



## CHAPTER 24

# Data Analysis. Probability Theory

We first show how to handle data numerically or in terms of graphs, and how to extract information (average size, spread of data, etc.) from them. If these data are influenced by “chance,” by factors whose effect we cannot predict exactly (e.g., weather data, stock prices, life spans of tires, etc.), we have to rely on **probability theory**. This theory originated in games of chance, such as flipping coins, rolling dice, or playing cards. Nowadays it gives mathematical models of chance processes called *random experiments* or, briefly, **experiments**. In such an experiment we observe a **random variable**  $X$ , that is, a function whose values in a **trial** (a performance of an experiment) occur “by chance” (Sec. 24.3) according to a **probability distribution** that gives the individual probabilities with which possible values of  $X$  may occur in the long run. (Example: Each of the six faces of a die should occur with the same probability,  $1/6$ .) Or we may simultaneously observe more than one random variable, for instance, height *and* weight of persons or hardness *and* tensile strength of steel. This is discussed in Sec. 24.9, which will also give the basis for the mathematical justification of the statistical methods in Chapter 25.

*Prerequisite:* Calculus.

*References and Answers to Problems:* App. 1 Part G, App. 2.

## 24.1 Data Representation. Average. Spread

Data can be represented numerically or graphically in various ways. For instance, your daily newspaper may contain tables of stock prices and money exchange rates, curves or bar charts illustrating economical or political developments, or pie charts showing how your tax dollar is spent. And there are numerous other representations of data for special purposes.

In this section we discuss the use of standard representations of data in statistics. (For these, software packages, such as DATA DESK, R, and MINITAB, are available, and Maple or Mathematica may also be helpful; see pp. 789 and 1009) We explain corresponding concepts and methods in terms of typical examples.

### EXAMPLE 1 Recording and Sorting

Sample values (observations, measurements) should be **recorded** in the order in which they occur. **Sorting**, that is, ordering the sample values by size, is done as a first step of investigating properties of the sample and graphing it. Sorting is a standard process on the computer; see Ref. [E35], listed in App. 1.

*Super alloys* is a collective name for alloys used in jet engines and rocket motors, requiring high temperature (typically 1800°F), high strength, and excellent resistance to oxidation. Thirty specimens of Hastelloy C (nickel-based steel, investment cast) had the tensile strength (in 1000 lb/sq in.), recorded in the order obtained and rounded to integer values,

(1)                      89 77 88 91 88 93 99 79 87 84 86 82 88 89 78  
                             90 91 81 90 83 83 92 87 89 86 89 81 87 84 89

Sorting gives

(2)                      77 78 79 81 81 82 83 83 84 84 86 86 87 87 87  
                             88 88 88 89 89 89 89 89 90 90 91 91 92 93 99

## Graphic Representation of Data

We shall now discuss standard graphic representations used in statistics for obtaining information on properties of data.

### EXAMPLE 2 Stem-and-Leaf Plot (Fig. 507)

This is one of the simplest but most useful representations of data. For (1) it is shown in Fig. 507. The numbers in (1) range from 78 to 99; see (2). We divide these numbers into 5 groups, 75–79, 80–84, 85–89, 90–94, 95–99. The integers in the tens position of the groups are 7, 8, 8, 9, 9. These form the *stem* in Fig. 507. The first leaf is 789, representing 77, 78, 79. The second leaf is 1123344, representing 81, 81, 82, 83, 83, 84, 84. And so on.

The number of times a value occurs is called its **absolute frequency**. Thus 78 has absolute frequency 1, the value 89 has absolute frequency 5, etc. The column to the extreme left in Fig. 507 shows the **cumulative absolute frequencies**, that is, the sum of the absolute frequencies of the values up to the line of the leaf. Thus, the number 10 in the second line on the left shows that (1) has 10 values up to and including 84. The number 23 in the next line shows that there are 23 values not exceeding 89, etc. Dividing the cumulative absolute frequencies by  $n$  ( $= 30$  in Fig. 507) gives the **cumulative relative frequencies** 0.1, 0.33, 0.76, 0.93, 1.00.

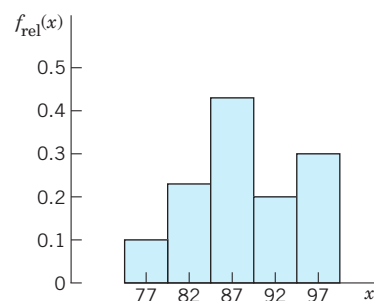
### EXAMPLE 3 Histogram (Fig. 508)

For large sets of data, histograms are better in displaying the distribution of data than stem-and-leaf plots. The principle is explained in Fig. 508. (An application to a larger data set is shown in Sec. 25.7). The bases of the rectangles in Fig. 508 are the  $x$ -intervals (known as **class intervals**) 74.5–79.5, 79.5–84.5, 84.5–89.5, 89.5–94.5, 94.5–99.5, whose midpoints (known as **class marks**) are  $x = 77, 82, 87, 92, 97$ , respectively. The height of a rectangle with class mark  $x$  is the **relative class frequency**  $f_{\text{rel}}(x)$ , defined as the number of data values in that class interval, divided by  $n$  ( $= 30$  in our case). Hence the areas of the rectangles are proportional to these relative frequencies, 0.10, 0.23, 0.43, 0.17, 0.07, so that histograms give a good impression of the distribution of data.

Leaf unit = 1.0

3	7	789
10	8	1123344
23	8	6677788899999
29	9	001123
30	9	9

**Fig. 507.** Stem-and-leaf plot of the data in Example 1



**Fig. 508.** Histogram of the data in Example 1 (grouped as in Fig. 507)

**EXAMPLE 4** Boxplot. Median. Interquartile Range. Outlier

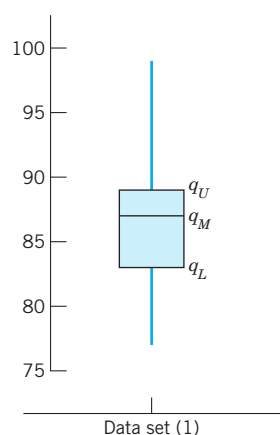
A **boxplot** of a set of data illustrates the average size and the spread of the values, in many cases the two most important quantities characterizing the set, as follows.

The average size is measured by the **median**, or *middle quartile*,  $q_M$ . If the number  $n$  of values of the set is *odd*, then  $q_M$  is the middlemost of the values when ordered as in (2). If  $n$  is *even*, then  $q_M$  is the average of the two middlemost values of the ordered set. In (2) we have  $n = 30$  and thus  $q_M = \frac{1}{2}(x_{15} + x_{16}) = \frac{1}{2}(87 + 88) = 87.5$ . (In general,  $q_M$  will be a fraction if  $n$  is even.)

The spread of values can be measured by the **range**  $R = x_{\max} - x_{\min}$ , the largest value minus the smallest one.

Better information on the spread gives the **interquartile range**  $IQR = q_U - q_L$ . Here  $q_U$  is the middlemost value (or the average of the two middlemost values) in the data *above* the median; and  $q_L$  is the middlemost value (or the average of the two middlemost values) in the data *below* the median. Hence in (2) we have  $q_U = x_{23} = 89$ ,  $q_L = x_8 = 83$ , and  $IQR = 89 - 83 = 6$ .

The box in Fig. 509 extends vertically from  $q_L$  to  $q_U$ ; it has height  $IQR = 6$ . The vertical lines below and above the box extend from  $x_{\min} = 77$  to  $x_{\max} = 99$ , so that they show  $R = 22$ .



**Fig. 509.** Boxplot of the data set (1)

The line above the box is suspiciously long. This suggests the concept of an **outlier**, a value that is more than 1.5 times the IQR away from either end of the box; here 1.5 is purely conventional. An outlier indicates that something might have gone wrong in the data collection. In (2) we have  $89 + 1.5 \text{ IQR} = 98$ , and we regard 99 as an outlier. ■

## Mean. Standard Deviation. Variance. Empirical Rule

Medians and quartiles are easily obtained by ordering and counting, practically without calculation. But they do not give full information on data: you can change data values to some extent without changing the median. Similarly for the quartiles.

The average size of the data values can be measured in a more refined way by the **mean**

$$(3) \quad \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j = \frac{1}{n} (x_1 + x_2 + \cdots + x_n).$$

This is the arithmetic mean of the data values, obtained by taking their sum and dividing by the data *size*  $n$ . Thus in (1),

$$\bar{x} = \frac{1}{30} (89 + 77 + \cdots + 89) = \frac{260}{3} \approx 86.7.$$

Every data value contributes, and changing one of them will change the mean.

Similarly, the spread (variability) of the data values can be measured in a more refined way by the **standard deviation**  $s$  or by its square, the **variance**

$$(4) \quad s^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2 = \frac{1}{n-1} [(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2].$$

Thus, to obtain the variance of the data, take the difference  $x_j - \bar{x}$  of each data value from the mean, square it, take the sum of these  $n$  squares, and divide it by  $n - 1$  (not  $n$ , as we motivate in Sec. 25.2). To get the standard deviation  $s$ , take the square root of  $s^2$ .

For example, using  $\bar{x} = 260/3$ , we get for the data (1) the variance

$$s^2 = \frac{1}{29} [(89 - \frac{260}{3})^2 + (77 - \frac{260}{3})^2 + \cdots + (89 - \frac{260}{3})^2] = \frac{2006}{87} \approx 23.06$$

Hence the standard deviation is  $s = \sqrt{2006/87} \approx 4.802$ . Note that the standard deviation has the same dimension as the data values ( $\text{kg}/\text{mm}^2$ , see at the beginning), which is an advantage. On the other hand, the variance is preferable to the standard deviation in developing statistical methods, as we shall see in Chap. 25.

**CAUTION!** Your CAS (Maple, for instance) may use  $1/n$  instead of  $1/(n - 1)$  in (4), but the latter is better when  $n$  is small (see Sec. 25.2).

Mean and standard deviation, introduced to give center and spread, actually give much more information according to this rule.

**Empirical Rule.** For any mound-shaped, nearly symmetric distribution of data the intervals

$$\bar{x} \pm s, \quad \bar{x} \pm 2s, \quad \bar{x} \pm 3s \quad \text{contain about} \quad 68\%, \quad 95\%, \quad 99.7\%,$$

respectively, of the data points.

### EXAMPLE 5 Empirical Rule and Outliers. z-Score

For (1), with  $\bar{x} = 86.7$  and  $s = 4.8$ , the three intervals in the Rule are  $81.9 \leq x \leq 91.5$ ,  $77.1 \leq x \leq 96.3$ ,  $72.3 \leq x \leq 101.1$  and contain 73% (22 values remain, 5 are too small, and 5 too large), 93% (28 values, 1 too small, and 1 too large), and 100%, respectively.

If we reduce the sample by omitting the outlier 99, mean and standard deviation reduce to  $\bar{x}_{\text{red}} = 86.2$ ,  $s_{\text{red}} = 4.3$ , approximately, and the percentage values become 67% (5 and 5 values outside), 93% (1 and 1 outside), and 100%.

Finally, the relative position of a value  $x$  in a set of mean  $\bar{x}$  and standard deviation  $s$  can be measured by the **z-score**

$$z(s) = \frac{x - \bar{x}}{s}.$$

This is the distance of  $x$  from the mean  $\bar{x}$  measured in multiples of  $s$ . For instance,  $z(83) = (83 - 86.7)/4.8 = -0.77$ . This is negative because 83 lies below the mean. By the Empirical Rule, the extreme  $z$ -values are about  $-3$  and  $3$ . ■

## PROBLEM SET 24.1

### 1–10 DATA REPRESENTATIONS

Represent the data by a stem-and-leaf plot, a histogram, and a boxplot:

1. Length of nails [mm]

19 21 19 20 19 20 21 20

2. Phone calls per minute in an office between 9:00 A.M. and 9:10 A.M.

6 6 4 2 1 7 0 4 6 7

3. Systolic blood pressure of 15 female patients of ages 20–22

156 158 154 133 141 130 144 137  
151 146 156 138 138 149 139

4. Iron content [%] of 15 specimens of hermatite ( $\text{Fe}_2\text{O}_3$ )

72.8 70.4 71.2 69.2 70.3 68.9 71.1 69.8  
71.5 69.7 70.5 71.3 69.1 70.9 70.6

5. Weight of filled bags [g] in an automatic filling

203 199 198 201 200 201 201

6. Gasoline consumption [miles per gallon, rounded] of six cars of the same model under similar conditions

15.0 15.5 14.5 15.0 15.5 15.0

7. Release time [sec] of a relay

1.3 1.2 1.4 1.5 1.3 1.3 1.4 1.1 1.5 1.4  
1.6 1.3 1.5 1.1 1.4 1.2 1.3 1.5 1.4 1.4

8. Foundrax test of Brinell hardness (2.5 mm steel ball, 62.5 kg load, 30 sec) of 20 copper plates (values in  $\text{kg}/\text{mm}^2$ )

86 86 87 89 76 85 82 86 87 85  
90 88 89 90 88 80 84 89 90 89

9. Efficiency [%] of seven Voith Francis turbines of runner diameter 2.3 m under a head range of 185 m

91.8 89.1 89.9 92.5 90.7 91.2 91.0

10.  $-0.51$   $0.12$   $-0.47$   $0.95$   $0.25$   $-0.18$   $-0.54$

### 11–16 AVERAGE AND SPREAD

Find the mean and compare it with the median. Find the standard deviation and compare it with the interquartile range.

11. For the data in Prob. 1

12. For the phone call data in Prob. 2

13. For the medical data in Prob. 3

14. For the iron contents in Prob. 4

15. For the release times in Prob. 7

16. For the Brinell hardness data in Prob. 8

17. **Outlier, reduced data.** Calculate  $s$  for the data 4 1 3 10 2. Then reduce the data by deleting the outlier and calculate  $s$ . Comment.

18. **Outlier, reduction.** Do the same tasks as in Prob. 17 for the hardness data in Prob. 8.

19. Construct the simplest possible data with  $\bar{x} = 100$  but  $q_M = 0$ . What is the point of this problem?

20. **Mean.** Prove that  $\bar{x}$  must always lie between the smallest and the largest data values.

## 24.2 Experiments, Outcomes, Events

We now turn to **probability theory**. This theory has the purpose of providing mathematical models of situations affected or even governed by “chance effects,” for instance, in weather forecasting, life insurance, quality of technical products (computers, batteries, steel sheets, etc.), traffic problems, and, of course, games of chance with cards or dice. And the accuracy of these models can be tested by suitable observations or experiments—this is a main purpose of **statistics** to be explained in Chap. 25.

We begin by defining some standard terms. An **experiment** is a process of measurement or observation, in a laboratory, in a factory, on the street, in nature, or wherever; so “experiment” is used in a rather general sense. Our interest is in experiments that involve **randomness**, chance effects, so that we cannot predict a result exactly. A **trial** is a single performance of an experiment. Its result is called an **outcome** or a **sample point**.  $n$  trials then give a **sample** of **size**  $n$  consisting of  $n$  sample points. The **sample space**  $S$  of an experiment is the set of all possible outcomes.

