < 1) = \Phi(1) - \Phi(0.25) = 0.2426$

(g) $P(|X| \le 2) = P(-2 \le X \le 2) = P(-1.5 \le Z \le 0.5) = P(Z \le 0.5) - P(Z \le -1.5) = P(Z \le 0.5) - P(Z \le 1.5)$
$= \Phi(0.5) - [1 - \Phi(1.5)] = 0.6247$

(h) $P(1 \le -2X + 3 \le 8) = P(-2 \le -2X \le 5) = P(-2.5 \le X \le 1) = P(-1.75 \le Z \le 0) = P(0 \le Z \le 1.75)$
$= \Phi(1.75) - \Phi(0) = 0.4599$

Example E3.54

Assume X is normally distributed with a mean of 10 and a standard deviation of 2. Determine the following:

(a) $P(X < 13)$

(b) $P(X > 9)$

(c) $P(6 < X < 14)$

(d) $P(2 < X < 4)$

(e) $P(-2 < X < 8)$

SOLUTION:

Refer to the table in Appendix-*E*.

(*a*) $P(X < 13) = P\left(Z < \dfrac{13-10}{2}\right) = P(Z < 1.5) = 0.93319$

(*b*) $P(X > 9) = 1 - P(X < 9) = 1 - P(Z < -0.5) = 0.69146$

(*c*) $P(6 < X < 14) = P\left(\dfrac{6-10}{2} < Z < \dfrac{14-10}{2}\right) = P(-2 < Z < 2) = [P(Z < 2) - P(Z < -2)] = 0.9545$

(*d*) $P(2 < X < 4) = P\left[\dfrac{2-10}{2} < Z < \dfrac{4-10}{2}\right] = P(-4 < Z < -3) = P(Z < -3) - P(Z < -4) = 0.00132$

(*e*) $P(-2 < X < 8) = P(X < 8) - P(X < -2) = P\left(Z < \dfrac{8-10}{2}\right) - P\left(Z < \dfrac{-2-10}{2}\right)$

$= P(Z < -1) - P(Z < -6) = 0.15866$

Example E3.55

The compressive strength of samples of cement can be modeled by a normal distribution with a mean of 7000 kg/cm^2 and a standard deviation of 100 kg/cm^2.

(*a*) find the probability of a sample's strength that is less than 7250 kg/cm^2

(*b*) find the probability that a sample's strength is between 6800 and 6900 kg/cm^2

(*c*) what strength is exceeded by 95% of the samples?

SOLUTION:

Refer to the table in Appendix-*E*.

(*a*) $P(X < 7250) = P\left(Z < \dfrac{7250-7000}{100}\right) = P(Z < 2.5) = 0.99379$

(*b*) $P(6800 < X < 6900) = P\left[\dfrac{6800-7000}{100} < Z < \dfrac{6900-7000}{100}\right] = P(-2 < Z < -1)$

$= P(Z < -1) - P(Z < -2) = 0.13591$

(*c*) $P(X > x) = P\left(Z > \dfrac{x-7000}{100}\right) = 0.95$

Hence, $\dfrac{x-7000}{100} = -1.65$ and $x = 6835$.

Example E3.56

(*a*) A process manufactures ball bearings whose diameters are normally distributed with mean 3.505 cm and standard deviation 0.008 cm. Specifications call for the diameter to be in the interval 3.5 ± 0.01 cm. What proportion of the ball bearings will meet the specification?

(*b*) Suppose the process can be recalibrated so that the mean will equal to 3.5 cm, the centre of the specification interval. The standard deviation of the process remains 0.008 cm. What proportion of the diameter will meet the specifications?

SOLUTION:

Let X represents the diameter of a randomly chosen ball bearing. Then, $X \sim N(3.505, 0.008^2)$. Figure E3.56(*a*) shows the probability density function of the $N(3.505, 0.008^2)$ population. The shaded area represents $P(3.49 < X < 3.51)$, which is the proportion of ball bearings that meet the specification.

(*a*) We compute the z-scores of 3.49 and 3.51 as follows:

$$z = \frac{3.49 - 3.505}{0.008} = -1.88 \text{ and } z = \frac{3.51 - 3.505}{0.008} = 0.63$$

The area to the left of $z = -1.88$ is 0.0301. The area to the left of $z = 0.63$ is 0.7357. The area between $z = 0.63$ and $z = -1.88$ is 0.7357 – 0.0301 = 0.7056. Hence, approximately 70.56% of the diameters of the ball bearings will meet the specifications.

(*b*) The mean is 3.5000 rather than 3.505. The calculations are as follows: (see Fig. E3.56(*b*)).

$$z = \frac{3.49 - 3.50}{0.008} = -1.25 \text{ and } z = \frac{3.51 - 3.50}{0.008} = 1.25$$

The area to the left of $z = -1.25$ is 0.1056. The area of the left of $z = 1.25$ is 0.8944. The area between $z = 1.25$ and $z = -1.25$ is 0.8944 – 0.1056 = 0.7888. Hence, recalibrating will increase the proportion of diameters that meet the specification to 78.88%.

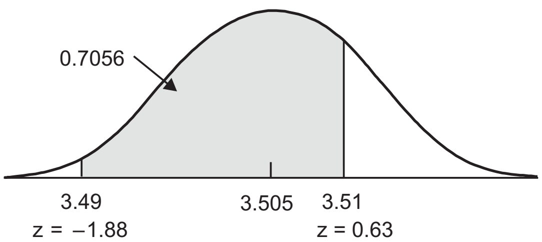


Fig. E3.56(a)

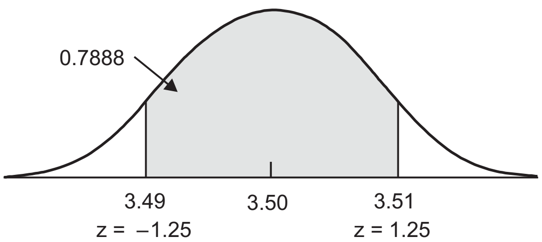


Fig. E3.56(b)

Example E3.57

Assume X is normally distributed with a mean of 5 and a standard deviation of 1. Determine the value of x that solves each of the following:

(*a*) $P(X > x) = 0.5$

(*b*) $P(X > x) = 0.95$

(*c*) $P(x < X < 5) = 0.2$

(*d*) $P[-x < (X - 5) < x] = 0.95$

(*e*) $P[-x < (X - 5) < x] = 0.99$

SOLUTION:

(a) $P(X < x) = P\left(Z > \frac{x-5}{1}\right) = 0.5$

Therefore, $\frac{x-5}{1} = 0$ and $x = 5$.

(b) $P(X < x) = P\left(Z > \frac{x-5}{1}\right) = 1 - P\left(Z < \frac{x-5}{1}\right) = 0.95$

Therefore, $P\left(Z < \frac{x-5}{1}\right) = 0.05$ and $\frac{x-5}{2} = -\,1.64$ and hence $x = 3.36$.

(c) $P(x < X < 5) = P\left(\frac{x-5}{1} < Z < 0\right) = 0.2$

Therefore, $P\left(Z < \frac{x-5}{1}\right) = 0.3$ and $\frac{x-5}{2} = -0.52$ and hence $x = 4.48$

(d) $P[5 - x < X < 5 + x] = P\left(\frac{-x}{2} < X < \frac{x}{2}\right) = 0.95$

Therefore, $x/2 = 0.98$

and $x = 1.96$

(e) $P[5 - x < X < 5 + x] = P[-x/2 < X < x/2] = 0.99$

Therefore, $x/2 = 1.79$

and $x = 2.58$.

Example E3.58

The diameter of a metal shaft for a precision instrument is assumed to be normally distributed with a mean of 0.5 mm and a standard deviation of 0.025 mm.

(a) what is the probability that shaft diameter is greater than 0.31 mm?

(b) what is the probability that shaft diameter is between 0.235 and 0.315 mm?

(c) the diameter of 90% of samples is below what value?

SOLUTION:

(a) $P(X > 0.31) = P\left(Z > \frac{0.315 - 0.25}{0.025}\right) = P(Z > 3) = 0.00135.$

(b) $P(0.235 < X < 0.315) = P(-0.6 < Z < 2.6) = P(Z < 2.6) - P(Z < -0.6) = 0.99534 - 0.27425 = 0.72109.$

(c) $P(X < x) = P\left(Z < \frac{x - 0.25}{0.025}\right) = 0.90$

Hence, $\frac{x - 0.25}{0.025} = 1.28$

and $x = 0.282$

Example E3.59

The length of a metal rod used in a machine system is normally distributed with a length of 45.10 mm and a standard deviation of 0.05 mm.

(a) what is the probability that a rod is longer than 45.15 mm or shorter than 49.85 mm?

(b) what should the process mean be set as to obtain the greatest number of rods between 44.85 and 45.15 mm?

(c) if the rods that are not between 44.85 and 45.15 mm are scrapped, what is the yield for the process mean that one would select in part *(b)*?

Solution:

(a) $$P(45.15 < X) = P\left[\left(\frac{45.15 - 45.10}{0.05}\right) < Z\right] = P(1 < Z) = 1 - 0.084134 = 0.15866$$

$$P(X < 49.85) = P\left(Z < \frac{49.85 - 45.10}{0.05}\right) = P(Z < -5) = 0$$

Hence, the answer is 0.159.

(b) The process mean should set at the centre of the specifications, that is, at $\mu = 45.0$

(c) $$P(44.85 < X < 45.15) = P\left[\frac{44.85 - 45.0}{0.05}\right] < Z < \left[\frac{45.15 - 45.0}{0.05}\right] = P(-3 < Z < 3) = 0.9973$$

Example E3.60

Refer to Example E3.59. If the process is centred so that the mean is 45 mm and the standard deviation is 0.05 mm and that 10 rods are measured and are assumed to be independent, determine

(a) the probability that all 10 rods measured are between 45.85 and 45.15 mm

(b) the expected number of the 10 rods that are between 44.85 and 45.15 mm.