**EXAMPLES 1–6****Random Experiments. Sample Spaces**

- (1) Inspecting a lightbulb.  $S = \{\text{Defective, Nondefective}\}$ .
- (2) Rolling a die.  $S = \{1, 2, 3, 4, 5, 6\}$ .
- (3) Measuring tensile strength of wire.  $S$  the numbers in some interval.
- (4) Measuring copper content of brass.  $S$ : 50% to 90%, say.
- (5) Counting daily traffic accidents in New York.  $S$  the integers in some interval.
- (6) Asking for opinion about a new car model.  $S = \{\text{Like, Dislike, Undecided}\}$ .

The subsets of  $S$  are called **events** and the outcomes **simple events**.

**EXAMPLE 7****Events**

In (2), events are  $A = \{1, 3, 5\}$  (“*Odd number*”),  $B = \{2, 4, 6\}$  (“*Even number*”),  $C = \{5, 6\}$ . etc. Simple events are  $\{1\}, \{2\}, \dots, \{6\}$ .

If, in a trial, an outcome  $a$  happens and  $a \in A$  ( $a$  is an *element of*  $A$ ), we say that  $A$  happens. For instance, if a die turns up a 3, the event  $A$ : *Odd number* happens. Similarly, if  $C$  in Example 7 happens (meaning 5 or 6 turns up), then, say,  $D = \{4, 5, 6\}$  happens. Also note that  $S$  happens in each trial, meaning that *some* event of  $S$  always happens. All this is quite natural.

## Unions, Intersections, Complements of Events

In connection with basic probability laws we shall need the following concepts and facts about events (subsets)  $A, B, C, \dots$  of a given sample space  $S$ .

The **union**  $A \cup B$  of  $A$  and  $B$  consists of all points in  $A$  or  $B$  or both.

The **intersection**  $A \cap B$  of  $A$  and  $B$  consists of all points that are in both  $A$  and  $B$ .

If  $A$  and  $B$  have no points in common, we write

$$A \cap B = \emptyset$$

where  $\emptyset$  is the *empty set* (set with no elements) and we call  $A$  and  $B$  **mutually exclusive** (or **disjoint**) because, in a trial, the occurrence of  $A$  *excludes* that of  $B$  (and conversely)—if your die turns up an odd number, it cannot turn up an even number in the same trial. Similarly, a coin cannot turn up *Head* and *Tail* at the same time.

**Complement**  $A^c$  of  $A$ . This is the set of all the points of  $S$  *not* in  $A$ . Thus,

$$A \cap A^c = \emptyset, \quad A \cup A^c = S.$$

In Example 7 we have  $A^c = B$ , hence  $A \cup A^c = \{1, 2, 3, 4, 5, 6\} = S$ .

Another notation for the complement of  $A$  is  $\bar{A}$  (instead of  $A^c$ ), but we shall not use this because in set theory  $\bar{A}$  is used to denote the *closure* of  $A$  (not needed in our work).

**Unions and intersections** of more events are defined similarly. The **union**

$$\bigcup_{j=1}^m A_j = A_1 \cup A_2 \cup \dots \cup A_m$$

of events  $A_1, \dots, A_m$  consists of all points that are in at least one  $A_j$ . Similarly for the union  $A_1 \cup A_2 \cup \dots$  of infinitely many subsets  $A_1, A_2, \dots$  of an *infinite* sample space  $S$  (that is,  $S$  consists of infinitely many points). The **intersection**

$$\bigcap_{j=1}^m A_j = A_1 \cap A_2 \cap \dots \cap A_m$$

of  $A_1, \dots, A_m$  consists of the points of  $S$  that are in each of these events. Similarly for the intersection  $A_1 \cap A_2 \cap \dots$  of infinitely many subsets of  $S$ .

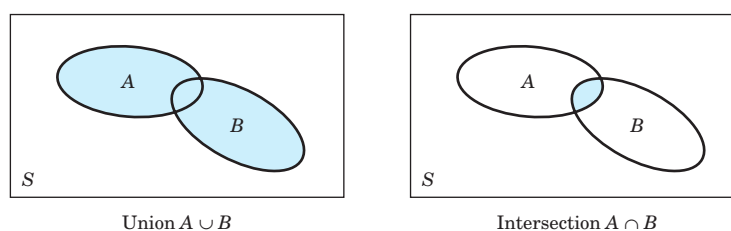
Working with events can be illustrated and facilitated by **Venn diagrams**<sup>1</sup> for showing unions, intersections, and complements, as in Figs. 510 and 511, which are typical examples that give the idea.

### EXAMPLE 8 Unions and Intersections of 3 Events

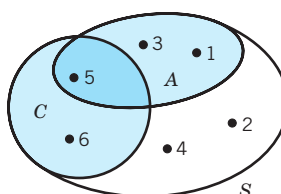
In rolling a die, consider the events

$A$ : Number greater than 3,     $B$ : Number less than 6,     $C$ : Even number.

Then  $A \cap B = \{4, 5\}$ ,  $B \cap C = \{2, 4\}$ ,  $C \cap A = \{4, 6\}$ ,  $A \cap B \cap C = \{4\}$ . Can you sketch a Venn diagram of this? Furthermore,  $A \cup B = S$ , hence  $A \cup B \cup C = S$  (why?).



**Fig. 510.** Venn diagrams showing two events  $A$  and  $B$  in a sample space  $S$  and their union  $A \cup B$  (colored) and intersection  $A \cap B$  (colored)



**Fig. 511.** Venn diagram for the experiment of rolling a die, showing  $S$ ,  $A = \{1, 3, 5\}$ ,  $C = \{5, 6\}$ ,  $A \cup C = \{1, 3, 5, 6\}$ ,  $A \cap C = \{5\}$

## PROBLEM SET 24.2

### 1–12 SAMPLE SPACES, EVENTS

Graph a sample space for the experiments:

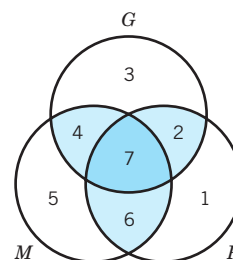
1. Drawing 3 screws from a lot of right-handed and left-handed screws
2. Tossing 2 coins
3. Rolling 2 dice
4. Rolling a die until the first *Six* appears
5. Tossing a coin until the first *Head* appears
6. Recording the lifetime of each of 3 lightbulbs

<sup>1</sup>JOHN VENN (1834–1923), English mathematician.

7. Recording the daily maximum temperature  $X$  and the daily maximum air pressure  $Y$  at Times Square in New York
8. Choosing a committee of 2 from a group of 5 people
9. Drawing gaskets from a lot of 10, containing one defective  $D$ , until  $D$  is drawn, one at a time and assuming **sampling without replacement**, that is, gaskets drawn are *not* returned to the lot. (More about this in Sec. 24.6)
10. In rolling 3 dice, are the events  $A$ : *Sum divisible by 3* and  $B$ : *Sum divisible by 5* mutually exclusive?
11. Answer the questions in Prob. 10 for rolling 2 dice.
12. List all 8 subsets of the sample space  $S = \{a, b, c\}$ .
13. In Prob. 3 circle and mark the events  $A$ : *Faces are equal*,  $B$ : *Sum of faces less than 5*,  $A \cup B$ ,  $A \cap B$ ,  $A^c$ ,  $B^c$ .
14. In drawing 2 screws from a lot of right-handed and left-handed screws, let  $A$ ,  $B$ ,  $C$ ,  $D$  mean at a least 1 right-handed, at least 1 left-handed, 2 right-handed, 2 left-handed, respectively. Are  $A$  and  $B$  mutually exclusive?  $C$  and  $D$ ?

#### 15–20 VENN DIAGRAMS

15. In connection with a trip to Europe by some students, consider the events  $P$  that they see Paris,  $G$  that they have a good time, and  $M$  that they run out of money, and describe in words the events 1,  $\dots$ , 7 in the diagram.



Problem 15

16. Show that, by the definition of complement, for any subset  $A$  of a sample space  $S$ .

$$(A^c)^c = A, \quad S^c = \emptyset, \quad \emptyset^c = S, \\ A \cup A^c = S, \quad A \cap A^c = \emptyset.$$

17. Using a Venn diagram, show that  $A \subseteq B$  if and only if  $A \cup B = B$ .
18. Using a Venn diagram, show that  $A \subseteq B$  if and only if  $A \cap B = A$ .
19. (**De Morgan's laws**) Using Venn diagrams, graph and check *De Morgan's laws*

$$(A \cup B)^c = A^c \cap B^c \\ (A \cap B)^c = A^c \cup B^c.$$

20. Using Venn diagrams, graph and check the rules

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \\ A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

## 24.3 Probability

The “probability” of an event  $A$  in an experiment is supposed to measure how frequently  $A$  is *about* to occur if we make many trials. If we flip a coin, then heads  $H$  and tails  $T$  will appear *about* equally often—we say that  $H$  and  $T$  are “**equally likely**.” Similarly, for a regularly shaped die of homogeneous material (“**fair die**”) each of the six outcomes  $1, \dots, 6$  will be equally likely. These are examples of experiments in which the sample space  $S$  consists of finitely many outcomes (points) that for reasons of some symmetry can be regarded as equally likely. This suggests the following definition.

#### DEFINITION

##### First Definition of Probability

If the sample space  $S$  of an experiment consists of finitely many outcomes (points) that are equally likely, then the probability  $P(A)$  of an event  $A$  is

$$(1) \quad P(A) = \frac{\text{Number of points in } A}{\text{Number of points in } S}.$$



From this definition it follows immediately that, in particular,

$$(2) \quad P(S) = 1.$$

### EXAMPLE 1 Fair Die

In rolling a fair die once, what is the probability  $P(A)$  of  $A$  of obtaining a 5 or a 6? The probability of  $B$ : “Even number”?

**Solution.** The six outcomes are equally likely, so that each has probability  $1/6$ . Thus  $P(A) = 2/6 = 1/3$  because  $A = \{5, 6\}$  has 2 points, and  $P(B) = 3/6 = 1/2$ . ■

Definition 1 takes care of many games as well as some practical applications, as we shall see, but certainly not of all experiments, simply because in many problems we do not have finitely many equally likely outcomes. To arrive at a more general definition of probability, we regard **probability as the counterpart of relative frequency**. Recall from Sec. 24.1 that the **absolute frequency**  $f(A)$  of an event  $A$  in  $n$  trials is the number of times  $A$  occurs, and the **relative frequency** of  $A$  in these trials is  $f(A)/n$ ; thus

$$(3) \quad f_{\text{rel}}(A) = \frac{f(A)}{n} = \frac{\text{Number of times } A \text{ occurs}}{\text{Number of trials}}.$$

Now if  $A$  did not occur, then  $f(A) = 0$ . If  $A$  always occurred, then  $f(A) = n$ . These are the extreme cases. Division by  $n$  gives

$$(4^*) \quad 0 \leq f_{\text{rel}}(A) \leq 1.$$

In particular, for  $A = S$  we have  $f(S) = n$  because  $S$  always occurs (meaning that some event always occurs; if necessary, see Sec. 24.2, after Example 7). Division by  $n$  gives

$$(5^*) \quad f_{\text{rel}}(S) = 1.$$

Finally, if  $A$  and  $B$  are mutually exclusive, they cannot occur together. Hence the absolute frequency of their union  $A \cup B$  must equal the sum of the absolute frequencies of  $A$  and  $B$ . Division by  $n$  gives the same relation for the relative frequencies,

$$(6^*) \quad f_{\text{rel}}(A \cup B) = f_{\text{rel}}(A) + f_{\text{rel}}(B) \quad (A \cap B = \emptyset).$$

We are now ready to extend the definition of probability to experiments in which equally likely outcomes are not available. Of course, the extended definition should include Definition 1. Since probabilities are supposed to be the theoretical counterpart of relative frequencies, we choose the properties in (4\*), (5\*), (6\*) as axioms. (Historically, such a choice is the result of a long process of gaining experience on what might be best and most practical.)

**DEFINITION 2****General Definition of Probability**

Given a sample space  $S$ , with each event  $A$  of  $S$  (subset of  $S$ ) there is associated a number  $P(A)$ , called the **probability** of  $A$ , such that the following **axioms of probability** are satisfied.

1. For every  $A$  in  $S$ ,

$$(4) \quad 0 \leq P(A) \leq 1.$$

2. The entire sample space  $S$  has the probability

$$(5) \quad P(S) = 1.$$

3. For mutually exclusive events  $A$  and  $B$  ( $A \cap B = \emptyset$ ; see Sec. 24.2),

$$(6) \quad P(A \cup B) = P(A) + P(B) \quad (A \cap B = \emptyset).$$

If  $S$  is infinite (has infinitely many points), Axiom 3 has to be replaced by **3'**. For mutually exclusive events  $A_1, A_2, \dots$ ,

$$(6') \quad P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$

In the infinite case the subsets of  $S$  on which  $P(A)$  is defined are restricted to form a so-called  $\sigma$ -algebra, as explained in Ref. [GenRef6] (not [G6]!) in App. 1. This is of no practical consequence to us.

## Basic Theorems of Probability

We shall see that the axioms of probability will enable us to build up probability theory and its application to statistics. We begin with three basic theorems. The first of them is useful if we can get the probability of the complement  $A^c$  more easily than  $P(A)$  itself.

**THEOREM 1****Complementation Rule**

For an event  $A$  and its complement  $A^c$  in a sample space  $S$ ,

$$(7) \quad P(A^c) = 1 - P(A).$$

**PROOF** By the definition of complement (Sec. 24.2), we have  $S = A \cup A^c$  and  $A \cap A^c = \emptyset$ . Hence by Axioms 2 and 3,

$$1 = P(S) = P(A) + P(A^c), \quad \text{thus} \quad P(A^c) = 1 - P(A). \quad \blacksquare$$

**EXAMPLE 2** Coin Tossing

Five coins are tossed simultaneously. Find the probability of the event  $A$ : *At least one head turns up*. Assume that the coins are fair.

**Solution.** Since each coin can turn up heads or tails, the sample space consists of  $2^5 = 32$  outcomes. Since the coins are fair, we may assign the same probability ( $1/32$ ) to each outcome. Then the event  $A^c$  (*No heads turn up*) consists of only 1 outcome. Hence  $P(A^c) = 1/32$ , and the answer is  $P(A) = 1 - P(A^c) = 31/32$ . ■

The next theorem is a simple extension of Axiom 3, which you can readily prove by induction.

**THEOREM 2****Addition Rule for Mutually Exclusive Events**

For mutually exclusive events  $A_1, \dots, A_m$  in a sample space  $S$ ,

$$(8) \quad P(A_1 \cup A_2 \cup \dots \cup A_m) = P(A_1) + P(A_2) + \dots + P(A_m).$$

**EXAMPLE 3** Mutually Exclusive Events

If the probability that on any workday a garage will get 10–20, 21–30, 31–40, over 40 cars to service is 0.20, 0.35, 0.25, 0.12, respectively, what is the probability that on a given workday the garage gets at least 21 cars to service?

**Solution.** Since these are mutually exclusive events, Theorem 2 gives the answer  $0.35 + 0.25 + 0.12 = 0.72$ . Check this by the complementation rule. ■

In many cases, events will not be mutually exclusive. Then we have

**THEOREM 3****Addition Rule for Arbitrary Events**

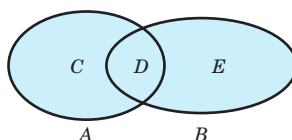
For events  $A$  and  $B$  in a sample space,

$$(9) \quad P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

**PROOF**  $C, D, E$  in Fig. 512 make up  $A \cup B$  and are mutually exclusive (disjoint). Hence by Theorem 2,

$$P(A \cup B) = P(C) + P(D) + P(E).$$

This gives (9) because on the right  $P(C) + P(D) = P(A)$  by Axiom 3 and disjointness; and  $P(E) = P(B) - P(D) = P(B) - P(A \cap B)$ , also by Axiom 3 and disjointness. ■



**Fig. 512.** Proof of Theorem 3

Note that for mutually exclusive events  $A$  and  $B$  we have  $A \cap B = \emptyset$  by definition and, by comparing (9) and (6),

$$(10) \quad P(\emptyset) = 0.$$

(Can you also prove this by (5) and (7)?)

#### EXAMPLE 4 Union of Arbitrary Events

In tossing a fair die, what is the probability of getting an odd number or a number less than 4?

**Solution.** Let  $A$  be the event “Odd number” and  $B$  the event “Number less than 4.” Then Theorem 3 gives the answer

$$P(A \cup B) = \frac{3}{6} + \frac{3}{6} - \frac{2}{6} = \frac{2}{3}$$

because  $A \cap B = \text{“Odd number less than 4”} = \{1, 3\}$ . ■

## Conditional Probability. Independent Events

Often it is required to find the probability of an event  $B$  under the condition that an event  $A$  occurs. This probability is called the **conditional probability of  $B$  given  $A$**  and is denoted by  $P(B|A)$ . In this case  $A$  serves as a new (reduced) sample space, and that probability is the fraction of  $P(A)$  which corresponds to  $A \cap B$ . Thus

$$(11) \quad P(B|A) = \frac{P(A \cap B)}{P(A)} \quad [P(A) \neq 0].$$

Similarly, the *conditional probability of  $A$  given  $B$*  is

$$(12) \quad P(A|B) = \frac{P(A \cap B)}{P(B)} \quad [P(B) \neq 0].$$

Solving (11) and (12) for  $P(A \cap B)$ , we obtain

#### THEOREM 4

##### Multiplication Rule

If  $A$  and  $B$  are events in a sample space  $S$  and  $P(A) \neq 0$ ,  $P(B) \neq 0$ , then

$$(13) \quad P(A \cap B) = P(A)P(B|A) = P(B)P(A|B).$$

#### EXAMPLE 5 Multiplication Rule

In producing screws, let  $A$  mean “screw too slim” and  $B$  “screw too short.” Let  $P(A) = 0.1$  and let the conditional probability that a slim screw is also too short be  $P(B|A) = 0.2$ . What is the probability that a screw that we pick randomly from the lot produced will be both too slim and too short?

**Solution.**  $P(A \cap B) = P(A)P(B|A) = 0.1 \cdot 0.2 = 0.02 = 2\%$ , by Theorem 4. ■

**Independent Events.** If events  $A$  and  $B$  are such that

$$(14) \quad P(A \cap B) = P(A)P(B),$$

they are called **independent events**. Assuming  $P(A) \neq 0$ ,  $P(B) \neq 0$ , we see from (11)–(13) that in this case

$$P(A|B) = P(A), \quad P(B|A) = P(B).$$

This means that the probability of  $A$  does not depend on the occurrence or nonoccurrence of  $B$ , and conversely. This justifies the term “independent.”

**Independence of  $m$  Events.** Similarly,  $m$  events  $A_1, \dots, A_m$  are called **independent** if

$$(15a) \quad P(A_1 \cap \dots \cap A_m) = P(A_1) \cdots P(A_m)$$

as well as for every  $k$  different events  $A_{j_1}, A_{j_2}, \dots, A_{j_k}$ .

$$(15b) \quad P(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}) = P(A_{j_1})P(A_{j_2}) \cdots P(A_{j_k})$$

where  $k = 2, 3, \dots, m - 1$ .

Accordingly, three events  $A, B, C$  are independent if and only if

$$(16) \quad \begin{aligned} P(A \cap B) &= P(A)P(B), \\ P(B \cap C) &= P(B)P(C), \\ P(C \cap A) &= P(C)P(A), \\ P(A \cap B \cap C) &= P(A)P(B)P(C). \end{aligned}$$

**Sampling.** Our next example has to do with randomly drawing objects, *one at a time*, from a given set of objects. This is called **sampling from a population**, and there are two ways of sampling, as follows.

1. In **sampling with replacement**, the object that was drawn at random is placed back to the given set and the set is mixed thoroughly. Then we draw the next object at random.
2. In **sampling without replacement** the object that was drawn is put aside.

#### EXAMPLE 6 Sampling With and Without Replacement

A box contains 10 screws, three of which are defective. Two screws are drawn at random. Find the probability that neither of the two screws is defective.

**Solution.** We consider the events

$A$ : First drawn screw nondefective.

$B$ : Second drawn screw nondefective.

Clearly,  $P(A) = \frac{7}{10}$  because 7 of the 10 screws are nondefective and we sample at random, so that each screw has the same probability ( $\frac{1}{10}$ ) of being picked. If we sample with replacement, the situation before the second drawing is the same as at the beginning, and  $P(B) = \frac{7}{10}$ . The events are independent, and the answer is

$$P(A \cap B) = P(A)P(B) = 0.7 \cdot 0.7 = 0.49 = 49\%.$$

If we sample without replacement, then  $P(A) = \frac{7}{10}$ , as before. If  $A$  has occurred, then there are 9 screws left in the box, 3 of which are defective. Thus  $P(B|A) = \frac{6}{9} = \frac{2}{3}$ , and Theorem 4 yields the answer

$$P(A \cap B) = \frac{7}{10} \cdot \frac{2}{3} = 47\%.$$

Is it intuitively clear that this value must be smaller than the preceding one? ■

### PROBLEM SET 24.3

1. In rolling 3 fair dice, what is the probability of obtaining a sum not greater than 16?
2. In rolling 2 fair dice, what is the probability of a sum greater than 3 but not exceeding 6?
3. Three screws are drawn at random from a lot of 100 screws, 10 of which are defective. Find the probability of the event that all 3 screws drawn are nondefective, assuming that we draw (a) with replacement, (b) without replacement.
4. In Prob. 3 find the probability of  $E$ : *At least 1 defective* (i) directly, (ii) by using complements; in both cases (a) and (b).
5. If a box contains 10 left-handed and 20 right-handed screws, what is the probability of obtaining at least one right-handed screw in drawing 2 screws with replacement?
6. Will the probability in Prob. 5 increase or decrease if we draw without replacement. First guess, then calculate.
7. Under what conditions will it make *practically* no difference whether we sample with or without replacement?
8. If a certain kind of tire has a life exceeding 40,000 miles with probability 0.90, what is the probability that a set of these tires on a car will last longer than 40,000 miles?
9. If we inspect photocopy paper by randomly drawing 5 sheets without replacement from every pack of 500, what is the probability of getting 5 clean sheets although 0.4% of the sheets contain spots?
10. Suppose that we draw cards repeatedly and with replacement from a file of 100 cards, 50 of which refer to male and 50 to female persons. What is the probability of obtaining the second “female” card before the third “male” card?
11. A batch of 200 iron rods consists of 50 oversized rods, 50 undersized rods, and 100 rods of the desired length. If two rods are drawn at random without replacement, what is the probability of obtaining (a) two rods of the desired length, (b) exactly one of the desired length, (c) none of the desired length?
12. If a circuit contains four automatic switches and we want that, with a probability of 99%, during a given time interval the switches to be all working, what probability of failure per time interval can we admit for a single switch?
13. A pressure control apparatus contains 3 electronic tubes. The apparatus will not work unless all tubes are operative. If the probability of failure of each tube during some interval of time is 0.04, what is the corresponding probability of failure of the apparatus?
14. Suppose that in a production of spark plugs the fraction of defective plugs has been constant at 2% over a long time and that this process is controlled every half hour by drawing and inspecting two just produced. Find the probabilities of getting (a) no defectives, (b) 1 defective, (c) 2 defectives. What is the sum of these probabilities?
15. What gives the greater probability of hitting at least once: (a) hitting with probability  $1/2$  and firing 1 shot, (b) hitting with probability  $1/4$  and firing 2 shots, (c) hitting with probability  $1/8$  and firing 4 shots? First guess.
16. You may wonder whether in (16) the last relation follows from the others, but the answer is no. To see this, imagine that a chip is drawn from a box containing 4 chips numbered 000, 011, 101, 110, and let  $A$ ,  $B$ ,  $C$  be the events that the first, second, and third digit, respectively, on the drawn chip is 1. Show that then the first three formulas in (16) hold but the last one does not hold.
17. Show that if  $B$  is a subset of  $A$ , then  $P(B) \leq P(A)$ .
18. Extending Theorem 4, show that  $P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$ .
19. Make up an example similar to Prob. 16, for instance, in terms of divisibility of numbers.

## 24.4 Permutations and Combinations

Permutations and combinations help in finding probabilities  $P(A) = a/k$  by **systematically counting** the number  $a$  of points of which an event  $A$  consists; here,  $k$  is the number of points of the sample space  $S$ . The practical difficulty is that  $a$  may often be surprisingly large, so that actual counting becomes hopeless. For example, if in assembling some instrument you need 10 different screws in a certain order and you want to draw them

randomly from a box (which contains nothing else) the probability of obtaining them in the required order is only  $1/3,628,800$  because there are

$$10! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 = 3,628,800$$

orders in which they can be drawn. Similarly, in many other situations the numbers of orders, arrangements, etc. are often incredibly large. (If you are unimpressed, take 20 screws—how much bigger will the number be?)

## Permutations

A **permutation** of given things (*elements* or *objects*) is an arrangement of these things in a row in some order. For example, for three letters  $a, b, c$  there are  $3! = 1 \cdot 2 \cdot 3 = 6$  permutations:  $abc, acb, bac, bca, cab, cba$ . This illustrates (a) in the following theorem.

### THEOREM 1

#### Permutations

(a) **Different things.** The number of permutations of  $n$  different things taken all at a time is

$$(1) \quad n! = 1 \cdot 2 \cdot 3 \cdots n \quad (\text{read “}n \text{ factorial”}).$$

(b) **Classes of equal things.** If  $n$  given things can be divided into  $c$  classes of alike things differing from class to class, then the number of permutations of these things taken all at a time is

$$(2) \quad \frac{n!}{n_1! n_2! \cdots n_c!} \quad (n_1 + n_2 + \cdots + n_c = n)$$

Where  $n_j$  is the number of things in the  $j$ th class.

**PROOF** (a) There are  $n$  choices for filling the first place in the row. Then  $n - 1$  things are still available for filling the second place, etc.

(b)  $n_1$  alike things in class 1 make  $n_1!$  permutations collapse into a single permutation (those in which class 1 things occupy the same  $n_1$  positions), etc., so that (2) follows from (1). ■

### EXAMPLE 1

#### Illustration of Theorem 1(b)

If a box contains 6 red and 4 blue balls, the probability of drawing first the red and then the blue balls is

$$P = 6!4!/10! = 1/210 \approx 0.5\%. \quad \blacksquare$$

A **permutation of  $n$  things taken  $k$  at a time** is a permutation containing only  $k$  of the  $n$  given things. Two such permutations consisting of the same  $k$  elements, in a different order, are different, by definition. For example, there are 6 different permutations of the three letters  $a, b, c$ , taken two letters at a time,  $ab, ac, bc, ba, ca, cb$ .

A **permutation of  $n$  things taken  $k$  at a time with repetitions** is an arrangement obtained by putting any given thing in the first position, any given thing, including a repetition of the one just used, in the second, and continuing until  $k$  positions are filled. For example, there

are  $3^2 = 9$  different such permutations of  $a, b, c$  taken 2 letters at a time, namely, the preceding 6 permutations and  $aa, bb, cc$ . You may prove (see Team Project 14):

**THEOREM 2****Permutations**

*The number of different permutations of  $n$  different things taken  $k$  at a time without repetitions is*

$$(3a) \quad n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

*and with repetitions is*

$$(3b) \quad n^k.$$

**EXAMPLE 2****Illustration of Theorem 2**

In an encrypted message the letters are arranged in groups of five letters, called *words*. From (3b) we see that the number of different such words is

$$26^5 = 11,881,376.$$

From (3a) it follows that the number of different such words containing each letter no more than once is

$$26!/(26-5)! = 26 \cdot 25 \cdot 24 \cdot 23 \cdot 22 = 7,893,600. \quad \blacksquare$$

## Combinations

In a permutation, the order of the selected things is essential. In contrast, a **combination** of given things means any selection of one or more things *without regard to order*. There are two kinds of combinations, as follows.

The number of **combinations of  $n$  different things, taken  $k$  at a time, without repetitions** is the number of sets that can be made up from the  $n$  given things, each set containing  $k$  different things and no two sets containing exactly the same  $k$  things.

The number of **combinations of  $n$  different things, taken  $k$  at a time, with repetitions** is the number of sets that can be made up of  $k$  things chosen from the given  $n$  things, each being used as often as desired.

For example, there are three combinations of the three letters  $a, b, c$ , taken two letters at a time, without repetitions, namely,  $ab, ac, bc$ , and six such combinations with repetitions, namely,  $ab, ac, bc, aa, bb, cc$ .

**THEOREM 3****Combinations**

*The number of different combinations of  $n$  different things taken,  $k$  at a time, without repetitions, is*

$$(4a) \quad \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{1 \cdot 2 \cdots k},$$

*and the number of those combinations with repetitions is*

$$(4b) \quad \binom{n+k-1}{k}.$$



**PROOF** The statement involving (4a) follows from the first part of Theorem 2 by noting that there are  $k!$  *permutations* of  $k$  things from the given  $n$  things that differ by the order of the elements (see Theorem 1), but there is only a single *combination* of those  $k$  things of the type characterized in the first statement of Theorem 3. The last statement of Theorem 3 can be proved by induction (see Team Project 14). ■

### EXAMPLE 3 Illustration of Theorem 3

The number of samples of five lightbulbs that can be selected from a lot of 500 bulbs is [see (4a)]

$$\binom{500}{5} = \frac{500!}{5!495!} = \frac{500 \cdot 499 \cdot 498 \cdot 497 \cdot 496}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 255,244,687,600. \quad \blacksquare$$

## Factorial Function

In (1)–(4) the **factorial function** is basic. By definition,

$$(5) \quad 0! = 1.$$

Values may be computed recursively from given values by

$$(6) \quad (n + 1)! = (n + 1)n!.$$

For large  $n$  the function is very large (see Table A3 in App. 5). A convenient approximation for large  $n$  is the **Stirling formula**<sup>2</sup>

$$(7) \quad n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \quad (e = 2.718\cdots)$$

where  $\sim$  is read “**asymptotically equal**” and means that the ratio of the two sides of (7) approaches 1 as  $n$  approaches infinity.

### EXAMPLE 4 Stirling Formula

$n!$	By (7)	Exact Value	Relative Error
4!	23.5	24	2.1%
10!	3,598,696	3,628,800	0.8%
20!	$2.42279 \cdot 10^{18}$	2,432,902,008,176,640,000	0.4%

## Binomial Coefficients

The **binomial coefficients** are defined by the formula

$$(8) \quad \binom{a}{k} = \frac{a(a-1)(a-2)\cdots(a-k+1)}{k!} \quad (k \geq 0, \text{ integer}).$$

<sup>2</sup>JAMES STIRLING (1692–1770), Scots mathematician.