Solution:

(a) $P(44.85 < X < 45.15) = 0.9973$ from Example E3.59. Therefore, by independence, the probability of 10 rods are within the given limit is $0.9973^{10} = 0.9733$.

(b) Let Y denotes the number of rods from the 10 selected that are within the given limits. Then, Y is binomial with $n = 10$ and $p = 0.9973$.

Hence, $E(Y) = 9.973$.

Example E3.61

The weight of a electronic component is normally distributed with a mean of 6 ounces and a standard deviation of 0.25 ounce.

(a) find the probability that the electronic component weighs more than 6.5 ounces.

(b) what must the standard deviation of weight be in order for the company to state that 99% of its electronic components are less than 6.5 ounces?

(c) if the standard deviation stays at 0.25 ounces, what must the mean weight be in order for the company to state that 99.9% of its electronic components are less than 6.5 ounces?

SOLUTION:

(a) $P(X > 6.5) = P\left(Z > \dfrac{6.5-6}{0.25}\right) = P(Z > 2) = 0.02275$

(b) If $P(X < 13) = 0.999$, then $P\left(Z < \dfrac{6.5-6}{\sigma}\right) = 0.999$

Hence, $\dfrac{0.5}{\sigma} = 3.09$

$\sigma = 0.1618$

(c) If $P(X < 6.5) = 0.999$, then $P\left(Z < \dfrac{6.5-\mu}{0.25}\right) = 0.999$

Therefore, $\dfrac{6.5-\mu}{0.25} = 3.09$ and $\mu = 5.7275$

Example E3.62

The weight of a mechanical component is normally distributed with a mean of 22 oz and a standard deviation of 0.5 oz.

(a) what is the probability that a component weighs more than 23 oz?

(b) what must the standard deviation of weight be in order for the company to state that 99.9% of its components are less than 23 oz?

(c) if the standard deviation remains at 0.5 oz, what must the mean weight be in order for the company to state that 99.9% of its components are less than 23 oz?

SOLUTION:

Refer to the table in Appendix-*E*.

(a) $P(X > 23) = P\left(Z > \dfrac{23-22}{0.5}\right) = P(Z > 2) = 1 - P(Z \le 2) = 1 - 0.977250 = 0.022750.$

(b) If $P(X < 23) = 0.999$, then $P\left[Z < \dfrac{23-22}{\sigma}\right] = 0.999$

Hence, $\dfrac{1}{\sigma} = 3.09$ and $\sigma = \dfrac{1}{3.09} = 0.324$

(c) If $P(X < 23) = 0.999$, then $P\left[Z < \dfrac{23-\mu}{0.5}\right] = 0.999$

Hence, $\dfrac{23-\mu}{0.5} = 3.09$

or $\mu = 21.455.$

Example E3.63

Refer to the Example E3.56. Assume that the process has been recalibrated so that mean diameter is now 3.5 cm. To what value must the standard deviation be lowered so that 95% of the diameters will meet the specifications?

Solution:

The specification interval is 3.49 — 3.51 cm. We must find a value for σ so that this interval spans the middle 95% of the population of ball bearing diameters, as shown in Fig. E3.63. The z-score that has 2.5% of the area to its left is $z = -1.96$. The z-score that has 2.5% of the area to its right is $z = 1.96$ (from the symmetry of the curve). It follows that the lower specification limit, 3.49, has a z-score of –1.96, while the upper limit of 3.51 has a score of 1.96. Either of these facts may be used to find σ.

Now $$z = \frac{x-\mu}{\sigma}$$

or $$1.96 = \frac{3.51-3.50}{\sigma}$$

or $$\sigma = 0.005102 \text{ cm}$$

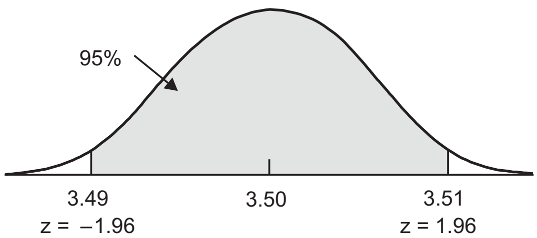


Fig. E3.63

Example E3.64

An experiment needs a 2.41 cm diameter steel rod. Suppose that the diameter of a steel rod has a normal distribution with a mean of 2.41 cm and a standard deviation of 0.01 cm

(a) determine the probability that a diameter is greater than 2.42 cm

(b) what diameter is exceeded by 95% of the samples?

(c) if the specifications require that the diameter is between 2.39 cm and 2.43 cm, what proportion of the samples meet specifications?

Solution:

Refer to the table in Appendix-*E.*

(a) Let X denotes the diameter of the steel rod.

$X \sim N(2.41, 0.01^2)$

$$P(X > 2.42) = 1 - P(X \le 2.42) = 1 - \Phi\left[\frac{2.42-2.41}{0.01}\right] = 1 - \Phi(1) = 0.1587$$

(b) $P(X > x) = P\left[Z > \frac{x - 2.41}{0.01}\right] = 0.95$

Hence, $\frac{x - 2.41}{0.01} = 1.645$ and $x = 2.3936$ cm

(c) $P(2.39 < X < 2.43) = \Phi\left[\frac{2.43 - 2.41}{0.01}\right] - \Phi\left[\frac{2.39 - 2.41}{0.01}\right] = \Phi(2) - \Phi(-2) = 0.9545$

3.5 APPROXIMATING PROBABILITY DISTRIBUTIONS

In some engineering quality control applications, it is quite useful to approximate one probability distribution by another. Figure 3.8 shows the approximation guide, that is, the conditions under which one distribution may be approximated by another.

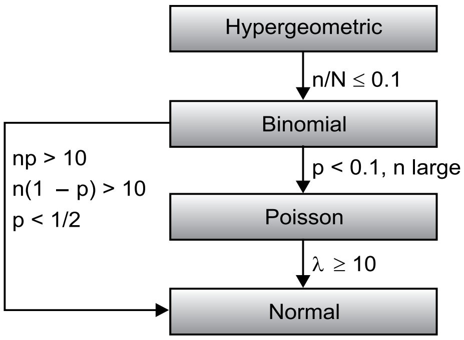


Fig. 3.8: Approximation guide

3.5.1 Binomial Approximation to the Hypergeometric

When the ratio n/N is small, that is, say $n/N \leq 0.1$, then the binomial distribution parameter $p = D/N$ can be used as a good approximation to the hypergeometric distribution. The smaller the function n/N, the better the approximation.

3.5.2 Poisson Approximation to the Binomial

The Poisson distribution can be used as an approximation to the binomial distribution when the sample size n is large and the probability of success p is small (the same applies when q is small, p and q being of course interchangeable), i.e., when the binomial distribution is highly skewed. As a guide, we can say that a good approximation is obtained when $n \geq 20$ and $p \leq 0.05$, and the approximation improves with a decrease in p.

The advantage of using the Poisson distribution to approximate the binomial distribution is that the Poisson distribution, having only one parameter, is well tabulated.

When the parameter is small, say, $p < 0.1$, and n is large, with $\lambda = np$ constant, the Poisson distribution is used as an approximation to the binomial distribution. The larger the value of n and the smaller the value of p, the better the approximation.

3.5.3 Normal Approximation to the Binomial

Recalling that if $X \sim \text{Bin}(n, p)$, then $X = Y_1 + Y_2 + \cdots + Y_n$, where $Y_1 + Y_2 + \cdots + Y_n$ is a sample from a Bernoulli (p) population. Hence, X is the sum of the sample observations.

The sample proportion is

$$\hat{p} = \frac{X}{n} = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$$

which is also the sample mean $\overline{Y}$. The Bernoulli population has mean $\mu = p$ and variance $\sigma^2 = p(1 - p)$. It follows from the central limit theorem (see section 3.8) that if the number of trials n is large, then $X \sim N(np, np(1 - p))$, and $\hat{p} \sim N(p, p(1 - p/n))$.

In the binomial case, the accuracy of the normal approximation depends on the mean number of successes np and the mean number of failures $n(1 - p)$. The larger the value of np and $n(1 - p)$, the better the approximation. Usually, the normal distribution is used as an approximation to the binomial distribution when np and $n(1 - p)$ are both greater than 5. A better and more conservative rule is to use the normal approximation whenever $np > 10$ and $n(1 - p) > 10$.

Hence if $X \sim \text{Bin}(n, p)$ and if $np > 10$ and $n(1 - p) > 10$, then

$$X \sim N(np, np(1-p)) \qquad \text{approximately}$$

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right) \qquad \text{approximately}$$

Hence, if X is a binomial random variable with parameter n and p

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

is approximately a standard normal random variable.

If n is large, then one can justify the approximation of the binomial distribution by the normal distribution with mean np and variance $np(1 - p)$. Noting that the binomial distribution is discrete and the normal distribution is continuous, we can make a continuity correction such that

$$p(X = a) = \Phi\left[\frac{(a + 1/2) - np}{\sqrt{np(1-p)}}\right] - \Phi\left[\frac{(a - 1/2) - np}{\sqrt{np(1-p)}}\right] \tag{3.82}$$

where $\Phi(\cdot)$ denotes the cdf of the standard normal distribution. We also write probability statements as

$$P(a \le X \le b) = \Phi\left[\frac{(b + 1/2) - np}{\sqrt{np(1-p)}}\right] - \Phi\left[\frac{(a - 1/2) - np}{\sqrt{np(1-p)}}\right] \tag{3.83}$$

Although p should be about 0.5, the approximation can be used without excessive loss of accuracy for $0.1 \le p \le 0.9$. If p is close to 0.5 and $n > 10$, the approximation is fairly good. For other values of p, the value of n should be larger.

In general, the approximation is good as long as $np > 5$ for $p \le 0.5$ or when $nq > 5$ when $p > 0.5$.

The normal distribution is also used to approximate the sample fraction nonconforming, $\hat{p}$. The random variable $\hat{p}$ is normally distributed with mean p and variance $p(1 – p)/n$, so that we have

$$p\left[\frac{a}{n} \le \frac{X}{n} \le \frac{b}{n}\right] = P\left[\frac{a}{n} \le \hat{p} \le \frac{b}{n}\right]$$

which indicates that we divide all terms in Eq.(3.83) by n to obtain

$$P\left(\frac{a}{n} \le \hat{p} \le \frac{b}{n}\right) = \Phi\left[\frac{\frac{b+1/2}{n} - p}{\sqrt{\frac{p(1-p)}{n}}}\right] - \Phi\left[\frac{\frac{a-1/2}{n} - p}{\sqrt{\frac{p(1-p)}{n}}}\right] \tag{3.84}$$

Continuity Correction: The binomial distribution is discrete, while the normal distribution is continuous. The *continuous correction* is an adjustment, made when approximating a discrete distribution with a continuous one that can improve the accuracy of the approximation.

3.5.4 Normal Approximation to the Poisson

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

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

is approximately a standard normal random variable. The approximation is good for $\lambda > 10$.

The normal distribution with mean $\mu = \sigma^2 = \lambda$ can be used as an approximation to the Poisson distribution if the mean λ of the distribution is large, say 10 or more.

The continuity correction used is given by

$$P(a \le X \le b) = \Phi\left[\frac{(b+1/2) - \lambda}{\sqrt{\lambda}}\right] - \Phi\left[\frac{(a-1/2) - \lambda}{\sqrt{\lambda}}\right] \tag{3.85}$$

Example E3.65

The number of widgets made per shift is equal to 1000. It is known that 8% of them are nonconforming. A sample of 20 widgets is taken. What is the probability that 2 or fewer widgets in the sample are nonconforming?

(a) set up the equation using the hypergoemetric distribution
(b) solve it using the binomial approximation
(c) solve it using the Poisson approximation.

Solution:

$$D = 80, N = 1000, n = 20, p = 0.08$$

(a) $$P(X \le 2) = \sum_{x=0}^{2} \frac{\binom{80}{x}\binom{920}{20-x}}{\binom{1000}{20}}$$

(*b*) $$P(X \le 2) = \sum_{x=0}^{2} \binom{20}{x} (0.08)^x (0.92)^{20-x}$$

$$= \binom{20}{0}(0.08)^0(0.92)^{20} + \binom{20}{1}(0.08)^1(0.92)^{19} + \binom{20}{2}(0.08)^2(0.92)^{18}$$

$$= 0.189 + 0.328 + 0.271 = 0.788$$

(*c*) $\lambda = np = 20(0.08) = 1.6$

$P(x \le 2) = 0.783$

Example E3.66

If 20% of the electronic components made in a company are nonconforming, find the probability that in a random sample of 100 such electronic components selected and using normal approximation.

(*a*) at most 16 will be nonconforming

(*b*) exactly 16 will be nonconforming

(*c*) between 16 and 22 will be nonconforming.

Solution:

Normal approximation

$$np = 100(0.20) = 20 \cdot \sqrt{np(1-p)} = 4$$

(*a*) $$P(x \le 16) = \Phi\left[\frac{16.5-20}{4}\right] = \Phi(-0.809) = 1 - 0.809 = 0.191$$

(*b*) $$P(x = 16) = \Phi\left[\frac{16.5-20}{4}\right] - \Phi\left[\frac{15.5-20}{4}\right] = 0.191 - \phi(-1.125) = 0.191 - (1 - 0.855) = 0.046$$

(*c*) $$P(16 \le x \le 22) = \Phi\left[\frac{22.5-20}{4}\right] - \Phi\left[\frac{15.5-20}{4}\right] = \phi(0.65) - 0.14 = 0.734 - 0.145 = 0.589$$

Example E3.67

Use the normal approximation to the binomial distribution to determine the probability of getting 7 heads and 7 tails in the 14 tosses of a balanced coin.

Solution:

We must thus determine the area under the curve between 6.5 and 7.5. $n = 14$ and $x = 7$, $\mu = 14\left(\frac{1}{2}\right) = 7$ and

$$\sigma = \sqrt{14\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = \sqrt{3.5} = 1.871$$

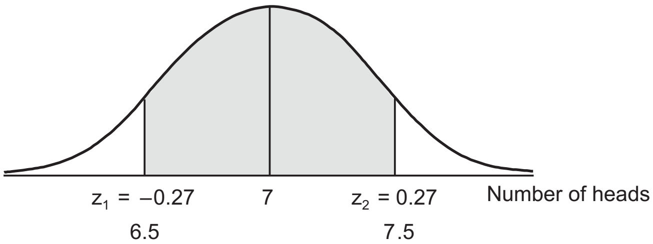


Fig. E3.67

$$z_1 = \frac{6.5-7}{1.871} = -0.27 \text{ and } z_2 = \frac{7.5-7}{1.871} = 0.27$$

The table for normal distribution in Appendix-*E* corresponding to $z = -0.27$ and 0.27 are 0.393580 and 0.60642 and we find that the normal approximation to the probability of "7 heads and 7 tails" is 0.60642 – 0.393580 = 0.2128.

The table for binomial approximation in Appendix-*C* gives 0.2095. Hence, the error of the approximation is

0.2095 – 0.2128 = –0.0033 and the percentage error is $\frac{0.0033}{0.2095}$ (100) = 1.5752% in the absolute value.

Example E3.68

Fifteen per cent of the parts produced in a manufacturing process fail a standardized quality test.

(a) using the Poisson formula, find the probability that in a random sample of 100 parts which went through this test exactly 20 will fail.

(b) using the Poisson probabilities table, find the probability that the number of parts which fail this test in a randomly selected 100 parts is

(i) at most 9

(ii) 10 to 16

(iii) at least 20.

Solution:

Let x denotes the number of parts in a random sample of 100 which fail the test. Since, on average, 15% of the parts fail the test, $\lambda = 0.15 \times 100 = 15$.

(a) $P(\text{exactly 20 fail}) = P(x = 20) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(15)^{20} e^{-15}}{20!} = 0.0418$

(b) *(i)* $P(\text{at most 9 fail}) = P(x \le 9) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5) + P(6) + P(7) + P(8) + P(9)$

$= 0.0 + 0.0 + 0.0 + 0.0002 + 0.0006 + 0.0019 + 0.0048 + 0.0104 = 0.0194 + 0.0324 = 0.0697$

(ii) $P(\text{10 to 16 fail}) = P(10 \le x \le 16) = P(10) + P(11) + P(12) + P(13) + P(14) + P(15) + P(16)$

$= 0.0486 + 0.0663 + 0.0829 + 0.0956 + 0.1024 + 0.01024 + 0.0960 = 0.5942$

(iii) $P(\text{at least 20}) = P(x \ge 20) = P(20) + P(21) + P(22) + \cdots$

$= 0.0418 + 0.299 + 0.0204 + 0.0133 + 0.0083 + 0.0050 + 0.0029 + 0.0016 + 0.0009$

$+ 0.0004 + 0.0002 + 0.0001 + 0.0001 = 0.1249$

Example E3.69

If the probability is 0.10 that a certain mortgage company will refuse a loan application, use the normal approximation to determine the probability that the mortgage company will refuse at most 40 of 450 mortgage loan applications.

SOLUTION:

$$n = 450, p = 0.10, \mu = np = 450(0.1) = 45, \ \sigma = \sqrt{np(1-p)} = \sqrt{450(0.1)(0.9)} = \sqrt{40.5} = 6.3640$$

$$z = \frac{40.5 - 45}{6.3640} = -0.7071$$

Referring to the table for normal distribution in Appendix-*E* corresponding to $z = -0.7071$, we find that the normal approximation to the probability is 0.22065.

Hence, the probability that the mortgage company will refuse the load application is 0.22065.

Example E3.70

The number of nonconforming manufactured parts per shift is Poisson distributed, with a mean of 16. Find the probability that there are between 14 and 20 nonconforming parts on a shift that is 14, 15, 16, 17, 18, 19 or 20.

(*a*) use the Poisson distribution

(*b*) use the normal approximation.

SOLUTION:

(*a*) $\lambda = 15$

$P[14 \le x \le 20] = F(20) - F(13) = 0.917 - 0.363 = 0.554$

(*b*) $P[14 \le x \le 20] = \phi\left[\frac{20.5-15}{4}\right] - \phi\left[\frac{13.5-15}{4}\right] = \phi(1.375) - \phi(-0.375) = 0.9154 - 0.3538 = 0.5616$

Example E3.71

Suppose we want to use the normal approximation to the binomial distribution to determine $B(1; 150, 0.05)$

(*a*) is it justified in using the approximation?

(*b*) make the approximation.

SOLUTION:

(*a*) Here $n = 100$, $p = 0.05$ and mean = 150(0.05) = 75 and $n(1-p) = 150 - 7.5 = 142.5$

Yes. It is justified in using the approximation since np is > 5 and $n(1-p)$ is > 100.

(*b*) $\mu = 7.5$, $\sigma^2 = np(1-p) = 150(0.05)(0.95) = 7.125$ or $\sigma = 2.67$.

$$z_1 = \frac{0.5 - 7.5}{2.67} = -2.62 \text{ and } z_2 = \frac{1.5 - 7.5}{2.67} = -2.25$$

The entries in table for normal distribution in Appendix-*E* corresponding to $z_1 = -2.62$ and $z_2 = -2.25$, we find that the normal approximation to the probability is

$$0.012224 - 0.004396 = 0.007828$$

Example E3.72

Refer to Example E3.71. Make the Poisson approximation.