The numerator has  $k$  factors. Furthermore, we define

$$(9) \quad \binom{a}{0} = 1, \quad \text{in particular,} \quad \binom{0}{0} = 1.$$

For integer  $a = n$  we obtain from (8)

$$(10) \quad \binom{n}{k} = \binom{n}{n-k} \quad (n \geq 0, 0 \leq k \leq n).$$

Binomial coefficients may be computed recursively, because

$$(11) \quad \binom{a}{k} + \binom{a}{k+1} = \binom{a+1}{k+1} \quad (k \geq 0, \text{integer}).$$

Formula (8) also yields

$$(12) \quad \binom{-m}{k} = (-1)^k \binom{m+k-1}{k} \quad (k \geq 0, \text{integer}) \\ (m > 0).$$

There are numerous further relations; we mention two important ones,

$$(13) \quad \sum_{s=0}^{n-1} \binom{k+s}{k} = \binom{n+k}{k+1} \quad (k \geq 0, n \geq 1, \\ \text{both integer})$$

and

$$(14) \quad \sum_{k=0}^r \binom{p}{k} \binom{q}{r-k} = \binom{p+q}{r} \quad (r \geq 0, \text{integer}).$$

## PROBLEM SET 24.4

Note the large numbers in the answers to some of these problems, which would make *counting cases hopeless!*

1. In how many ways can a company assign 10 drivers to  $n$  buses, one driver to each bus and conversely?
2. List (a) all permutations, (b) all combinations without repetitions, (c) all combinations with repetitions, of 5 letters  $a, e, i, o, u$  taken 2 at a time.
3. If a box contains 4 rubber gaskets and 2 plastic gaskets, what is the probability of drawing (a) first the plastic and then the rubber gaskets, (b) first the rubber and then the plastic ones? Do this by using a theorem and checking it by multiplying probabilities.
4. An urn contains 2 green, 3 yellow, and 5 red balls. We draw 1 ball at random and put it aside. Then we draw the next ball, and so on. Find the probability of drawing

at first the 2 green balls, then the 3 yellow ones, and finally the red ones.

5. In how many different ways can we select a committee consisting of 3 engineers, 2 physicists, and 2 computer scientists from 10 engineers, 5 physicists, and 6 computer scientists? First guess.
6. How many different samples of 4 objects can we draw from a lot of 50?
7. Of a lot of 10 items, 2 are defective. (a) Find the number of different samples of 4. Find the number of samples of 4 containing (b) no defectives, (c) 1 defective, (d) 2 defectives.
8. Determine the number of different bridge hands. (A bridge hand consists of 13 cards selected from a full deck of 52 cards.)

9. In how many different ways can 6 people be seated at a round table?
10. If a cage contains 100 mice, 3 of which are male, what is the probability that the 3 male mice will be included if 10 mice are randomly selected?
11. How many automobile registrations may the police have to check in a hit-and-run accident if a witness reports KDP7 and cannot remember the last two digits on the license plate but is certain that all three digits were different?
12. If 3 suspects who committed a burglary and 6 innocent persons are lined up, what is the probability that a witness who is not sure and has to pick three persons will pick the three suspects by chance? That the witness picks 3 innocent persons by chance?
13. **CAS PROJECT. Stirling formula.** (a) Using (7), compute approximate values of  $n!$  for  $n = 1, \dots, 20$ . (b) Determine the relative error in (a). Find an empirical formula for that relative error. (c) An upper bound for that relative error is  $e^{1/12n} - 1$ . Try to relate your empirical formula to this. (d) Search through the literature for further information on Stirling's formula. Write a short essay about your

findings, arranged in logical order and illustrated with numeric examples.

**14. TEAM PROJECT. Permutations, Combinations.**

- (a) Prove Theorem 2.
- (b) Prove the last statement of Theorem 3.
- (c) Derive (11) from (8).
- (d) By the **binomial theorem**,

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k},$$

so that  $a^k b^{n-k}$  has the coefficient  $\binom{n}{k}$ . Can you conclude this from Theorem 3 or is this a mere coincidence?

- (e) Prove (14) by using the binomial theorem.
- (f) Collect further formulas for binomial coefficients from the literature and illustrate them numerically.

15. **Birthday problem.** What is the probability that in a group of 20 people (that includes no twins) at least two have the same birthday, if we assume that the probability of having birthday on a given day is  $1/365$  for every day. First guess. *Hint.* Consider the complementary event.

## 24.5 Random Variables. Probability Distributions

In Sec. 24.1 we considered frequency distributions of data. These distributions show the absolute or relative frequency of the data values. Similarly, a **probability distribution** or, briefly, a **distribution**, shows the probabilities of events in an experiment. The quantity that we observe in an experiment will be denoted by  $X$  and called a **random variable** (or **stochastic variable**) because the value it will assume in the next trial depends on chance, on **randomness**—if you roll a die, you get one of the numbers from 1 to 6, but you don't know which one will show up next. Thus  $X = \text{Number a die turns up}$  is a random variable. So is  $X = \text{Elasticity of rubber}$  (elongation at break). ("Stochastic" means related to chance.)

If we **count** (cars on a road, defective screws in a production, tosses until a die shows the first Six), we have a **discrete random variable and distribution**. If we **measure** (electric voltage, rainfall, hardness of steel), we have a **continuous random variable and distribution**. Precise definitions follow. In both cases the distribution of  $X$  is determined by the **distribution function**

$$(1) \quad F(x) = P(X \leq x);$$

this is the probability that in a trial,  $X$  will assume any value not exceeding  $x$ .

**CAUTION!** The terminology is not uniform.  $F(x)$  is sometimes also called the **cumulative distribution function**.

For (1) to make sense in both the discrete and the continuous case we formulate conditions as follows.

### DEFINITION

#### Random Variable

A **random variable**  $X$  is a function defined on the sample space  $S$  of an experiment. Its values are real numbers. For every number  $a$  the probability

$$P(X = a)$$

with which  $X$  assumes  $a$  is defined. Similarly, for any interval  $I$  the probability

$$P(X \in I)$$

with which  $X$  assumes any value in  $I$  is defined.

Although this definition is very general, in practice only a very small number of distributions will occur over and over again in applications.

From (1) we obtain the fundamental formula for the probability corresponding to an interval  $a < x \leq b$ ,

$$(2) \quad P(a < X \leq b) = F(b) - F(a).$$

This follows because  $X \leq a$  (“ $X$  assumes any value not exceeding  $a$ ”) and  $a < X \leq b$  (“ $X$  assumes any value in the interval  $a < x \leq b$ ”) are mutually exclusive events, so that by (1) and Axiom 3 of Definition 2 in Sec. 24.3

$$\begin{aligned} F(b) &= P(X \leq b) = P(X \leq a) + P(a < X \leq b) \\ &= F(a) + P(a < X \leq b) \end{aligned}$$

and subtraction of  $F(a)$  on both sides gives (2).

## Discrete Random Variables and Distributions

By definition, a random variable  $X$  and its distribution are **discrete** if  $X$  assumes only finitely many or at most countably many values  $x_1, x_2, x_3, \dots$ , called the **possible values** of  $X$ , with positive probabilities  $p_1 = P(X = x_1), p_2 = P(X = x_2), p_3 = P(X = x_3), \dots$ , whereas the probability  $P(X \in I)$  is zero for any interval  $I$  containing no possible value.

Clearly, the discrete distribution of  $X$  is also determined by the **probability function**  $f(x)$  of  $X$ , defined by

$$(3) \quad f(x) = \begin{cases} p_j & \text{if } x = x_j \\ 0 & \text{otherwise} \end{cases} \quad (j = 1, 2, \dots),$$

From this we get the values of the **distribution function**  $F(x)$  by taking sums,

$$(4) \quad F(x) = \sum_{x_j \leq x} f(x_j) = \sum_{x_j \leq x} p_j$$

where for any given  $x$  we sum all the probabilities  $p_j$  for which  $x_j$  is smaller than or equal to that of  $x$ . This is a **step function** with upward jumps of size  $p_j$  at the possible values  $x_j$  of  $X$  and constant in between.

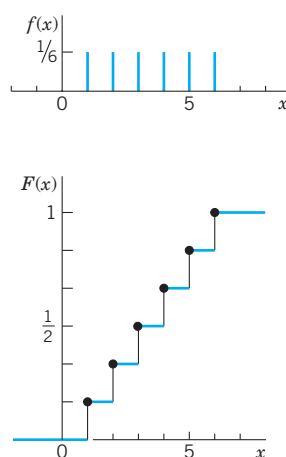
### EXAMPLE 1 Probability Function and Distribution Function

Figure 513 shows the probability function  $f(x)$  and the distribution function  $F(x)$  of the discrete random variable

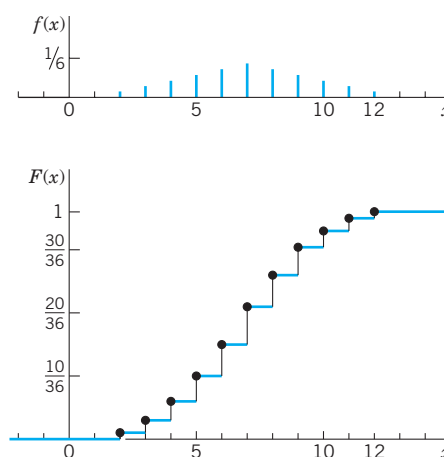
$X = \text{Number a fair die turns up.}$

$X$  has the possible values  $x = 1, 2, 3, 4, 5, 6$  with probability  $1/6$  each. At these  $x$  the distribution function has upward jumps of magnitude  $1/6$ . Hence from the graph of  $f(x)$  we can construct the graph of  $F(x)$  and conversely.

In Figure 513 (and the next one) at each jump the *fat dot* indicates the *function value at the jump!* ■



**Fig. 513.** Probability function  $f(x)$  and distribution function  $F(x)$  of the random variable  $X = \text{Number obtained in tossing a fair die once}$



**Fig. 514.** Probability function  $f(x)$  and distribution function  $F(x)$  of the random variable  $X = \text{Sum of the two numbers obtained in tossing two fair dice once}$

### EXAMPLE 2 Probability Function and Distribution Function

The random variable  $X = \text{Sum of the two numbers two fair dice turn up}$  is discrete and has the possible values  $2 (= 1 + 1), 3, 4, \dots, 12 (= 6 + 6)$ . There are  $6 \cdot 6 = 36$  equally likely outcomes  $(1, 1), (1, 2), \dots, (6, 6)$ , where the first number is that shown on the first die and the second number that on the other die. Each such outcome has probability  $1/36$ . Now  $X = 2$  occurs in the case of the outcome  $(1, 1)$ ;  $X = 3$  in the case of the two outcomes  $(1, 2)$  and  $(2, 1)$ ;  $X = 4$  in the case of the three outcomes  $(1, 3), (2, 2), (3, 1)$ ; and so on. Hence  $f(x) = P(X = x)$  and  $F(x) = P(X \leq x)$  have the values

$x$	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36
$F(x)$	1/36	3/36	6/36	10/36	15/36	21/36	26/36	30/36	33/36	35/36	36/36

Figure 514 shows a bar chart of this function and the graph of the distribution function, which is again a step function, with jumps (of different height!) at the possible values of  $X$ . ■

Two useful formulas for discrete distributions are readily obtained as follows. For the probability corresponding to intervals we have from (2) and (4)

$$(5) \quad P(a < X \leq b) = F(b) - F(a) = \sum_{a < x_j \leq b} p_j \quad (X \text{ discrete}).$$

This is the sum of all probabilities  $p_j$  for which  $x_j$  satisfies  $a < x_j \leq b$ . (Be careful about  $<$  and  $\leq$  !) From this and  $P(S) = 1$  (Sec. 24.3) we obtain the following formula.

$$(6) \quad \sum_j p_j = 1 \quad (\text{sum of all probabilities}).$$

### EXAMPLE 3 Illustration of Formula (5)

In Example 2, compute the probability of a sum of at least 4 and at most 8.

**Solution.**  $P(3 < X \leq 8) = F(8) - F(3) = \frac{26}{36} - \frac{3}{36} = \frac{23}{36}$ . ■

### EXAMPLE 4 Waiting Time Problem. Countably Infinite Sample Space

In tossing a fair coin, let  $X = \text{Number of trials until the first head appears}$ . Then, by independence of events (Sec. 24.3),

$$P(X = 1) = P(H) = \frac{1}{2} \quad (H = \text{Head})$$

$$P(X = 2) = P(TH) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \quad (T = \text{Tail})$$

$$P(X = 3) = P(TTH) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}, \quad \text{etc.}$$

and in general  $P(X = n) = \left(\frac{1}{2}\right)^n$ ,  $n = 1, 2, \dots$ . Also, (6) can be confirmed by the sum formula for the geometric series,

$$\begin{aligned} \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots &= -1 + \frac{1}{1 - \frac{1}{2}} \\ &= -1 + 2 = 1. \end{aligned} \quad \text{■}$$

## Continuous Random Variables and Distributions

Discrete random variables appear in experiments in which we **count** (defectives in a production, days of sunshine in Chicago, customers standing in a line, etc.). Continuous random variables appear in experiments in which we **measure** (lengths of screws, voltage in a power line, Brinell hardness of steel, etc.). By definition, a random variable  $X$  and its distribution are *of continuous type* or, briefly, **continuous**, if its distribution function  $F(x)$  [defined in (1)] can be given by an integral

$$(7) \quad F(x) = \int_{-\infty}^x f(v) dv$$

(we write  $v$  because  $x$  is needed as the upper limit of the integral) whose integrand  $f(x)$ , called the **density** of the distribution, is nonnegative, and is continuous, perhaps except for finitely many  $x$ -values. Differentiation gives the relation of  $f$  to  $F$  as

$$(8) \quad f(x) = F'(x)$$

for every  $x$  at which  $f(x)$  is continuous.

From (2) and (7) we obtain the very important formula for the probability corresponding to an interval:

$$(9) \quad P(a < X \leq b) = F(b) - F(a) = \int_a^b f(v) dv.$$

This is the analog of (5).

From (7) and  $P(S) = 1$  (Sec. 24.3) we also have the analog of (6):

$$(10) \quad \int_{-\infty}^{\infty} f(v) dv = 1.$$

Continuous random variables are **simpler than discrete ones** with respect to intervals. Indeed, in the continuous case the four probabilities corresponding to  $a < X \leq b$ ,  $a < X < b$ ,  $a \leq X < b$ , and  $a \leq X \leq b$  with any fixed  $a$  and  $b$  ( $b > a$ ) are all the same. Can you see why? (*Answer.* This probability is the area under the density curve, as in Fig. 515, and does not change by adding or subtracting a single point in the interval of integration.) This is different from the discrete case! (Explain.)

The next example illustrates notations and typical applications of our present formulas.

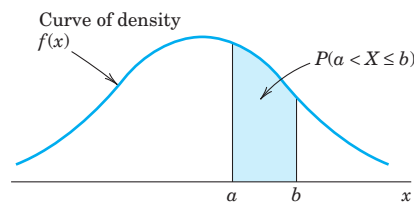


Fig. 515. Example illustrating formula (9)

### EXAMPLE 5 Continuous Distribution

Let  $X$  have the density function  $f(x) = 0.75(1 - x^2)$  if  $-1 \leq x \leq 1$  and zero otherwise. Find the distribution function. Find the probabilities  $P(-\frac{1}{2} \leq X \leq \frac{1}{2})$  and  $P(\frac{1}{4} \leq X \leq 2)$ . Find  $x$  such that  $P(X \leq x) = 0.95$ .