Solution:

Here $\lambda = 7.5$ and $P(7.5) = \dfrac{e^{-\lambda}\lambda^x}{x!} = \dfrac{e^{-7.5}(7.5)^1}{1!} = (0.00055)7.5 = 0.0041$

Example E3.73

In a survey it was found that 1.5% of the adult population in a large city suffer from drug addition. Of 100 randomly selected adult population, let X denotes the number who have drug addiction problem.

(a) what are the parameters for the appropriate normal distribution?

(b) what are the parameters for approximating Poisson distribution?

(c) compute the individual probabilities for the binomial distribution in part *(a)*

(d) compute the individual probabilities for the Poisson distribution in part *(b)*. Find the probabilities until they are zero to 4 decimal places.

(e) compare the probabilities found in parts *(c)* and *(d)*.

(f) apply both the binomial probabilities and Poisson probabilities found in parts (*c*) and *(d)* to determine the probability that the number who suffer from drug addiction is exactly 3; between 2 and 5 (inclusive); less than 4% of those surveyed; more than 2. Compare the two results in each case.

Solution:

(a) Here $n = 100$ and $p = 1.5/100 = 0.015$

(b) $\lambda = np = 100(0.015) = 1.5$

(c) The binomial probability function is

$$P(X = x) = \binom{100}{x} (0.015)^x(0.985)100^{-x} \qquad \text{for } x = 0, 1, 2, \ldots, 100$$

Starting at $x = 0$ and evaluating the function until it is zero to four decimal places, we obtain the following Table E3.73(a).

Table E3.73(a)

x	P(X = x)
0	0.2206
1	0.3360
2	0.2532
3	0.1260
4	0.0465
5	0.0136
6	0.0033
7	0.0007
8	0.0001
9	0.0000

(*d*) The Poisson probability function is

$$P(X = x) = e^{1.5}\frac{1.5^x}{x!} \qquad \text{for } x = 0, 1, 2, \ldots$$

Starting at $x = 0$ and evaluating this function until it is zero to four decimal places, we obtain the following Table E3.73(b).

Table E3.73(b)

x	P(X = x)
0	0.2231
1	0.3347
2	0.2510
3	0.1255
4	0.0471
5	0.0141
6	0.0035
7	0.0008
8	0.0001
9	0.0000

(*e*) All the probabilities for $x = 0$ to 9 agree to 2 decimal places and are usually at most a couple of thousandths apart. The agreement is really good.

(*f*) See Table E3.73(c).

Table E3.73(c)

Event	Binomial probability	Poisson probability
X = 3	0.1260	0.1255
2 ≤ X ≤ 5	0.4393	0.4377
X < 4	0.9358	0.9343
X > 2	0.1902	0.1912

Each pair of probabilities in the above Table E3.73(c) agree to within 0.0016 or less.

3.6 CHEBYSHEV'S THEOREM

The standard deviation σ of a random variable *X* measures the weighted spread of the values of *X* about the mean μ. For smaller σ, we would expect that *X* will be closer to μ. A more precise statement of this expectation is given by Chebyshev (1821–1894) which is stated as follows:

Let *X* be a random variable with mean μ and standard deviation σ. Then, for any integer *k*, the probability that a value of *X* lies in the interval $[\mu - k\sigma, \mu + k\sigma]$ is at least $1 - \frac{1}{k^2}$.

Hence, $$P[\mu - k\sigma \le X \le \mu + k\sigma] \ge 1 - \frac{1}{k^2} \tag{3.86}$$

Proof: By definition

$$\sigma^2 = \text{Var}(X) = \Sigma(x_i - \mu)^2 p_i \tag{3.87}$$

Now, deleting all terms for which $|x_i - \mu| \le k\sigma$ and denoting the summation of the remaining terms by $\Sigma^*(x_i - \mu)^2 p_i$. then, we have

$$\sigma^2 \ge \Sigma^*(x_i - \mu)^2 p_i \ge \Sigma^* k^2\sigma^2 p_i = k^2\sigma^2\Sigma^* p_i = k^2\sigma^2 p(|X - \mu| > k\sigma) \tag{3.88}$$

or $$= k^2\sigma^2[1 - p(|X - \mu| \le k\sigma)] = k^2\sigma^2[1 - P(\mu - k\sigma \le X \le \mu + k\sigma)] \tag{3.89}$$

when $\sigma > 0$, and dividing Eq. (3.89) by $k^2\sigma^2$, we get

$$\frac{1}{k^2} \ge 1 - P(\mu - k\sigma \le X \le \mu + k\sigma) \tag{3.90}$$

or $$P(\mu - k\sigma \le X \le \mu + k\sigma) \le 1 - \frac{1}{k^2} \tag{3.91}$$

Equation (3.91) is the Chebyshev's inequality for $\sigma > 0$.

If $\sigma = 0$, then $x_i = \mu$ for all $p_i > 0$, and we have

$$P[\mu - k(0) \le X \le \mu + k(0)] = P(X = \mu) = 1 > 1 - \frac{1}{k^2} \tag{3.91a}$$

Chebyshev's theorem gives a lower bound for the area under a curve between two points that are an opposite sides of the mean and at the same distance from the mean. Chebyshev's theorem is stated as follows: For any number k greater than 1, at least $\left(1 - \frac{1}{k^2}\right)$ of the data values lie within k standard deviations of the mean. Chebyshev's theorem is illustrated in Fig. 3.9.

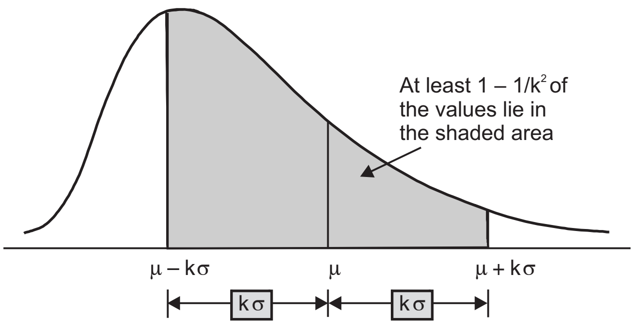


Fig. 3.9: Chebyshev's theorem

For $k = 2$

$$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 0.75 \text{ or } 75\%$$

That is, at least 75% of the values of a data set lie within two standard deviations of the mean as shown in Fig. 3.10.

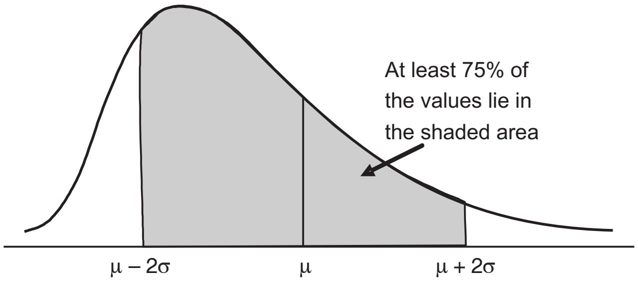


Fig. 3.10: Percentage of values within two standard deviations of the mean

For $k = 3$

$$1-\frac{1}{k^2}=1-\frac{1}{3^2}=0.89 \text{ or } 89\%$$

When $k = 3$, this inequality shows that regardless of the assumed distribution of X, the probability is no more than $1/3^2 = 1/9$ that the random variable takes on a value more than three standard deviations away from its mean (see Fig. 3.11).

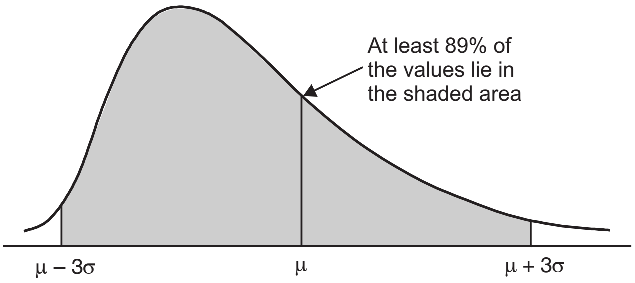


Fig. 3.11: Percentage of values within three standard deviations of the mean

Chebyshev's inequality is applicable to both sample and population data. Chebyshev's inequality is also applicable to distribution of any shape. It should be noted here that Chebyshev's inequality can be used only for $k > 1$ (since when $k = 1$, $1 - \frac{1}{k^2} = 0$ and when $k < 1$, $1 - \frac{1}{k^2}$ is negative).

Example E3.74

The mean and standard deviation for the final examination scores in *Economics* course are 80 and 7.5 respectively. Determine the percentage of students who scored between 65 and 95, using Chebyshev's theorem.

Solution:

From the given data

Mean $\mu = 80$ and standard deviation $\sigma = 10$.

Each of the two points, 60 and 95, is 15 units away from the mean.

Hence, $$k=\frac{15}{\sigma}=\frac{15}{7.5}=2$$

$$1-\frac{1}{k^2}=1-\frac{1}{2^2}=0.75 \text{ or } 75\%$$

That is, at least 75% of the students scored between 60 and 95 as shown in Fig. E3.74.