**Solution.** From (7) we obtain  $F(x) = 0$  if  $x \leq -1$ ,

$$F(x) = 0.75 \int_{-1}^x (1 - v^2) dv = 0.5 + 0.75x - 0.25x^3 \quad \text{if } -1 < x \leq 1,$$

and  $F(x) = 1$  if  $x > 1$ . From this and (9) we get

$$P(-\tfrac{1}{2} \leq X \leq \tfrac{1}{2}) = F(\tfrac{1}{2}) - F(-\tfrac{1}{2}) = 0.75 \int_{-1/2}^{1/2} (1 - v^2) dv = 68.75\%$$

(because  $P(-\frac{1}{2} \leq X \leq \frac{1}{2}) = P(-\frac{1}{2} < X \leq \frac{1}{2})$  for a continuous distribution) and

$$P(\frac{1}{4} \leq X \leq 2) = F(2) - F(\frac{1}{4}) = 0.75 \int_{1/4}^1 (1 - v^2) dv = 31.64\%.$$

(Note that the upper limit of integration is 1, not 2. Why?) Finally,

$$P(X \leq x) = F(x) = 0.5 + 0.75x - 0.25x^3 = 0.95.$$

Algebraic simplification gives  $3x - x^3 = 1.8$ . A solution is  $x = 0.73$ , approximately.

Sketch  $f(x)$  and mark  $x = -\frac{1}{2}, \frac{1}{2}, \frac{1}{4}$ , and  $0.73$ , so that you can see the results (the probabilities) as areas under the curve. Sketch also  $F(x)$ . ■

Further examples of continuous distributions are included in the next problem set and in later sections.

## PROBLEM SET 24.5

- Graph the probability function  $f(x) = kx^2$  ( $x = 1, 2, 3, 4, 5$ ;  $k$  suitable) and the distribution function.
- Graph the density function  $f(x) = kx^2$  ( $0 \leq x \leq 5$ ;  $k$  suitable) and the distribution function.
- Uniform distribution.** Graph  $f$  and  $F$  when the density of  $X$  is  $f(x) = k = \text{const}$  if  $-2 \leq x \leq 2$  and  $0$  elsewhere. Find  $P(0 \leq X \leq 2)$ .
- In Prob. 3 find  $c$  and  $\tilde{c}$  such that  $P(-c < X < c) = 95\%$  and  $P(0 < X < \tilde{c}) = 95\%$ .
- Graph  $f$  and  $F$  when  $f(-2) = f(2) = \frac{1}{8}$ ,  $f(-1) = f(1) = \frac{3}{8}$ . Can  $f$  have further positive values?
- A box contains 4 right-handed and 6 left-handed screws. Two screws are drawn at random without replacement. Let  $X$  be the number of left-handed screws drawn. Find the probabilities  $P(X = 0)$ ,  $P(X = 1)$ ,  $P(X = 2)$ ,  $P(1 < X < 2)$ ,  $P(X \leq 1)$ ,  $P(X \geq 1)$ ,  $P(X > 1)$ , and  $P(0.5 < X < 10)$ .
- Let  $X$  be the number of years before a certain kind of pump needs replacement. Let  $X$  have the probability function  $f(x) = kx^3$ ,  $x = 0, 1, 2, 3, 4$ . Find  $k$ . Sketch  $f$  and  $F$ .
- Graph the distribution function  $F(x) = 1 - e^{-3x}$  if  $x > 0$ ,  $F(x) = 0$  if  $x \leq 0$ , and the density  $f(x)$ . Find  $x$  such that  $F(x) = 0.9$ .
- Let  $X$  [millimeters] be the thickness of washers. Assume that  $X$  has the density  $f(x) = kx$  if  $0.9 < x < 1.1$  and  $0$  otherwise. Find  $k$ . What is the probability that a washer will have thickness between  $0.95$  mm and  $1.05$  mm?
- If the diameter  $X$  of axles has the density  $f(x) = k$  if  $119.9 \leq x \leq 120.1$  and  $0$  otherwise, how many defectives will a lot of 500 axles approximately contain if defectives are axles slimmer than  $119.91$  or thicker than  $120.09$ ?
- Find the probability that none of three bulbs in a traffic signal will have to be replaced during the first 1500 hours of operation if the lifetime  $X$  of a bulb is a random variable with the density  $f(x) = 6[0.25 - (x - 1.5)^2]$  when  $1 \leq x \leq 2$  and  $f(x) = 0$  otherwise, where  $x$  is measured in multiples of 1000 hours.
- Let  $X$  be the ratio of sales to profits of some company. Assume that  $X$  has the distribution function  $F(x) = 0$  if  $x < 2$ ,  $F(x) = (x^2 - 4)/5$  if  $2 \leq x < 3$ ,  $F(x) = 1$  if  $x \geq 3$ . Find and sketch the density. What is the probability that  $X$  is between  $2.5$  (40% profit) and  $5$  (20% profit)?
- Suppose that in an automatic process of filling oil cans, the content of a can (in gallons) is  $Y = 100 + X$ , where  $X$  is a random variable with density  $f(x) = 1 - |x|$  when  $|x| \leq 1$  and  $0$  when  $|x| > 1$ . Sketch  $f(x)$  and  $F(x)$ . In a lot of 1000 cans, about how many will contain 100 gallons or more? What is the probability that a can will contain less than 99.5 gallons? Less than 99 gallons?
- Find the probability function of  $X = \text{Number of times a fair die is rolled until the first Six appears}$  and show that it satisfies (6).
- Let  $X$  be a random variable that can assume every real value. What are the complements of the events  $X \leq b$ ,  $X < b$ ,  $X \geq c$ ,  $X > c$ ,  $b \leq X \leq c$ ,  $b < X \leq c$ ?



## 24.6 Mean and Variance of a Distribution

The mean  $\mu$  and variance  $\sigma^2$  of a random variable  $X$  and of its distribution are the theoretical counterparts of the mean  $\bar{x}$  and variance  $s^2$  of a frequency distribution in Sec. 24.1 and serve a similar purpose. Indeed, the mean characterizes the central location and the variance the spread (the variability) of the distribution. The **mean**  $\mu$  (mu) is defined by

$$(1) \quad \begin{aligned} (a) \quad \mu &= \sum_j x_j f(x_j) && \text{(Discrete distribution)} \\ (b) \quad \mu &= \int_{-\infty}^{\infty} x f(x) dx && \text{(Continuous distribution)} \end{aligned}$$

and the **variance**  $\sigma^2$  (sigma square) by

$$(2) \quad \begin{aligned} (a) \quad \sigma^2 &= \sum_j (x_j - \mu)^2 f(x_j) && \text{(Discrete distribution)} \\ (b) \quad \sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx && \text{(Continuous distribution).} \end{aligned}$$

$\sigma$  (the positive square root of  $\sigma^2$ ) is called the **standard deviation** of  $X$  and its distribution.  $f$  is the probability function or the density, respectively, in (a) and (b).

The mean  $\mu$  is also denoted by  $E(X)$  and is called the **expectation** of  $X$  because it gives the average value of  $X$  to be expected in many trials. Quantities such as  $\mu$  and  $\sigma^2$  that measure certain properties of a distribution are called **parameters**.  $\mu$  and  $\sigma^2$  are the two most important ones. From (2) we see that

$$(3) \quad \sigma^2 > 0$$

(except for a discrete “distribution” with only one possible value, so that  $\sigma^2 = 0$ ). We assume that  $\mu$  and  $\sigma^2$  exist (are finite), as is the case for practically all distributions that are useful in applications.

### EXAMPLE 1 Mean and Variance

The random variable  $X = \text{Number of heads in a single toss of a fair coin}$  has the possible values  $X = 0$  and  $X = 1$  with probabilities  $P(X = 0) = \frac{1}{2}$  and  $P(X = 1) = \frac{1}{2}$ . From (1a) we thus obtain the mean  $\mu = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$ , and (2a) yields the variance

$$\sigma^2 = (0 - \frac{1}{2})^2 \cdot \frac{1}{2} + (1 - \frac{1}{2})^2 \cdot \frac{1}{2} = \frac{1}{4}.$$

### EXAMPLE 2 Uniform Distribution. Variance Measures Spread

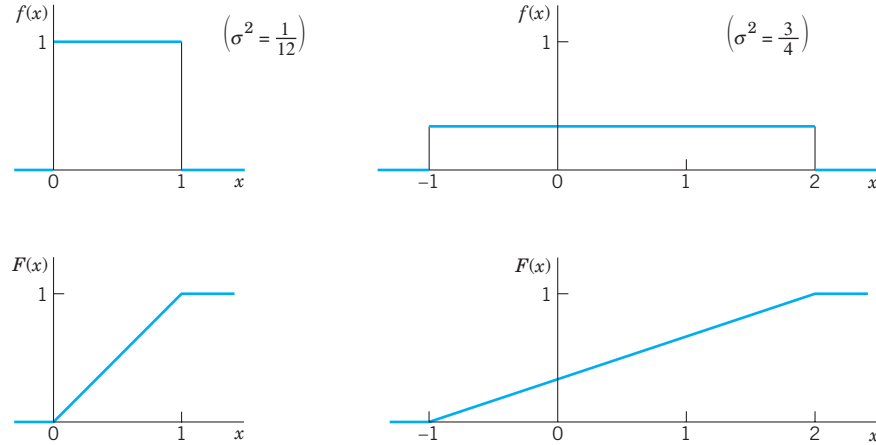
The distribution with the density

$$f(x) = \frac{1}{b-a} \quad \text{if} \quad a < x < b$$

and  $f = 0$  otherwise is called the **uniform distribution** on the interval  $a < x < b$ . From (1b) (or from Theorem 1, below) we find that  $\mu = (a + b)/2$ , and (2b) yields the variance

$$\sigma^2 = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} dx = \frac{(b-a)^2}{12}.$$

Figure 516 illustrates that the spread is large if and only if  $\sigma^2$  is large. ■



**Fig. 516.** Uniform distributions having the same mean (0.5) but different variances  $\sigma^2$

**Symmetry.** We can obtain the mean  $\mu$  without calculation if a distribution is symmetric. Indeed, you may prove

### THEOREM 1

#### Mean of a Symmetric Distribution

If a distribution is **symmetric** with respect to  $x = c$ , that is,  $f(c - x) = f(c + x)$ , then  $\mu = c$ . (Examples 1 and 2 illustrate this.)

## Transformation of Mean and Variance

Given a random variable  $X$  with mean  $\mu$  and variance  $\sigma^2$ , we want to calculate the mean and variance of  $X^* = a_1 + a_2X$ , where  $a_1$  and  $a_2$  are given constants. This problem is important in statistics, where it often appears.

### THEOREM 2

#### Transformation of Mean and Variance

(a) If a random variable  $X$  has mean  $\mu$  and variance  $\sigma^2$ , then the random variable

$$(4) \quad X^* = a_1 + a_2X \quad (a_2 > 0)$$

has the mean  $\mu^*$  and variance  $\sigma^{*2}$ , where

$$(5) \quad \mu^* = a_1 + a_2\mu \quad \text{and} \quad \sigma^{*2} = a_2^2\sigma^2.$$

(b) In particular, the **standardized random variable**  $Z$  corresponding to  $X$ , given by

$$(6) \quad Z = \frac{X - \mu}{\sigma}$$

has the mean 0 and the variance 1.

**PROOF** We prove (5) for a continuous distribution. To a small interval  $I$  of length  $\Delta x$  on the  $x$ -axis there corresponds the probability  $f(x)\Delta x$  [approximately; the area of a rectangle of base  $\Delta x$  and height  $f(x)$ ]. Then the probability  $f(x)\Delta x$  must equal that for the corresponding interval on the  $x^*$ -axis, that is,  $f^*(x^*)\Delta x^*$ , where  $f^*$  is the density of  $X^*$  and  $\Delta x^*$  is the length of the interval on the  $x^*$ -axis corresponding to  $I$ . Hence for differentials we have  $f^*(x^*) dx^* = f(x) dx$ . Also,  $x^* = a_1 + a_2 x$  by (4), so that (1b) applied to  $X^*$  gives

$$\begin{aligned} \mu^* &= \int_{-\infty}^{\infty} x^* f^*(x^*) dx^* \\ &= \int_{-\infty}^{\infty} (a_1 + a_2 x) f(x) dx \\ &= a_1 \int_{-\infty}^{\infty} f(x) dx + a_2 \int_{-\infty}^{\infty} x f(x) dx. \end{aligned}$$

On the right the first integral equals 1, by (10) in Sec. 24.5. The second integral is  $\mu$ . This proves (5) for  $\mu^*$ . It implies

$$x^* - \mu^* = (a_1 + a_2 x) - (a_1 + a_2 \mu) = a_2(x - \mu).$$

From this and (2) applied to  $X^*$ , again using  $f^*(x^*) dx^* = f(x) dx$ , we obtain the second formula in (5),

$$\sigma^{*2} = \int_{-\infty}^{\infty} (x^* - \mu^*)^2 f^*(x^*) dx^* = a_2^2 \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = a_2^2 \sigma^2.$$

For a discrete distribution the proof of (5) is similar.

Choosing  $a_1 = -\mu/\sigma$  and  $a_2 = 1/\sigma$  we obtain (6) from (4), writing  $X^* = Z$ . For these  $a_1, a_2$  formula (5) gives  $\mu^* = 0$  and  $\sigma^{*2} = 1$ , as claimed in (b). ■

## Expectation, Moments

Recall that (1) defines the expectation (the mean) of  $X$ , the value of  $X$  to be expected on the average, written  $\mu = E(X)$ . More generally, if  $g(x)$  is nonconstant and continuous for all  $x$ , then  $g(X)$  is a random variable. Hence its *mathematical expectation* or, briefly, its

**expectation**  $E(g(X))$  is the value of  $g(X)$  to be expected on the average, defined [similarly to (1)] by

$$(7) \quad E(g(X)) = \sum_j g(x_j)f(x_j) \quad \text{or} \quad E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x) dx.$$

In the first formula,  $f$  is the probability function of the discrete random variable  $X$ . In the second formula,  $f$  is the density of the continuous random variable  $X$ . Important special cases are the  **$k$ th moment** of  $X$  (where  $k = 1, 2, \dots$ )

$$(8) \quad E(X^k) = \sum_j x_j^k f(x_j) \quad \text{or} \quad \int_{-\infty}^{\infty} x^k f(x) dx$$

and the  **$k$ th central moment** of  $X$  ( $k = 1, 2, \dots$ )

$$(9) \quad E([X - \mu]^k) = \sum_j (x_j - \mu)^k f(x_j) \quad \text{or} \quad \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx.$$

This includes the first moment, the **mean** of  $X$

$$(10) \quad \mu = E(X) \quad \text{[(8) with } k = 1].$$

It also includes the second central moment, the **variance** of  $X$

$$(11) \quad \sigma^2 = E([X - \mu]^2) \quad \text{[(9) with } k = 2].$$

For later use you may prove

$$(12) \quad E(1) = 1.$$

## PROBLEM SET 24.6

### 1-8 MEAN, VARIANCE

Find the mean and variance of the random variable  $X$  with probability function or density  $f(x)$ .

1.  $f(x) = kx$  ( $0 \leq x \leq 2$ ,  $k$  suitable)
2.  $X$  = Number a fair die turns up
3. Uniform distribution on  $[0, 2\pi]$
4.  $Y = \sqrt{3}(X - \mu)/\pi$  with  $X$  as in Prob. 3
5.  $f(x) = 4e^{-4x}$  ( $x \geq 0$ )
6.  $f(x) = k(1 - x^2)$  if  $-1 \leq x \leq 1$  and 0 otherwise
7.  $f(x) = Ce^{-x/2}$  ( $x = 0$ )
8.  $X$  = Number of times a fair coin is flipped until the first Head appears. (Calculate  $\mu$  only.)
9. If the diameter  $X$  [cm] of certain bolts has the density  $f(x) = k(x - 0.9)(1.1 - x)$  for  $0.9 < x < 1.1$  and 0 for other  $x$ , what are  $k$ ,  $\mu$ , and  $\sigma^2$ ? Sketch  $f(x)$ .

10. If, in Prob. 9, a defective bolt is one that deviates from 1.00 cm by more than 0.06 cm, what percentage of defectives should we expect?
11. For what choice of the maximum possible deviation from 1.00 cm shall we obtain 10% defectives in Probs. 9 and 10?
12. What total sum can you expect in rolling a fair die 20 times? Do the experiment. Repeat it a number of times and record how the sum varies.
13. What is the expected daily profit if a store sells  $X$  air conditioners per day with probability  $f(10) = 0.1$ ,  $f(11) = 0.3$ ,  $f(12) = 0.4$ ,  $f(13) = 0.2$  and the profit per conditioner is \$55?
14. Find the expectation of  $g(X) = X^2$ , where  $X$  is uniformly distributed on the interval  $-1 \leq x \leq 1$ .

15. A small filling station is supplied with gasoline every **Saturday** afternoon. Assume that its volume  $X$  of sales in ten thousands of gallons has the probability density  $f(x) = 6x(1-x)$  if  $0 \leq x \leq 1$  and 0 otherwise. Determine the mean, the variance, and the standardized variable.
16. What capacity must the tank in Prob. 15 have in order that the probability that the tank will be emptied in a given week be 5%?
17. James rolls 2 fair dice, and Harry pays  $k$  cents to James, where  $k$  is the product of the two faces that show on the dice. How much should James pay to Harry for each game to make the game fair?
18. What is the mean life of a lightbulb whose life  $X$  [hours] has the density  $f(x) = 0.001e^{-0.001x}$  ( $x \geq 0$ )?
19. Let  $X$  be discrete with probability function  $f(0) = f(3) = \frac{1}{8}$ ,  $f(1) = f(2) = \frac{3}{8}$ . Find the expectation of  $X^3$ .
20. **TEAM PROJECT. Means, Variances, Expectations.**
- (a) Show that  $E(X - \mu) = 0$ ,  $\sigma^2 = E(X^2) - \mu^2$ .
- (b) Prove (10)–(12).
- (c) Find all the moments of the uniform distribution on an interval  $a \leq x \leq b$ .
- (d) The **skewness**  $\gamma$  of a random variable  $X$  is defined by

$$(13) \quad \gamma = \frac{1}{\sigma^3} E[(X - \mu)^3].$$

Show that for a symmetric distribution (whose third central moment exists) the skewness is zero.

(e) Find the skewness of the distribution with density  $f(x) = xe^{-x}$  when  $x > 0$  and  $f(x) = 0$  otherwise. Sketch  $f(x)$ .

(f) Calculate the skewness of a few simple discrete distributions of your own choice.

(g) Find a *nonsymmetric* discrete distribution with 3 possible values, mean 0, and skewness 0.

## 24.7 Binomial, Poisson, and Hypergeometric Distributions

These are the three most important *discrete* distributions, with numerous applications.

### Binomial Distribution

The **binomial distribution** occurs in games of chance (rolling a die, see below, etc.), quality inspection (e.g., counting of the number of defectives), opinion polls (counting number of employees favoring certain schedule changes, etc.), medicine (e.g., recording the number of patients who recovered on a new medication), and so on. The conditions of its occurrence are as follows.

We are interested in the number of times an event  $A$  occurs in  $n$  independent trials. In each trial the event  $A$  has the same probability  $P(A) = p$ . Then in a trial,  $A$  will *not* occur with probability  $q = 1 - p$ . In  $n$  trials the random variable that interests us is

$X = \text{Number of times the event } A \text{ occurs in } n \text{ trials.}$

$X$  can assume the values  $0, 1, \dots, n$ , and we want to determine the corresponding probabilities. Now  $X = x$  means that  $A$  occurs in  $x$  trials and in  $n - x$  trials it does not occur. This may look as follows.

$$(1) \quad \underbrace{A \ A \cdots A}_{x \text{ times}} \quad \underbrace{B \ B \cdots B}_{n - x \text{ times}}$$

Here  $B = A^c$  is the complement of  $A$ , meaning that  $A$  does not occur (Sec. 24.2). We now use the assumption that the trials are independent, that is, they do not influence each other. Hence (1) has the probability (see Sec. 24.3 on independent events)

$$(1^*) \quad \underbrace{pp \cdots p}_{x \text{ times}} \cdot \underbrace{qq \cdots q}_{n-x \text{ times}} = p^x q^{n-x}.$$

Now (1) is just one order of arranging  $x$  A's and  $n - x$  B's. We now use Theorem 1(b) in Sec. 24.4, which gives the number of permutations of  $n$  things (the  $n$  outcomes of the  $n$  trials) consisting of 2 classes, class 1 containing the  $n_1 = x$  A's and class 2 containing the  $n - n_1 = n - x$  B's. This number is

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

Accordingly, (1\*), multiplied by this binomial coefficient, gives the probability  $P(X = x)$  of  $X = x$ , that is, of obtaining A precisely  $x$  times in  $n$  trials. Hence  $X$  has the probability function

$$(2) \quad f(x) = \binom{n}{x} p^x q^{n-x} \quad (x = 0, 1, \dots, n)$$

and  $f(x) = 0$  otherwise. The distribution of  $X$  with probability function (2) is called the **binomial distribution** or *Bernoulli distribution*. The occurrence of A is called *success* (regardless of what it actually is; it may mean that you miss your plane or lose your watch) and the nonoccurrence of A is called *failure*. Figure 517 shows typical examples. Numeric values can be obtained from Table A5 in App. 5 or from your CAS.

The mean of the binomial distribution is (see Team Project 16)

$$(3) \quad \mu = np$$

and the variance is (see Team Project 16)

$$(4) \quad \sigma^2 = npq.$$

For the *symmetric case* of equal chance of success and failure ( $p = q = \frac{1}{2}$ ) this gives the mean  $n/2$ , the variance  $n/4$ , and the probability function

$$(2^*) \quad f(x) = \binom{n}{x} \left(\frac{1}{2}\right)^n \quad (x = 0, 1, \dots, n).$$

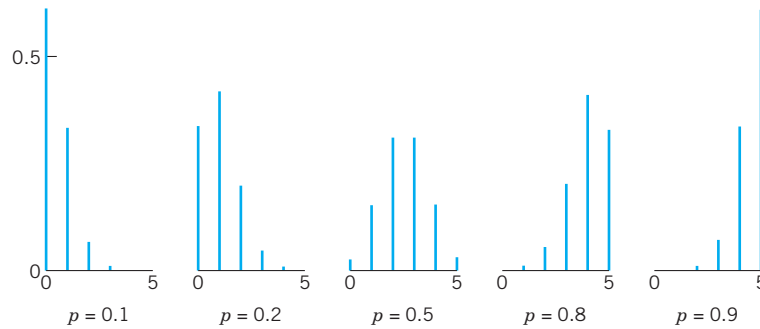


Fig. 517. Probability function (2) of the binomial distribution for  $n = 5$  and various values of  $p$

**EXAMPLE 1 Binomial Distribution**

Compute the probability of obtaining at least two “Six” in rolling a fair die 4 times.

**Solution.**  $p = P(A) = P(\text{“Six”}) = \frac{1}{6}$ ,  $q = \frac{5}{6}$ ,  $n = 4$ . The event “At least two ‘Six’” occurs if we obtain 2 or 3 or 4 “Six.” Hence the answer is

$$\begin{aligned} P = f(2) + f(3) + f(4) &= \binom{4}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 + \binom{4}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right) + \binom{4}{4} \left(\frac{1}{6}\right)^4 \\ &= \frac{1}{6^4} (6 \cdot 25 + 4 \cdot 5 + 1) = \frac{171}{1296} = 13.2\%. \end{aligned}$$

**Poisson Distribution**

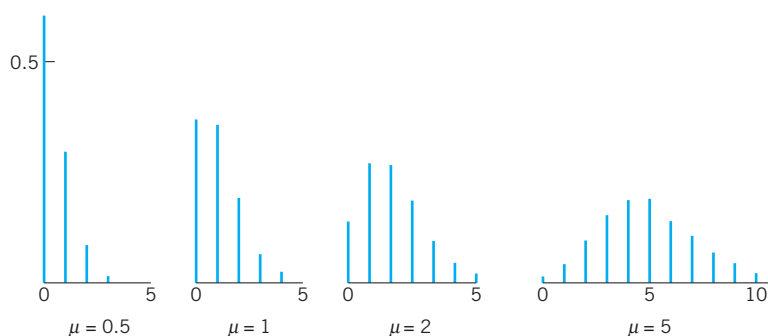
The discrete distribution with infinitely many possible values and probability function

$$(5) \quad f(x) = \frac{\mu^x}{x!} e^{-\mu} \quad (x = 0, 1, \dots)$$

is called the **Poisson distribution**, named after S. D. Poisson (Sec. 18.5). Figure 518 shows (5) for some values of  $\mu$ . It can be proved that this distribution is obtained as a limiting case of the binomial distribution, if we let  $p \rightarrow 0$  and  $n \rightarrow \infty$  so that the mean  $\mu = np$  approaches a finite value. (For instance,  $\mu = np$  may be kept constant.) The Poisson distribution has the mean  $\mu$  and the variance (see Team Project 16)

$$(6) \quad \sigma^2 = \mu.$$

Figure 518 gives the impression that, with increasing mean, the spread of the distribution increases, thereby illustrating formula (6), and that the distribution becomes more and more (approximately) symmetric.



**Fig. 518.** Probability function (5) of the Poisson distribution for various values of  $\mu$

**EXAMPLE 2 Poisson Distribution**

If the probability of producing a defective screw is  $p = 0.01$ , what is the probability that a lot of 100 screws will contain more than 2 defectives?

**Solution.** The complementary event is  $A^c$ : *Not more than 2 defectives*. For its probability we get, from the binomial distribution with mean  $\mu = np = 1$ , the value [see (2)]

$$P(A^c) = \binom{100}{0} 0.99^{100} + \binom{100}{1} 0.01 \cdot 0.99^{99} + \binom{100}{2} 0.01^2 \cdot 0.99^{98}.$$

Since  $p$  is very small, we can approximate this by the much more convenient Poisson distribution with mean  $\mu = np = 100 \cdot 0.01 = 1$ , obtaining [see (5)]

$$\begin{aligned} P(A^c) &\approx e^{-1} \left(1 + 1 + \frac{1}{2}\right) \\ &= 91.97\%. \end{aligned}$$

Thus  $P(A) = 8.03\%$ . Show that the binomial distribution gives  $P(A) = 7.94\%$ , so that the Poisson approximation is quite good. ■

### EXAMPLE 3 Parking Problems. Poisson Distribution

If on the average, 2 cars enter a certain parking lot per minute, what is the probability that during any given minute 4 or more cars will enter the lot?

**Solution.** To understand that the Poisson distribution is a model of the situation, we imagine the minute to be divided into very many short time intervals, let  $p$  be the (constant) probability that a car will enter the lot during any such short interval, and assume independence of the events that happen during those intervals. Then we are dealing with a binomial distribution with very large  $n$  and very small  $p$ , which we can approximate by the Poisson distribution with

$$\mu = np = 2,$$

because 2 cars enter on the average. The complementary event of the event “4 cars or more during a given minute” is “3 cars or fewer enter the lot” and has the probability

$$\begin{aligned} f(0) + f(1) + f(2) + f(3) &= e^{-2} \left( \frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} \right) \\ &= 0.857. \end{aligned}$$

Answer: 14.3%. (Why did we consider that complement?) ■

## Sampling with Replacement

This means that we draw things from a given set one by one, and after each trial we replace the thing drawn (put it back to the given set and mix) before we draw the next thing. This guarantees independence of trials and leads to the **binomial distribution**. Indeed, if a box contains  $N$  things, for example, screws,  $M$  of which are defective, the probability of drawing a defective screw in a trial is  $p = M/N$ . Hence the probability of drawing a nondefective screw is  $q = 1 - p = 1 - M/N$ , and (2) gives the probability of drawing  $x$  defectives in  $n$  trials in the form

$$(7) \quad f(x) = \binom{n}{x} \left(\frac{M}{N}\right)^x \left(1 - \frac{M}{N}\right)^{n-x} \quad (x = 0, 1, \dots, n).$$

## Sampling without Replacement. Hypergeometric Distribution

**Sampling without replacement** means that we return no screw to the box. Then we no longer have independence of trials (why?), and instead of (7) the probability of drawing  $x$  defectives in  $n$  trials is



$$(8) \quad f(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \quad (x = 0, 1, \dots, n).$$

The distribution with this probability function is called the **hypergeometric distribution** (because its moment generating function (see Team Project 16) can be expressed by the hypergeometric function defined in Sec. 5.4, a fact that we shall not use).

**Derivation of (8).** By (4a) in Sec. 24.4 there are

- (a)  $\binom{N}{n}$  different ways of picking  $n$  things from  $N$ ,
- (b)  $\binom{M}{x}$  different ways of picking  $x$  defectives from  $M$ ,
- (c)  $\binom{N-M}{n-x}$  different ways of picking  $n-x$  nondefectives from  $N-M$ ,

and each way in (b) combined with each way in (c) gives the total number of mutually exclusive ways of obtaining  $x$  defectives in  $n$  drawings without replacement. Since (a) is the total number of outcomes and we draw at random, each such way has the probability  $1/\binom{N}{n}$ . From this, (8) follows. ■

The hypergeometric distribution has the mean (Team Project 16)

$$(9) \quad \mu = n \frac{M}{N}$$

and the variance

$$(10) \quad \sigma^2 = \frac{nM(N-M)(N-n)}{N^2(N-1)}.$$

#### EXAMPLE 4 Sampling with and without Replacement

We want to draw random samples of two gaskets from a box containing 10 gaskets, three of which are defective. Find the probability function of the random variable  $X = \text{Number of defectives in the sample}$ .

**Solution.** We have  $N = 10$ ,  $M = 3$ ,  $N - M = 7$ ,  $n = 2$ . For sampling with replacement, (7) yields

$$f(x) = \binom{2}{x} \left(\frac{3}{10}\right)^x \left(\frac{7}{10}\right)^{2-x}, \quad f(0) = 0.49, \quad f(1) = 0.42, \quad f(2) = 0.09.$$

For sampling without replacement we have to use (8), finding

$$f(x) = \frac{\binom{3}{x} \binom{7}{2-x}}{\binom{10}{2}}, \quad f(0) = f(1) = \frac{21}{45} \approx 0.47, \quad f(2) = \frac{3}{45} \approx 0.07. \quad \blacksquare$$

If  $N$ ,  $M$ , and  $N - M$  are large compared with  $n$ , then it does not matter too much whether we sample with or without replacement, and in this case the hypergeometric distribution may be approximated by the binomial distribution (with  $p = M/N$ ), which is somewhat simpler.