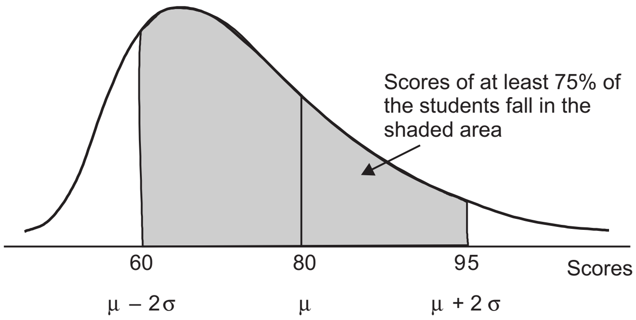


Fig. E3.74: Percentage of students who scores between 60 and 95

Example E3.75

A random variable X has mean $\mu = 35$ and standard deviation $\sigma = 2$. Apply Chebyshev's inequality to estimate

(*a*) $P(X \le 45)$

(*b*) $P(X \ge 30)$

Solution:

Chebyshev's inequality states:

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

(*a*) Substituting $\mu = 35$ and $\sigma = 2$ in $\mu + k\sigma$, we have

$$35 + 2k = 45 \text{ or } k = 5$$

Hence, $$1-\frac{1}{k^2}=1-\frac{1}{5^2}=0.96$$

Similarly, $\mu - k\sigma = 35 - 10 = 25$

Chebyshev's inequality gives

$P(25 \le X \le 45) \ge 0.96.$

The event corresponding to $X \le 45$ contains as a subset the event corresponding to $25 \le X \le 45$. Hence

$$P(X \le 45) \ge P(25 \le X \le 45) \ge 0.96$$

Therefore, the probability that X is less than or equal to 45 is at least 96%

(*b*) Substituting $\mu = 35$ and $\sigma = 2$ in $\mu + k\sigma$, we have

$35 - 2k = 30$ for k or $k = 2.5$

$$1-\frac{1}{k^2}=1-\frac{1}{2.5^2}=0.84$$

Since $\mu + k\sigma = 35 + 5 = 40$, Chebyshev's inequality gives

$$P(30 \le X \le 40) \ge 0.84$$

The event corresponding to $X \ge 30$ contains as a subset the event corresponding to $30 \le X \le 40$. Hence

$$P(X \ge 45) \ge P(30 \le X \le 40) \ge 0.84$$

Thus, the probability that X is greater than or equal to 30 is at least 84%.

Example E3.76

A random variable X has mean $\mu = 85$ and standard deviation $\sigma = 5$.

(a) what inferences can be made from Chebyshev's inequality for $k = 2$ and $k = 3$?

(b) estimate the probability that X lies between 65 and 105.

(c) determine an internal $[a, b]$ about the mean for which the probability that X lies in the interval is at least 99%.

Solution:

(a) We set $k = 2$ and obtain $\mu - k\sigma = 85 - 2(5) = 75$ and $\mu + k\sigma = 85 + 2(5) = 95$.

We can therefore conclude from Chebyshev's inequality that the probability that a value of X lies between 75 and 95 is at least $1 - \dfrac{1}{2^2} = 0.75$.

$$P(75 \le X \le 95) \ge 0.75$$

If $k = 3$, we find that the probability that X lies between 70 and 100 is at least $1 - \dfrac{1}{3^2} = 8/9 = 0.889$

(b) Here $k\sigma = 20$ since $85 - 20 = 65$ and $85 + 20 = 105$ or $k(5) = 20$ or $k = 4$.

Hence, by Chebyshev's inequality

$$P(65 \le X \le 105) \ge 1 - \frac{1}{4^2} = 0.94$$

Thus, the probability X lies between 65 and 105 is at least 94%.

(c) Setting $1 - \dfrac{1}{k^2} = 0.99$, we get $1 - 0.99 = 1/k^2$ or $k = 10$.

Hence the interval is $[85 - 10(5), 85 + 10(5)] = [35, 135]$.

3.7 EMPIRICAL RULE

If the distribution of the data is approximately bell-shaped, then empirical rule applies. The empirical rule is stated as follows:

For a bell-shaped distribution, approximately

1. 68% of the observations lie within one standard deviation of the mean.
2. 95% of the observations lie within two standard deviations of the mean.
3. 99.75% of the observations lie within three standard deviations of the mean.

The empirical rule is illustrated in Fig. 3.12. The bell-shaped distribution, also known as the normal distribution has been described earlier. The empirical rule is applicable to population data as well as sample data.

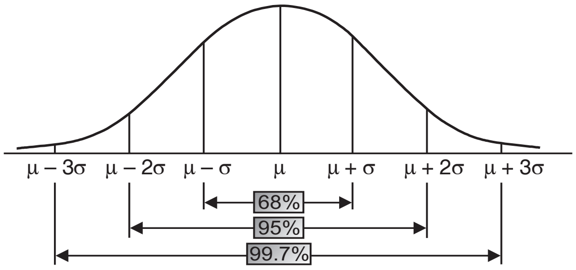


Fig. 3.12: Empirical rule

Example E3.77

The weight distribution of a sample of 10,000 male students in a particular university is bell-shaped with a mean of 40 kg and a standard deviation of 5 kg. Find the percentage of students who weigh approximately 30 to 50 kg.

SOLUTION:

From the given data, we have for the sample

Mean $\bar{x}$ = 40 kg and standard deviation s = 5 kg.

Each of the two points, 30 and 50, is 10 units away from the sample mean of 40 kg. Dividing 10 by the standard deviation (5 kg), we have $\frac{10}{\sigma} = \frac{10}{5} = 2$. The distance between 30 and 40 and between 40 and 50 kg is each equal to 2s as shown in Fig. E3.77, the area from 30 to 50 kg is the area from $\bar{x} - 2s$ to $\bar{x} + 2s$.

Applying the empirical rule, approximately 95% of the male students in the sample weigh between 30 and 50 kg.

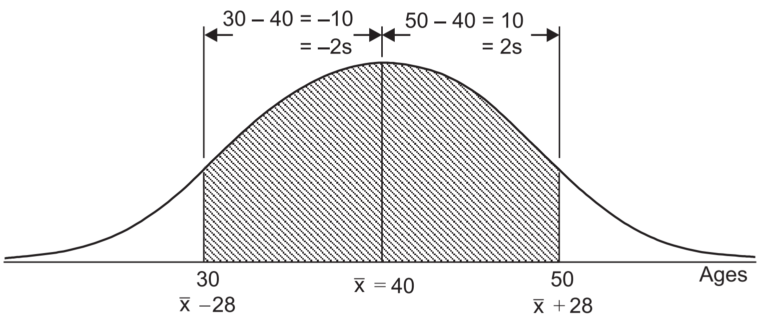


Fig. E3.77: Percentage of male students who weigh between 30 and 50 kg

3.8 THE CENTRAL LIMIT THEOREM

Many commonly used statistical methods depend on the central limit theorem. The central limit theorem is very important result in statistics. The central limit theorem states that if a large enough sample is drawn from a population, then the distribution of the sample mean is approximately normal, no matter what population the sample was drawn from. If the variable is normally distributed, so is the variable mean, $\bar{x}$. This key fact also holds approximately if x is not normally distributed, provided only that the sample size is relatively large.

Let $X_1, X_2, \ldots, X_n$ be a simple random sample from a population with mean μ and variance σ.

Let $\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ be the sample mean.

Let $S_n = X_1 + X_2 + \cdots + X_n$ be the sum of the sample observations.

Then, if n is sufficiently large,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ approximately}$$

and $$S_n \sim N(n\mu, n\sigma^2) \text{ approximately}$$

The statement of the *central limit theorem* specifies that $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}^2 = \dfrac{\sigma^2}{n}$, which hold for any sample mean. The sum of the sample items is equal to the mean multiplied by the sample size, that is, $S_n = n\bar{x}$. It follows that $\mu_{S_n} = n\mu$ and $\sigma_{S_n}^2 = n^2\sigma^2/n = n\sigma^2$.

Hence, for a relatively large sample size, the variable is approximately normally distributed, regardless of the distribution of the variable under consideration. The approximation becomes better with increasing sample size.

The further the variable under consideration is from being normally distributed, the larger the sample size must be for a normal distribution to provide an adequate approximation to the distribution of $\bar{x}$. Generally, however, a sample size of 30 or more ($n \geq 30$) is large enough.

The proof of the central limit theorem is difficult and very complex to be included here.

Summarising, the central limit theorem says that if $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\bar{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

as $n \to \infty$, is the standard normal distribution. The normal approximation for $\bar{X}$ depends on the sample size n. In many cases of practical interest, if $n \geq 30$, the normal approximation will be satisfactory regardless of the shape of the population. If $n < 30$, the central limit theorem will work if the distribution of the population is not severely non normal.

Example E3.77

A mechanical component has tensile strength that is normally distributed with mean 75.5 MPa and standard deviation 3.5 MPa. Find the probability for a random sample of $n = 6$, these component specimens will have sample mean tensile strength that exceeds 75.75 MPa.

SOLUTION:

Given $\mu_{\bar{X}} = 75.5$ MPa

Therefore, $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{3.5}{\sqrt{6}} = 1.429$

$$P(\bar{X} \geq 75.75) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right) = P(Z \geq 0.175) = 1 - P(Z \leq 0.175)$$

$$= 1 - 0.56945 = 0.43055$$

Example E3.78

A normal population has mean 80 and variance 25. How longer must the random sample be if we want the standard error of the sample average to be 1.5?

SOLUTION:

Given $\sigma^2 = 25$ or $\sigma = 5$ and $\sigma_{\bar{X}} = 1.5$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{n}} = 1.5$$

or $$n = \left(\frac{\sigma}{\sigma_{\bar{X}}}\right)^2 = \left(\frac{5}{1.5}\right)^2 = 11.11 \text{ or} \approx 12$$

PROBLEMS

Section 3.1 Random Variables

P3.1 The frequency distribution of the number of orders received per day by a distribution company during the past 3 months (90 days)

Number of orders received per day	2	3	4	5	6
Number of days	11	20	23	22	14

(a) construct a probability distribution table for the number of orders received per day

(b) are the probabilities in *(a)* exact or approximate?

(c) if x denotes the number of orders received on any given day, calculate the following probabilities

(i) $P(x = 3)$

(ii) $P(x \geq 3)$

(iii) $P(2 \leq x \leq 4)$

(iv) $P(x < 4)$.

P3.2 The following table gives the probability distribution of a discrete random variable x.

x	0	1	2	3	4	5	6
P(x)	0.11	0.14	0.18	0.15	0.12	0.09	0.05

Find the following probabilities:

(*a*) $P(x = 4)$

(*b*) $P(x \geq 3)$

(*c*) $P(x < 3)$

(*d*) $P(3 \leq x \leq 5)$

P3.3 The following table gives the probability distribution of automobiles sold on a given day in an automobile dealership.