Hence, in sampling from an indefinitely large population (“infinite population”), we may use the binomial distribution, regardless of whether we sample with or without replacement.

## PROBLEM SET 24.7

1. Mark the positions of  $\mu$  in Fig. 517. Comment.
2. Graph (2) for  $n = 8$  as in Fig. 517 and compare with Fig. 517.
3. In Example 3, if 5 cars enter the lot on the average, what is the probability that during any given minute 6 or more cars will enter? First guess. Compare with Example 3.
4. How do the probabilities in Example 4 of the text change if you double the numbers: drawing 4 gaskets from 20, 6 of which are defective? First guess.
5. Five fair coins are tossed simultaneously. Find the probability function of the random variable  $X = \text{Number of heads}$  and compute the probabilities of obtaining no heads, precisely 1 head, at least 1 head, not more than 4 heads.
6. Suppose that 4% of steel rods made by a machine are defective, the defectives occurring at random during production. If the rods are packaged 100 per box, what is the Poisson approximation of the probability that a given box will contain  $x = 0, 1, \dots, 5$  defectives?
7. Let  $X$  be the number of cars per minute passing a certain point of some road between 8 A.M. and 10 A.M. on a Sunday. Assume that  $X$  has a Poisson distribution with mean 5. Find the probability of observing 4 or fewer cars during any given minute.
8. Suppose that a telephone switchboard of some company on the average handles 300 calls per hour, and that the board can make at most 10 connections per minute. Using the Poisson distribution, estimate the probability that the board will be overtaxed during a given minute. (Use Table A6 in App. 5 or your CAS.)
9. **Rutherford–Geiger experiments.** In 1910, E. Rutherford and H. Geiger showed experimentally that the number of alpha particles emitted per second in a radioactive process is a random variable  $X$  having a Poisson distribution. If  $X$  has mean 0.5, what is the probability of observing two or more particles during any given second?
10. Let  $p = 2\%$  be the probability that a certain type of lightbulb will fail in a 24-hour test. Find the probability that a sign consisting of 15 such bulbs will burn 24 hours with no bulb failures.
11. Guess how much less the probability in Prob. 10 would be if the sign consisted of 100 bulbs. Then calculate.
12. Suppose that a certain type of magnetic tape contains, on the average, 2 defects per 100 meters. What is the probability that a roll of tape 300 meters long will contain (a)  $x$  defects, (b) no defects?
13. Suppose that a test for extrasensory perception consists of naming (in any order) 3 cards randomly drawn from a deck of 13 cards. Find the probability that by chance alone, the person will correctly name (a) no cards, (b) 1 card, (c) 2 cards, (d) 3 cards.
14. If a ticket office can serve at most 4 customers per minute and the average number of customers is 120 per hour, what is the probability that during a given minute customers will have to wait? (Use the Poisson distribution, Table 6 in Appendix 5.)
15. Suppose that in the production of 60-ohm radio resistors, nondefective items are those that have a resistance between 58 and 62 ohms and the probability of a resistor’s being defective is 0.1%. The resistors are sold in lots of 200, with the guarantee that all resistors are nondefective. What is the probability that a given lot will violate this guarantee? (Use the Poisson distribution.)
16. **TEAM PROJECT. Moment Generating Function.** The moment generating function  $G(t)$  is defined by
 
$$G(t) = E(e^{tX_j}) = \sum_j e^{tx_j} f(x_j)$$
 or
 
$$G(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$
 where  $X$  is a discrete or continuous random variable, respectively.
 

(a) Assuming that termwise differentiation and differentiation under the integral sign are permissible, show

that  $E(X^k) = G^{(k)}(0)$ , where  $G^{(k)} = d^k G/dt^k$ , in particular,  $\mu = G'(0)$ .

(b) Show that the binomial distribution has the moment generating function

$$G(t) = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^n \binom{n}{x} (pe^t)^x q^{n-x} = (pe^t + q)^n.$$

(c) Using (b), prove (3).

(d) Prove (4).

(e) Show that the Poisson distribution has the moment generating function  $G(t) = e^{-\mu} e^{\mu e^t}$  and prove (6).

(f) Prove  $x \binom{M}{x} = M \binom{M-1}{x-1}$ .

Using this, prove (9).

17. **Multinomial distribution.** Suppose a trial can result in precisely one of  $k$  mutually exclusive events

$A_1, \dots, A_k$  with probabilities  $p_1, \dots, p_k$ , respectively, where  $p_1 + \dots + p_k = 1$ . Suppose that  $n$  independent trials are performed. Show that the probability of getting  $x_1 A_1$ 's,  $\dots$ ,  $x_k A_k$ 's is

$$f(x_1, \dots, x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

where  $0 \leq x_j \leq n$ ,  $j = 1, \dots, k$ , and  $x_1 + \dots + x_k = n$ . The distribution having this probability function is called the *multinomial distribution*.

18. A process of manufacturing screws is checked every hour by inspecting  $n$  screws selected at random from that hour's production. If one or more screws are defective, the process is halted and carefully examined. How large should  $n$  be if the manufacturer wants the probability to be about 95% that the process will be halted when 10% of the screws being produced are defective? (Assume independence of the quality of any screw from that of the other screws.)

## 24.8 Normal Distribution

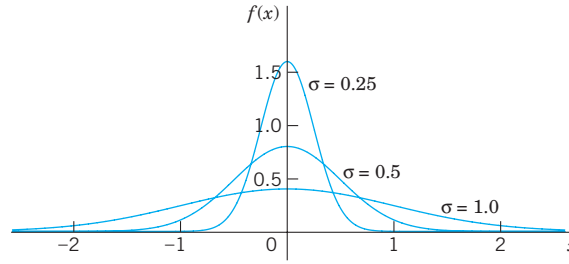
Turning from discrete to continuous distributions, in this section we discuss the normal distribution. This is the most important continuous distribution because in applications many random variables are **normal random variables** (that is, they have a normal distribution) or they are approximately normal or can be transformed into normal random variables in a relatively simple fashion. Furthermore, the normal distribution is a useful approximation of more complicated distributions, and it also occurs in the proofs of various statistical tests.

The **normal distribution** or *Gauss distribution* is defined as the distribution with the density

$$(1) \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right] \quad (\sigma > 0)$$

where  $\exp$  is the exponential function with base  $e = 2.718 \dots$ . This is simpler than it may at first look.  $f(x)$  has these features (see also Fig. 519).

1.  $\mu$  is the mean and  $\sigma$  the standard deviation.
2.  $1/(\sigma\sqrt{2\pi})$  is a constant factor that makes the area under the curve of  $f(x)$  from  $-\infty$  to  $\infty$  equal to 1, as it must be by (10), Sec. 24.5.
3. The curve of  $f(x)$  is symmetric with respect to  $x = \mu$  because the exponent is quadratic. Hence for  $\mu = 0$  it is symmetric with respect to the  $y$ -axis  $x = 0$  (Fig. 519, "bell-shaped curves").
4. The exponential function in (1) goes to zero very fast—the faster the smaller the standard deviation  $\sigma$  is, as it should be (Fig. 519).



**Fig. 519.** Density (1) of the normal distribution with  $\mu = 0$  for various values of  $\sigma$

## Distribution Function $F(x)$

From (7) in Sec. 24.5 and (1) we see that the normal distribution has the **distribution function**

$$(2) \quad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp \left[ -\frac{1}{2} \left( \frac{v - \mu}{\sigma} \right)^2 \right] dv.$$

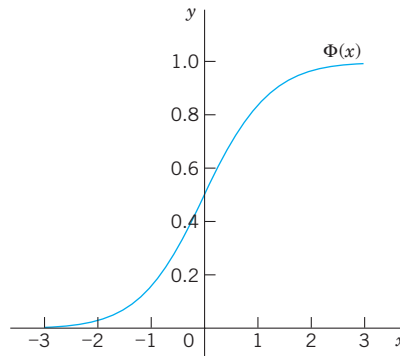
Here we needed  $x$  as the upper limit of integration and wrote  $v$  (instead of  $x$ ) in the integrand.

For the corresponding **standardized normal distribution** with mean 0 and standard deviation 1 we denote  $F(x)$  by  $\Phi(z)$ . Then we simply have from (2)

$$(3) \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du.$$

This integral cannot be integrated by one of the methods of calculus. But this is no serious handicap because its values can be obtained from Table A7 in App. 5 or from your CAS. These values are needed in working with the normal distribution. The curve of  $\Phi(z)$  is S-shaped. It increases monotone (why?) from 0 to 1 and intersects the vertical axis at  $\frac{1}{2}$  (why?), as shown in Fig. 520.

**Relation Between  $F(x)$  and  $\Phi(z)$ .** Although your CAS will give you values of  $F(x)$  in (2) with any  $\mu$  and  $\sigma$  directly, it is important to comprehend that and why any such an  $F(x)$  can be expressed in terms of the tabulated standard  $\Phi(z)$ , as follows.



**Fig. 520.** Distribution function  $\Phi(z)$  of the normal distribution with mean 0 and variance 1

**THEOREM 1****Use of the Normal Table A7 in App. 5**

The distribution function  $F(x)$  of the normal distribution with any  $\mu$  and  $\sigma$  [see (2)] is related to the standardized distribution function  $\Phi(z)$  in (3) by the formula

$$(4) \quad F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

**PROOF** Comparing (2) and (3) we see that we should set

$$u = \frac{v - \mu}{\sigma}. \quad \text{Then } v = x \text{ gives } u = \frac{x - \mu}{\sigma}$$

as the new upper limit of integration. Also  $v - \mu = \sigma u$ , thus  $dv = \sigma du$ . Together, since  $\sigma$  drops out,

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-u^2/2} \sigma du = \Phi\left(\frac{x - \mu}{\sigma}\right). \quad \blacksquare$$

Probabilities corresponding to intervals will be needed quite frequently in statistics in Chap. 25. These are obtained as follows.

**THEOREM 2****Normal Probabilities for Intervals**

The probability that a normal random variable  $X$  with mean  $\mu$  and standard deviation  $\sigma$  assume any value in an interval  $a < x \leq b$  is

$$(5) \quad P(a < X \leq b) = F(b) - F(a) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

**PROOF** Formula (2) in Sec. 24.5 gives the first equality in (5), and (4) in this section gives the second equality. ■

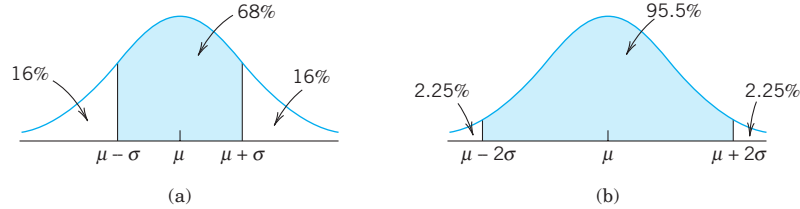
## Numeric Values

In practical work with the normal distribution it is good to remember that about  $\frac{2}{3}$  of all values of  $X$  to be observed will lie between  $\mu \pm \sigma$ , about 95% between  $\mu \pm 2\sigma$ , and practically all between the **three-sigma limits**  $\mu \pm 3\sigma$ . More precisely, by Table A7 in App. 5,

$$(6) \quad \begin{aligned} (a) \quad & P(\mu - \sigma < X \leq \mu + \sigma) \approx 68\% \\ (b) \quad & P(\mu - 2\sigma < X \leq \mu + 2\sigma) \approx 95.5\% \\ (c) \quad & P(\mu - 3\sigma < X \leq \mu + 3\sigma) \approx 99.7\%. \end{aligned}$$

Formulas (6a) and (6b) are illustrated in Fig. 521.

The formulas in (6) show that a value deviating from  $\mu$  by more than  $\sigma$ ,  $2\sigma$ , or  $3\sigma$  will occur in one of about 3, 20, and 300 trials, respectively.



**Fig. 521.** Illustration of formula (6)

In tests (Chap. 25) we shall ask, conversely, for the intervals that correspond to certain given probabilities; practically most important are the probabilities of 95%, 99%, and 99.9%. For these, Table A8 in App. 5 gives the answers  $\mu \pm 2\sigma$ ,  $\mu \pm 2.6\sigma$ , and  $\mu \pm 3.3\sigma$ , respectively. More precisely,

- (7)      (a)  $P(\mu - 1.96\sigma < X \leq \mu + 1.96\sigma) = 95\%$   
             (b)  $P(\mu - 2.58\sigma < X \leq \mu + 2.58\sigma) = 99\%$   
             (c)  $P(\mu - 3.29\sigma < X \leq \mu + 3.29\sigma) = 99.9\%.$

## Working with the Normal Tables A7 and A8 in App. 5

There are two normal tables in App. 5, Tables A7 and A8. If you want probabilities, use Table A7. If probabilities are given and corresponding intervals or  $x$ -values are wanted, use Table A8. The following examples are typical. Do them with care, verifying all values, and don't just regard them as dull exercises for your software. Make sketches of the density to see whether the results look reasonable.

### EXAMPLE 1 Reading Entries from Table A7

If  $X$  is standardized normal (so that  $\mu = 0$ ,  $\sigma = 1$ ), then

$$P(X \leq 2.44) = 0.9927 \approx 99\frac{1}{4}\%$$

$$P(X \leq -1.16) = 1 - \Phi(1.16) = 1 - 0.8770 = 0.1230 = 12.3\%$$

$$P(X \geq 1) = 1 - P(X \leq 1) = 1 - 0.8413 = 0.1587 \text{ by (7), Sec. 24.3}$$

$$P(1.0 \leq X \leq 1.8) = \Phi(1.8) - \Phi(1.0) = 0.9641 - 0.8413 = 0.1228. \quad \blacksquare$$

### EXAMPLE 2 Probabilities for Given Intervals, Table A7

Let  $X$  be normal with mean 0.8 and variance 4 (so that  $\sigma = 2$ ). Then by (4) and (5)

$$P(X \leq 2.44) = F(2.44) = \Phi\left(\frac{2.44 - 0.80}{2}\right) = \Phi(0.82) = 0.7939 \approx 80\%$$

or, if you like it better, (similarly in the other cases)

$$P(X \leq 2.44) = P\left(\frac{X - 0.80}{2} \leq \frac{2.44 - 0.80}{2}\right) = P(Z \leq 0.82) = 0.7939$$

$$P(X \geq 1) = 1 - P(X \leq 1) = 1 - \Phi\left(\frac{1 - 0.8}{2}\right) = 1 - 0.5398 = 0.4602$$

$$P(1.0 \leq X \leq 1.8) = \Phi(0.5) - \Phi(0.1) = 0.6915 - 0.5398 = 0.1517. \quad \blacksquare$$

**EXAMPLE 3** Unknown Values  $c$  for Given Probabilities, Table A8

Let  $X$  be normal with mean 5 and variance 0.04 (hence standard deviation 0.2). Find  $c$  or  $k$  corresponding to the given probability

$$P(X \leq c) = 95\%, \quad \Phi\left(\frac{c-5}{0.2}\right) = 95\%, \quad \frac{c-5}{0.2} = 1.645, \quad c = 5.329$$

$$P(5-k \leq X \leq 5+k) = 90\%, \quad 5+k = 5.329 \quad (\text{as before; why?})$$

$$P(X \geq c) = 1\%, \quad \text{thus } P(X \leq c) = 99\%, \quad \frac{c-5}{0.2} = 2.326, \quad c = 5.465. \quad \blacksquare$$

**EXAMPLE 4** Defectives

In a production of iron rods let the diameter  $X$  be normally distributed with mean 2 in. and standard deviation 0.008 in.