Automobile sold	0	1	2	3	4	5	6
Probability	0.05	0.10	0.20	0.30	0.20	0.15	0.05

Calculate the mean and standard deviation for this probability distribution.

P3.4 Table P3.4 gives the probability distribution where X represents the number of times the manufacturing process is recalibrated during a week whenever the quality of the component produced falls below certain specifications. Assume that X has the probability mass function as given in Table P3.4.

Table P3.4

x	0	1	2	3	4
p(x)	0.35	0.25	0.20	0.15	0.05

(*a*) determine the mean of X

(*b*) determine the variance and standard deviation for the random variable X.

P3.5 Table P3.5 shows the probability distribution of the number of breakdowns per week for a machine.

Table P3.5

Breakdowns per week	0	1	2	3
Probability	0.15	0.20	0.35	0.30

(*a*) represent the probability graphically

(*b*) find the probability that the number of breakdowns for this machine during a given week is

(*i*) exactly 2

(*ii*) 0 to 2

(*iii*) more than 1

(*iv*) at most 1

P3.6 For each of the following, indicate whether it is or is not a random variable.

(a) the number of automobiles sold at a dealership during a given month

(b) the height of a person

(c) the number of heart beats per minute

(d) the price of an automobile

(e) the time needed to complete a written examination

(f) the number of automobiles a household owns

(g) the number of tails obtained in four tosses of a balanced coin.

P3.7 The number of telephone calls received in a particular sales office between 10:00 am and 11:00 am has the probability function given in Table P3.7.

Table P3.7

Number of telephone calls received, x	0	1	2	3	4	5	6
Probability, P(x)	0.05	0.20	0.25	0.20	0.10	0.15	0.05

(a) is this a probability function?

(b) find the probability that there will be three or more calls

(c) find the probability that there will be an even number of calls.

P3.8 Table P3.8 gives the probability distribution of x, where x denotes the number of defective mechanical components produced in a manufacturing company.

Table P3.8

x	0	1	2	3	4	5
P(x)	0.02	0.20	0.30	0.30	0.10	0.08

Determine the mean and standard deviation of x.

P3.9 Table P3.9 gives the number of telephone calls received per hour in an office and the distribution.

Table P3.9

Number of telephone calls received, x	10	11	12	13	14	15
f (x)	0.08	0.15	0.30	0.20	0.20	0.07

Determine the mean and standard deviation of the number of telephone calls received per hour in that office.

P3.10 Table P3.10 gives the results of the experiment of rolling a die with the discrete random variable number of dots.

Table P3.10

Number of dots, x	1	2	3	4	5	6
Probability, f (x)	1/6	1/6	1/6	1/6	1/6	1/6

Determine the mean and variance of that random variable x.

P3.11 The time to failure of an electronic component is a continuous random variable known to have the density function $0.5e^{-0.5t}$ where t is in years. What is the probability that this component will fail within the first year of operation?

P3.12 Find the mean μ for the probability density function, ρ(x), of the life of a projector bulb, random variable x, is given as:

and
$$\rho(x) = \begin{cases} 0 & ; \quad \text{for } x < 0 \\ 1/900\, e^{-x/900} & ; \quad \text{for } x \geq 0 \end{cases}$$

P3.13 The density function for a continuous random variable x is given as:

$f(x) = 0.25(x - 2)$ for $2 \leq x \leq 5$

Sketch the density and distribution functions.

P3.14 The daily consumption of an electric power of a certain machine (in units of power) is a random variable whose probability density is given by:

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

Determine the probabilities that on a given day

(a) the consumption of this machine is no more than 6 units

(b) the power supply is inadequate in the daily capacity if the supply is 9 units.

P3.15 The total lifetime (in years) of a certain machine is a random variable whose distribution function is given by:

$$f(x) = \begin{cases} 0 & ; \quad \text{for } x \leq 5 \\ 1 - 25/2^2 & ; \quad \text{for } x > 5 \end{cases}$$

Find the probabilities that such a machine system will have life

(a) beyond 10 years

(b) less than 8 years

(c) anywhere from 12 to 15 years.

P3.16 The probability density function of X is given by

$$f(x) = \begin{cases} 1.25(1 - x^4) & ; \quad 0 < x < 1 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

where the random variable X denotes the clearance (in mm). The clearance is the difference between the radius of the hole drilled in a flat sheet-metal plate and a shaft inserted through the hole. Components with clearances larger than 0.8 mm are to be scrapped. What proportion of components are scrapped?

P3.17 Refer to Problem P3.16. Determine the mean clearance and the variance of the clearance.

P3.18 Determine the mean and variance for the function $f(x)$

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

where X is the uniform random variable on the interval [a, b].

P3.19 A certain radioactive mass emits alpha particles from time to time. The time between emissions (in seconds) is random with the probability density function given by

$$f(x) = \begin{cases} 0.1e^{-0.1x} & ; \ x > 0 \\ 0 & ; \ x \le 0 \end{cases}$$

Determine the median time between emissions.

P3.20 Determine the mean and variance for the following continuous random variable X with probability density function $f(x)$ given by

$$f(x) = \begin{cases} \dfrac{1}{\lambda} e^{-x/\lambda} & ; \ x \ge 0 \text{ and } \lambda > 0 \\ 0 & ; \ \text{otherwise} \end{cases}$$

Section 3.2 Permutations and Combinations

P3.21 *(a)* Determine the number of permutations of the 26 letters of the alphabet.

(b) Determine the number of ways of 6 distinct products that can be lined up on a display shelf.

P3.22 A committee is to be formed consisting of 2 men and 2 women. There are 6 men and 7 women qualified and available to fill this position. How many different committees can be formed out of the finalists?

P3.23 Develop a table of the number of possibilities when 2, 3, 4 and 5 letters of the alphabet are used. Generate the possibilities. English alphabet letters are used with and without repetition.

P3.24 How many 5 digit numbers can be formed with the 10 digits 0, 1, 2, 3, …, 9 if

(a) repetitions are allowed

(b) repetitions are not allowed

(c) the last digit must be zero and repetitions are not allowed.

P3.25 How many ways can an executive committee of 5 can be chosen from a pool of 15 members?

P3.26 A company wants to purchase 4 electronic systems. After all the system models were reviewed, 8 foreign made and 10 U.S. made systems were considered to satisfy all the security requirements for the company.

(a) if the systems are chosen at random, find the probability that 2 of the systems selected are foreign made.

(b) what is the probability that all the four systems selected are U.S. made?

(c) what is the probability that all of the 4 systems selected are foreign made?

(d) what is the probability that at least 2 of the systems are U.S. made?

Permutations

P3.27 A pianist knows five pieces but will have enough stage time to play only four of them. Pieces played in a different order constitute a different program. How many different programs can be arranged?

P3.28 There are six persons on a sinking boat. There are four life jackets on board. How many combinations of survivors are there?

P3.29 A violinist knows four pieces but will have enough time to play only three of them. Pieces played in a different order constitute a different program. Determine the number of different programs that can be arranged.

P3.30 A bin contains 5 white balls, 2 red balls, and 3 green balls. Find the probability of getting either a white ball or a red ball in one drawn from the bin.

P3.31 Five defective parts were unintentionally mixed with 45 non-defective (good) ones. After a thorough mixing, 5 parts are picked simultaneously from the collection of 50.

(a) determine the probability that all 5 parts selected are non-defective ones

(b) what is the probability that no more than 1 part selected is defective?

(c) what is the probability that all 5 parts selected are defective?

P3.32 From 12 items, in how many ways can a selection of 6 be made?

(a) when a specified item is always included

(b) when a specified item is always excluded.

P3.33 Out of 6 manufacturing engineers and 9 design engineers, a committee consisting of 3 manufacturing engineers and 5 design engineers is to be formed. In how many ways can this be done if:

(a) any manufacturing engineer and any design engineer can be included

(b) one particular design engineer must be on the committee

(c) two particular manufacturing engineers cannot be on the committee.

P3.34 Determine the number of ways of splitting 10 persons into 2 groups containing 4 and 6 persons respectively.

P3.35 A committee of 2 men and 3 women needs to be formed out of 7 men and 7 women. Determine the number of ways this can be done if

(a) any man and any woman can be included

(b) one particular woman must be on the committee

(c) two particular men cannot be on the committee.

P3.36 Three cards are drawn in succession from a deck of 52 cards. Determine the number of ways this can be done

(a) with replacement

(b) without replacement.

P3.37 Determine the number of committees of 3 that can be formed from 8 persons.

P3.38 Determine the number of combinations of 4 taken 3 at a time.

P3.39 Find the number of ways of selecting two objects out of four objects.

P3.40 If 3 persons are to be selected randomly from 5 persons for a committee, determine the different possible combinations.

Section 3.3 Discrete Distributions

P3.41 A lot contains 140 electronic components and 20 are selected without replacement for quality testing.

(a) if 20 components are defective, what is the probability that at least one defective component is in the sample?

(b) if 5 components are defective, what is the probability that at least one defective component appears in the sample?

P3.42 A lot of mechanical parts contains 24 in a package. It was determined to test 4 parts in each package and if all 4 parts pass the tests, then they are all accepted. If in a package of parts in which 3 are non-conforming, what is the probability of rejection of these parts?

P3.43 A die is rolled three times. Find the probability of getting one 4 in the three rolls. Also find the probability of getting two 4'*s*, three 4'*s*, and no 4'*s* in three rolls.

P3.44 In a production process the defective rate is 15 per cent. Assuming a random sample of 10 items is drawn from this process, find the probability that two of them are defective.

P3.45 Batches of 50 shock absorbers from a production process are tested for conformance to quality requirements. The mean number of non-conforming absorbers in a batch is 5. Assume that the number of non-conforming shock absorbers in a batch, denoted as x, is a binomial random variable.

(a) find n and p

(b) find $p(x \leq 2)$

(c) find $p(x \geq 49)$.

P3.46 A production process manufactures certain mechanical components for a machine system. On average, 1.5% of the components will not perform up to specifications. When a shipment of 100 components is received at the plant, they are tested, and if more than 2 are defective, the shipment is returned to the manufacturer. What is the probability of returning a shipment?

P3.47 A quality test engineer claims that 1 in 10 of certain manufactured parts is due to material defects. Using binomial distribution and rounding to four decimals, find the probability that at least of 3 out of 5 of the tested parts are due to material defects.

P3.48 At the ABC House Delivery Service, providing high-quality service to its customers is the top priority of the company. The company guarantees a refund of all charges if a package it is delivering does not arrive at it's destination by the specified time. It is known from past data that despite all efforts, 3% of the packages mailed through this company do not arrive at their destinations within the specified time. A corporation mailed 10 packages through ABC House Delivery Service on Monday.

(a) find the probability that exactly one of these 10 packages will not arrive at it's destination within the specified time.

(b) find the probability that at most one of these 10 packages will not arrive at it's destination within the specified time.

P3.49 Five per cent of a large batch of high-strength steel components purchased for a mechanical system are defective.

(a) if seven components are randomly selected, find the probability that exactly three will be defective

(b) find the probability that two or more components will be defective.

P3.50 The number of customers arriving a bank in the next period is a Poisson distribution having a mean of eight. Find the probability that exactly six customers will arrive in the next period.

P3.51 If the probability of a concrete beam failing in compression is 0.05, use the Poisson approximation to obtain the probability that from a sample of 50 beams

(a) at least three will fail in compression

(b) no beam will fail in compression.

P3.52 The number of defects on an electronic component which is used in a computerized system has been found to follow the Poisson distribution with $\lambda = 3$. Find the probability that a randomly selected electronic component will have two or less defects.

P3.53 The number of telephone calls made to a certain company's operator is a Poisson random variable with a mean of 5 calls per hour.

(a) what is the probability that 5 calls are received in one hour?

(b) what is the probability that 10 calls are received in 1.5 hour?

(c) what is the probability that less than 2 calls are received in 1 – 1/2 hours?

P3.54 A photocopying machine in an office breaks down an average of three times per month. Using the Poisson probability distribution formula, find the probability that during the next month this machine will have

(a) exactly two breakdowns

(b) at most one breakdown.

P3.55 The proportion of mechanical manufactured parts that are non-conforming is 0.04. Obtain the Poisson approximation to the binomial distribution for the probability of three or fewer non-conforming parts in a sample of 100.

P3.56 An examination consists of five questions, and to pass the examination a student has to answer at least four questions correctly. Each question has three possible answers, of which only one is correct. If a student guesses on each question, what is the probability that the student will pass the test?

P3.57 The probability that a person undergoes a heart operation will recover is 0.6. Determine the probability that of the six patients who undergo similar heart operation:

(a) none will recover

(b) all will recover

(c) half will recover

(d) at least half will recover.

P3.58 In a certain manufacturing process, it was found from quality control inspection that 20% of the machine components produced by the process are defective. Determine the probability that out of 4 machine components selected at random *(a)* 1, *(b)* 0, *(c)* at most 2 machine components will be defective.

P3.59 Given that the probability of an individual patient suffers a bad reaction from injection of a particular serum is 0.001. Determine the probability that out of 2000 individual patients.

(a) exactly 3 individuals will suffer a bad reaction

(b) more than 2 individuals will suffer a bad reaction. Use Poisson distribution.

P3.60 Use binomial distribution and repeat Problem 3.59.

Section 3.4 Continuous Probability Distributions

P3.61 The inside diameter of a finished shaft of uniform diameter is normally distributed with a mean of 4.50 cm and a standard deviation of 0.01 cm. What is the probability of obtaining a diameter exceeding 4.52 cm?

P3.62 The resistance of a foil strain gauge is normally distributed with a mean of 100 ohms and a standard deviation of 0.8 ohm. The specification limits are 100 ±1.0 ohms. What percentage of gauges will be defective?

P3.63 The measurement of the diameter of a special steel pipe is normally distributed with a mean of 5.01 cm and standard deviation of 0.03 cm. The specification limits are 5.00 ±0.05 cm. What percentage of pipes is not acceptable?

P3.64 The yield strength of metal component manufactured is $N(192.36)$. A purchaser of the metal components requires strength of at least 180 psi. The probability that a metal component will meet or exceed the specifications is given by $P(x \geq 180)$. Determine the probability that a metal component meets or exceeds the specifications.

P3.65 Find the probability that the yield strength of the metal components in Problem P3.64 is between 180 and 198 psi.

P3.66 The diameter of a machine shaft produced in a manufacturing company is normally distributed with a mean diameter of 0.001 inches and a standard deviation of 0.0002 inches.

(a) what is the probability that the diameter of the shaft exceeds 0.0013 inches?

(b) what is the probability that a diameter of the shaft is between 0.0007 and 0.0013 inches?

(c) what standard deviation of diameters is needed so that the probability in part (*b*) is 0.995?

P3.67 The mass, m, of a particular machine part is normally distributed with a mean of 66 kg and a standard deviation of 5 kg.

(a) what percentage of the parts will have a mass less than 72 kg?

(b) what percentage of parts will have a mass in excess of 72 kg?

(c) what per cent of the parts will have a mass between 61 and 72 kg?

P3.68 It was determined experimentally that the load X required to break a plate is normally distributed with mean 2.5 and standard deviation 0.24. Determine the probability that

(a) the plate breaks at a load of 2.61 or less

(b) the plate breaks at a load of more than 2.39

(c) the breaking load is at least 2.86

(d) the breaking load is in the interval (2.61, 2.86).

P3.69 Use the Table in Appendix-*E* to determine the following probabilities for the standard normal random variable z.

(a) $P(z < 3.0)$

(b) $P(z > -2.15)$

(c) $P(-3 < z < 3)$

(d) $P(0 < z < 1)$

(e) $P(z > 3)$

P3.70 Assume z has a standard normal distribution. Use Appendix-*E* table to determine the value for z that solves each of the following:

(a) $P(Z < z) = 0.5$

(b) $P(Z > z) = 0.9$

(c) $P(-z < Z < z) = 0.99$

(d) $P(-z < Z < z) = 0.68$

(e) $P(-1.24 < Z < z) = 0.8$

P3.71 The diameter of a component in a machine system is normally distributed with mean 0.2508 cm and standard deviation 0.0005 cm. The specifications on the component are 0.25 ±0.0015 cm. Determine the proportion of components confirms to specifications.

P3.72 The mass, μ, of a particular electronic component is normally distributed with a mean of 66 g and a standard deviation of 5 g. Determine

(*a*) the per cent of components that will have a mass less than 72 g

(*b*) the per cent of components that will have a mass in excess of 72 kg

(*c*) the per cent of components that will have a mass between 61 and 72 g.

P3.73 Customers buying copper rods supplied by a certain manufacturer require that the rods be between 9.9 cm and 10.5 cm, inclusive. The manufacturing process is such that the actual rod lengths are well approximated by a normal distribution with mean 10.1 cm and standard deviation 0.20 cm. Determine the percentage of the manufacturer's production is acceptable to the customer.

P3.74 Refer to Problem P3.73 and determine what rod length is exceeded by 95% of the manufacturer's product.

Section 3.5 Approximating Probability Distributions

P3.75 In a production lot of 100 components, five of them are found to be non-conforming. Approximate the probability that a random sample of 10 components contains no more than 1 non-conforming component (hypergeometric distribution) by the binomial distribution.

P3.76 The proportion of components manufactured that are non-conforming is 0.04. The probability of three of fewer non-conforming components in a random sample of 100 is given by the binomial distribution. Make an approximation using Poisson distribution.

P3.77 Out of a batch of 2800 electronic components received 25% of them were defective or non-conforming. In a sample of 50 randomly selected, the probability that between 12 and 14 are given by the binomial distribution. Use normal distribution to this binomial distribution.

P3.78 A lot of electronic components are known to contain $p = 0.12$ non-conforming. In a sample of 100, the probability that the fraction non-conforming is $0.10 \le p \le 0.20$ is given by the normal distribution. Make an approximation.

P3.79 The number of telephone calls received in a specified time by an operator in a manufacturing company is Poisson distribution with mean $\lambda = 14$. Determine the probability that between 10 and 18 calls are received by the operator in that specified time using normal approximation.

P3.80 Determine the probability that in a sample of 10 machine components chosen at random, exactly two will be defective by using

(*a*) the binomial distribution

(*b*) the Poisson approximate to the binomial distribution.

Given that 10% of the machine components produced in that manufacturing process are defective.

P3.81 In some manufacturing process that a production of 200 components contains 5 components that do not meet the quality specifications. Determine the probability that a random sample of 10 components will contain no non-conforming components. Use binomial approximation to the hypergeometric.

P3.82 Assume that an examination has 10 questions of the type true or false. If the student taking such an examination guesses at all 10 questions, determine the probability that the student answers either seven or eight questions correctly.

P3.83 *(a)* If a fair coin is tossed 100 times, use the normal curve to approximate the probability that the number of heads is between 35 and 45 inclusive.

(b) If a fair coin is tossed 100 times, use the normal curve to approximate the probability that the number of heads is between 35 and 45 exclusive.

P3.84 In a certain university, 25% of the students are over 21 years of age. In a sample of 400 students, what is the probability that more than 110 of them are over 21 years of age?

P3.85 A soft drink company conducted a taste survey marketing a new soft drink. The results of this survey showed that 70% of the people who tried the drink liked it. Encouraged by this result, the company decided to market the new soft drink. Assume that 70% of all people like this drink. On a certain day, 100 customers bought this soft drink.

(a) find the probability that exactly 65 out of 100 customers will like this drink

(b) find the probability that exactly 60 or less of the 100 customers will like this drink

(c) find the probability that exactly 75 to 80 of the 100 customers will like this drink

P3.86 The length of a bolt manufactured in a certain manufacturing process has a mean of 50 mm and standard deviation of 0.45 mm. Determine the largest possible value for the probability that the length of the bolt is outside the interval 49.1 — 50.9 mm.

P3.87 In a mechanical engineering design class, the mean for the final examination scores is 75 and the standard deviation is 5. Using Chebyshev's theorem, find the percentage of students who scored between 60 and 90.

P3.88 Given that X is a random variable with the $N(\mu, \sigma^2)$ distribution. Denoting the standardised X by Z, make the comparisons between the actual and the Chebyschev's bounds for $Z = 1$, 2, and $3(k = 1$, 2, and 3).

P3.89 In mechanics class, the mean for the midterm scores is 65 and the standard deviation is 8. Using Chebyshev's theorem, find the percentage of students who scored between 49 and 81.

P3.90 The 2007 gross sales of all firms in a large city have a mean of 3.3 million and a standard deviation of 0.6 million. Using Chebyshev's theorem, find at least what percentage of firms in this city had 2007 gross sales of

(a) \$2.1 to \$4.5 million

(b) \$1.8 to \$4.8 million

(c) \$1.5 to \$5.1 million

P3.91 *(a)* Let X be a random variable with mean $\mu = 50$ and standard deviation $\sigma = 6$. Use Chebyshev's inequality to find a value for b for which $P(50 - b \le X \le 50 + b) \ge 0.95$.

(b) Let X be a random variable with mean $\mu = 70$ and unknown standard deviation σ. Use Chebyshev's inequality to find a value for σ for which $P(65 \le X \le 75) \ge 0.90$.

P3.92 The age distribution of a sample of 6000 persons is bell-shaped with a mean of 50 years and a standard deviation of 10 years. Determine the approximate percentage of people who are 30 to 70 years old.

P3.93 A manufacturing company manufactures steel rods of 100 cm length and a standard deviation of 10 cm. The distribution of the rod is normal. Find the probability that a random sample of $n = 25$ rods will have an average length less than 95 cm.

P3.94 The following table gives the probability mass function of X, where X denotes the number of defects in a 1 m length of aluminium wire. One hundred wires are sampled from this population. Find the probability that the average number of defects per wire in this sample is less than 0.5.

x	0	1	2	3
P(X = x)	0.48	0.39	0.12	0.01

P3.95 Suppose that a random sample of size $n = 12$ is taken from the uniform distribution on the interval [0, 1]. Determine the value of $P\left(\left|\bar{X}_n - \frac{1}{2}\right| \le 0.1\right)$.

P3.96 Suppose that a random variable X has a continuous uniform distribution

$$f(x) = \begin{cases} 1/2 \ ; & 4 \le x \le 6 \\ 0 \ ; & \text{otherwise} \end{cases}$$

Find the distribution of the sample mean of a random sample of size $n = 50$. Use central limit theorem.

P3.97 At a large university, the mean age of the students is 21.3 years, and the standard deviation is 4 years. A random sample of 64 students is drawn. What is the probability that the average age of these students is greater than 22 years?

P3.98 The GPAs of all 5540 students enrolled at a university have an approximate normal distribution with a mean of 3.02 and a standard deviation of 0.29. Let $\bar{x}$ be the mean GPA of a random sample of 48 students selected from this university. Determine the mean and standard deviation of $\bar{x}$ and comment on the shape of its sampling distribution.

REVIEW QUESTIONS

1. Explain the meaning of probability distribution of a discrete random variable. What are the various ways to present the probability distribution of a discrete random variable?
2. Briefly explain the two-characteristics (conditions) of the probability distribution of a discrete random variable.
3. Define the terms mean and standard deviation of a discrete random variable.
4. Briefly explain the following:
   (a) binomial experiment
   (b) trial
   (c) binomial random variable
5. Describe the parameters of the binomial probability distribution.
6. Define the following:
   (a) *n*-factorial
   (b) combinations
   (c) permutations

7. What are the conditions that must be satisfied to apply Poisson probability distribution?
8. What is the parameter of the Poisson distribution?
9. Define the following terms:
   (a) Bernoulli trial
   (b) Binomial parameters
   (c) Binomial probability distribution
   (d) continuous random variable
   (e) discrete random variable
   (f) random variable
   (g) z-value or z-score
10. Describe the difference between the probability distribution of a discrete random variable and that of a continuous random variable.
11. Explain the main characteristics of a normal distribution.
12. Describe the standard normal distribution curve.
13. Describe the parameters of the normal distribution.
14. Describe the conditions for the normal distribution to be used as an approximation to the binomial distribution.

STATE TRUE OR FALSE

1. A random variable is a quantitative variable whose value depends on chance. (True/False)
2. A discrete random variable is a random variable whose possible values cannot be listed. (True/False)
3. Probability distribution is a listing of the possible values and corresponding probabilities of a discrete random variable, or a formula for the probabilities. (True/False)
4. Probability histogram is a graph of the probability distribution that displays the possible values of a discrete random variable on the horizontal axis and the probabilities of those values on the vertical axis. (True/False)
5. For any discrete random variable, $\Sigma P(X = 1) = 1$. (True/False)
6. The term expected value and expectation are commonly used in place of the term mean. (True/False)
7. $k! = k(k-1) \cdots 2{\cdot}1$. (True/False)
8. $0! = 1$. (True/False)
9. In Bernoulli trials, the trials are not independent. (True/False)
10. In Bernoulli trials, the probability of a success, called the success probability, remains the same from trial to trial. (True/False)
11. In Bernoulli trials, the experiment (each trial) has more than two possibilities. (True/False)
12. The binomial distribution is the probability distribution for the number of failures in a sequence of Bernoulli trials. (True/False)
13. In Bernoulli trials, the number of outcomes that contain exactly x successes equals the binomial coefficient $\begin{Bmatrix} n \\ x \end{Bmatrix}$. (True/False)

14. The mean of a binomial random variable with parameters n and p is $\mu = np$. (True/False)
15. The standard deviation of binomial random variable with parameters n and p is $\sigma = \sqrt{np(1-p)}$. (True/False)
16. The mean of a Poisson random variable with parameter λ is $\mu = \lambda$. (True/False)
17. The standard deviation of a Poisson random variable with parameter λ is $\sigma = \sqrt{\lambda}$. (True/False)
18. A random variable is a quantitative variable whose value depends on chance. (True/False)
19. A discrete random variable is a random variable whose possible values cannot be listed. (True/False)
20. The sum of the probabilities of the possible values of a discrete random variable equals 0. (True/False)
21. The number of possible permutations of m objects among themselves is $m!$. (True/False)
22. The number of possible samples of size n from a population of size N is NC_n. (True/False)
23. For any two events, the probability that one or the other of the events occurs equal the sum of the two individual probabilities. (True/False)
24. For any event, the probability that it occurs equals 1 minus the probability that it does not occur. (True/False)
25. Data obtained by observing values of one variable of a population are called univariate data. (True/False)
26. Data obtained by observing values of two variables of a population are called bivariate data. (True/False)
27. A frequency distribution for bivariate data is called contingency table, or two-way table. (True/False)
28. The joint probability equals the product of the marginal probabilities. (True/False)
29. For a normally distributed variable, the percentage of all possible observations within any specified range equals the corresponding area under its associated normal curve, expressed as a percentage. (True/False)
30. A normally distributed variable having mean 0 and standard deviation 1 is said to have the standard normal distribution. (True/False)
31. A normal distribution is completely determined by the mean and standard deviation. (True/False)
32. The shape of a normal distribution is completely determined by its standard deviation.(True/False)
33. The total area under the standard normal curve is –1. (True/False)
34. The standard normal curve extends indefinitely in both directions, approaching, but never touches, the horizontal axis as it does so. (True/False)
35. The standard normal curve is symmetric about 0. (True/False)
36. Almost all the area under the standard normal curve lies between –1 and +1. (True/False)
37. For a normally distributed variable, one can determine the percentage of all possible observations that lie within any specified range by first converting z-scores and then obtaining the corresponding area under the standard normal curve. (True/False)
38. The rule of thumb for using the normal approximation to the binomial is that np and $n(1 - p)$ are 5 are greater. (True/False)

39. A variable is said to be normally distributed if its distribution has the shape of a normal curve. (True/False)
40. If a variable of a population is normally distributed and is the only variable under consideration, common practice is to say that the population is a normally distributed population. (True/False)
41. The parameters for a normal curve are the corresponding mean and standard deviation of the variable. (True/False)
42. Two variables that have the same mean and standard deviation have the same distribution. (True/False)
43. Two normally distributed variables that have the same mean and standard deviations have the same distribution. (True/False)
44. Two normal distributions that have the same mean are centred at the same place, regardless of the relationship between their standard deviations. (True/False)
45. Two normal distributions that have the same standard deviations have the same shape, regardless of the relationship between their means. (True/False)

ANSWERS TO STATE TRUE OR FALSE

1. **True**	2. **False**	3. **True**	4. **True**	5. **True**	6. **True**	7. **True**	8. **True**	9. **False**	10. **True**
11. **False**	12. **False**	13. **True**	14. **True**	15. **True**	16. **True**	17. **True**	18. **True**	19. **True**	20. **False**
21. **True**	22. **True**	23. **False**	24. **True**	25. **True**	26. **True**	27. **True**	28. **True**	29. **True**	30. **True**
31. **True**	32. **True**	33. **False**	34. **True**	35. **True**	36. **False**	37. **True**	38. **True**	39. **True**	40. **True**
41. **True**	42. **False**	43. **True**	44. **True**	45. **True**					