(a) What percentage of defectives can we expect if we set the tolerance limits at  $2 \pm 0.02$  in.?

(b) How should we set the tolerance limits to allow for 4% defectives?

**Solution.** (a)  $1\frac{1}{4}\%$  because from (5) and Table A7 we obtain for the complementary event the probability

$$\begin{aligned} P(1.98 \leq X \leq 2.02) &= \Phi\left(\frac{2.02-2.00}{0.008}\right) - \Phi\left(\frac{1.98-2.00}{0.008}\right) \\ &= \Phi(2.5) - \Phi(-2.5) \\ &= 0.9938 - (1 - 0.9938) \\ &= 0.9876 \\ &= 98\frac{3}{4}\%. \end{aligned}$$

(b)  $2 \pm 0.0164$  because, for the complementary event, we have

$$0.96 = P(2-c \leq X \leq 2+c)$$

or

$$0.98 = P(X \leq 2+c)$$

so that Table A8 gives

$$\begin{aligned} 0.98 &= \Phi\left(\frac{2+c-2}{0.008}\right), \\ \frac{2+c-2}{0.008} &= 2.054, \quad c = 0.0164. \quad \blacksquare \end{aligned}$$

## Normal Approximation of the Binomial Distribution

The probability function of the binomial distribution is (Sec. 24.7)

$$(8) \quad f(x) = \binom{n}{x} p^x q^{n-x} \quad (x = 0, 1, \dots, n).$$

If  $n$  is large, the binomial coefficients and powers become very inconvenient. It is of great practical (and theoretical) importance that, in this case, the normal distribution provides a good approximation of the binomial distribution, according to the following theorem, one of the most important theorems in all probability theory.

## THEOREM 3

**Limit Theorem of De Moivre and Laplace**

For large  $n$ ,

$$(9) \quad f(x) \sim f^*(x) \quad (x = 0, 1, \dots, n).$$

Here  $f$  is given by (8). The function

$$(10) \quad f^*(x) = \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-z^2/2}, \quad z = \frac{x - np}{\sqrt{npq}}$$

is the density of the normal distribution with mean  $\mu = np$  and variance  $\sigma^2 = npq$  (the mean and variance of the binomial distribution). The symbol  $\sim$  (read **asymptotically equal**) means that the ratio of both sides approaches 1 as  $n$  approaches  $\infty$ . Furthermore, for any nonnegative integers  $a$  and  $b$  ( $b > a$ ),

$$(11) \quad P(a \leq X \leq b) = \sum_{x=a}^b \binom{n}{x} p^x q^{n-x} \sim \Phi(\beta) - \Phi(\alpha),$$

$$\alpha = \frac{a - np - 0.5}{\sqrt{npq}}, \quad \beta = \frac{b - np + 0.5}{\sqrt{npq}}.$$

A proof of this theorem can be found in [G3] listed in App. 1. The proof shows that the term 0.5 in  $\alpha$  and  $\beta$  is a correction caused by the change from a discrete to a continuous distribution.

## PROBLEM SET 24.8

- Let  $X$  be normal with mean 10 and variance 4. Find  $P(X > 12)$ ,  $P(X < 10)$ ,  $P(X < 11)$ ,  $P(9 < X < 13)$ .
- Let  $X$  be normal with mean 105 and variance 25. Find  $P(X \leq 112.5)$ ,  $P(x > 100)$ ,  $P(110.5 < X < 111.25)$ .
- Let  $X$  be normal with mean 50 and variance 9. Determine  $c$  such that  $P(X < c) = 5\%$ ,  $P(X > c) = 1\%$ ,  $P(50 - c < X < 50 + c) = 50\%$ .
- Let  $X$  be normal with mean 3.6 and variance 0.01. Find  $c$  such that  $P(X \leq c) = 50\%$ ,  $P(X > c) = 10\%$ ,  $P(-c < X - 3.6 \leq c) = 99.9\%$ .
- If the lifetime  $X$  of a certain kind of automobile battery is normally distributed with a mean of 5 years and a standard deviation of 1 year, and the manufacturer wishes to guarantee the battery for 4 years, what percentage of the batteries will he have to replace under the guarantee?
- If the standard deviation in Prob. 5 were smaller, would that percentage be larger or smaller?
- A manufacturer knows from experience that the resistance of resistors he produces is normal with mean  $\mu = 150 \Omega$  and standard deviation  $\sigma = 5 \Omega$ . What percentage of the resistors will have resistance between  $148 \Omega$  and  $152 \Omega$ ? Between  $140 \Omega$  and  $160 \Omega$ ?
- The breaking strength  $X$  [kg] of a certain type of plastic block is normally distributed with a mean of 1500 kg and a standard deviation of 50 kg. What is the maximum load such that we can expect no more than 5% of the blocks to break?
- If the mathematics scores of the SAT college entrance exams are normal with mean 480 and standard deviation 100 (these are about the actual values over the past years) and if some college sets 500 as the minimum score for new students, what percent of students would not reach that score?
- A producer sells electric bulbs in cartons of 1000 bulbs. Using (11), find the probability that any given carton contains not more than 1% defective bulbs, assuming the production process to be a Bernoulli experiment with  $p = 1\%$  (= probability that any given bulb will be defective). First guess. Then calculate.



11. If sick-leave time  $X$  used by employees of a company in one month is (very roughly) normal with mean 1000 hours and standard deviation 100 hours, how much time  $t$  should be budgeted for sick leave during the next month if  $t$  is to be exceeded with probability of only 20%?
12. If the monthly machine repair and maintenance cost  $X$  in a certain factory is known to be normal with mean \$12,000 and standard deviation \$2000, what is the probability that the repair cost for the next month will exceed the budgeted amount of \$15,000?
13. If the resistance  $X$  of certain wires in an electrical network is normal with mean  $0.01 \Omega$  and standard deviation  $0.001 \Omega$ , how many of 1000 wires will meet the specification that they have resistance between  $0.009$  and  $0.011 \Omega$ ?
14. **TEAM PROJECT. Normal Distribution.** (a) Derive the formulas in (6) and (7) from the appropriate normal table.  
 (b) Show that  $\Phi(-z) = 1 - \Phi(z)$ . Give an example.  
 (c) Find the points of inflection of the curve of (1).  
 (d) Considering  $\Phi^2(\infty)$  and introducing polar coordinates in the double integral (a standard trick worth remembering), prove

$$(12) \quad \Phi(\infty) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} du = 1.$$

(e) Show that  $\sigma$  in (1) is indeed the standard deviation of the normal distribution. [Use (12).]

(f) **Bernoulli's law of large numbers.** In an experiment let an event  $A$  have probability  $p$  ( $0 < p < 1$ ), and let  $X$  be the number of times  $A$  happens in  $n$  independent trials. Show that for any given  $\epsilon > 0$ ,

$$P\left(\left|\frac{X}{n} - p\right| \leq \epsilon\right) \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

(g) **Transformation.** If  $X$  is normal with mean  $\mu$  and variance  $\sigma^2$ , show that  $X^* = c_1X + c_2$  ( $c_1 > 0$ ) is normal with mean  $\mu^* = c_1\mu + c_2$  and variance  $\sigma^{*2} = c_1^2\sigma^2$ .

15. **WRITING PROJECT. Use of Tables, Use of CAS.** Give a systematic discussion of the use of Tables A7 and A8 for obtaining  $P(X < b)$ ,  $P(X > a)$ ,  $P(a < X < b)$ ,  $P(X < c) = k$ ,  $P(X > c) = k$ , as well as  $P(\mu - c < X < \mu + c) = k$ ; include simple examples. If you have a CAS, describe to what extent it makes the use of those tables superfluous; give examples.

## 24.9 Distributions of Several Random Variables

Distributions of two or more random variables are of interest for two reasons:

1. They occur in experiments in which we observe several random variables, for example, carbon content  $X$  and hardness  $Y$  of steel, amount of fertilizer  $X$  and yield of corn  $Y$ , height  $X_1$ , weight  $X_2$ , and blood pressure  $X_3$  of persons, and so on.
2. They will be needed in the mathematical justification of the methods of statistics in Chap. 25.

In this section we consider two random variables  $X$  and  $Y$  or, as we also say, a **two-dimensional random variable**  $(X, Y)$ . For  $(X, Y)$  the outcome of a trial is a pair of numbers  $X = x$ ,  $Y = y$ , briefly  $(X, Y) = (x, y)$ , which we can plot as a point in the  $XY$ -plane.

The **two-dimensional probability distribution** of the random variable  $(X, Y)$  is given by the **distribution function**

$$(1) \quad F(x, y) = P(X \leq x, Y \leq y).$$

This is the probability that in a trial,  $X$  will assume any value not greater than  $x$  and in the same trial,  $Y$  will assume any value not greater than  $y$ . This corresponds to the blue region in Fig. 522, which extends to  $-\infty$  to the left and below.  $F(x, y)$  determines the

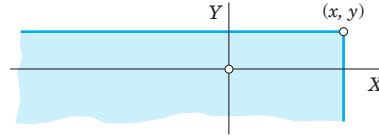


Fig. 522. Formula (1)

probability distribution uniquely, because in analogy to formula (2) in Sec. 24.5, that is,  $P(a < X \leq b) = F(b) - F(a)$ , we now have for a rectangle (see Prob. 16)

$$(2) \quad P(a_1 < X \leq b_1, \quad a_2 < Y \leq b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2).$$

As before, in the two-dimensional case we shall also have discrete and continuous random variables and distributions.

## Discrete Two-Dimensional Distributions

In analogy to the case of a single random variable (Sec. 24.5), we call  $(X, Y)$  and its distribution **discrete** if  $(X, Y)$  can assume only finitely many or at most countably infinitely many pairs of values  $(x_1, y_1), (x_2, y_2), \dots$  with positive probabilities, whereas the probability for any domain containing none of those values of  $(X, Y)$  is zero.

Let  $(x_i, y_j)$  be any of those pairs and let  $P(X = x_i, Y = y_j) = p_{ij}$  (where we admit that  $p_{ij}$  may be 0 for certain pairs of subscripts  $i, j$ ). Then we define the **probability function**  $f(x, y)$  of  $(X, Y)$  by

$$(3) \quad f(x, y) = p_{ij} \quad \text{if} \quad x = x_i, y = y_j \quad \text{and} \quad f(x, y) = 0 \quad \text{otherwise;}$$

here,  $i = 1, 2, \dots$  and  $j = 1, 2, \dots$  independently. In analogy to (4), Sec. 24.5, we now have for the distribution function the formula

$$(4) \quad F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f(x_i, y_j).$$

Instead of (6) in Sec. 24.5 we now have the condition

$$(5) \quad \sum_i \sum_j f(x_i, y_j) = 1.$$

### EXAMPLE 1 Two-Dimensional Discrete Distribution

If we simultaneously toss a dime and a nickel and consider

$X = \text{Number of heads the dime turns up,}$

$Y = \text{Number of heads the nickel turns up,}$

then  $X$  and  $Y$  can have the values 0 or 1, and the probability function is

$$f(0, 0) = f(1, 0) = f(0, 1) = f(1, 1) = \frac{1}{4}, \quad f(x, y) = 0 \quad \text{otherwise.}$$

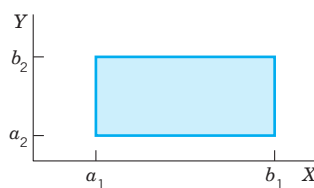


Fig. 523. Notion of a two-dimensional distribution

## Continuous Two-Dimensional Distributions

In analogy to the case of a single random variable (Sec. 24.5) we call  $(X, Y)$  and its distribution **continuous** if the corresponding distribution function  $F(x, y)$  can be given by a double integral

$$(6) \quad F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(x^*, y^*) dx^* dy^*$$

whose integrand  $f$ , called the **density** of  $(X, Y)$ , is nonnegative everywhere, and is continuous, possibly except on finitely many curves.

From (6) we obtain the probability that  $(X, Y)$  assume any value in a rectangle (Fig. 523) given by the formula

$$(7) \quad P(a_1 < X \leq b_1, a_2 < Y \leq b_2) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x, y) dx dy.$$

### EXAMPLE 2 Two-Dimensional Uniform Distribution in a Rectangle

Let  $R$  be the rectangle  $\alpha_1 < x \leq \beta_1, \alpha_2 < y \leq \beta_2$ . The density (see Fig. 524)

$$(8) \quad f(x, y) = 1/k \quad \text{if } (x, y) \text{ is in } R, \quad f(x, y) = 0 \text{ otherwise}$$

defines the so-called **uniform distribution in the rectangle**  $R$ ; here  $k = (\beta_1 - \alpha_1)(\beta_2 - \alpha_2)$  is the area of  $R$ . The distribution function is shown in Fig. 525. ■

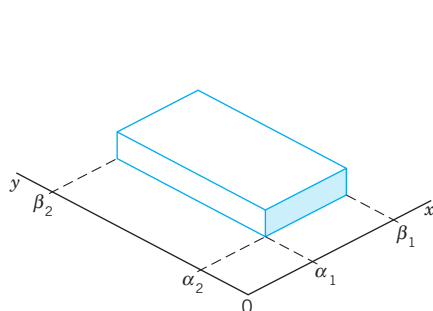


Fig. 524. Density function (8) of the uniform distribution

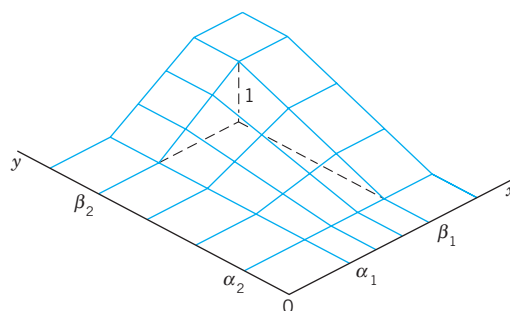


Fig. 525. Distribution function of the uniform distribution defined by (8)

## Marginal Distributions of a Discrete Distribution

This is a rather natural idea, without counterpart for a single random variable. It amounts to being interested only in one of the two variables in  $(X, Y)$ , say,  $X$ , and asking for its distribution, called the **marginal distribution** of  $X$  in  $(X, Y)$ . So we ask for the probability

$P(X = x, Y \text{ arbitrary})$ . Since  $(X, Y)$  is discrete, so is  $X$ . We get its probability function, call it  $f_1(x)$ , from the probability function  $f(x, y)$  of  $(X, Y)$  by summing over  $y$ :

$$(9) \quad f_1(x) = P(X = x, Y \text{ arbitrary}) = \sum_y f(x, y)$$

where we sum all the values of  $f(x, y)$  that are not 0 for that  $x$ .

From (9) we see that the distribution function of the marginal distribution of  $X$  is

$$(10) \quad F_1(x) = P(X \leq x, Y \text{ arbitrary}) = \sum_{x^* \leq x} f_1(x^*).$$

Similarly, the probability function

$$(11) \quad f_2(y) = P(X \text{ arbitrary}, Y \leq y) = \sum_x f(x, y)$$

determines the **marginal distribution** of  $Y$  in  $(X, Y)$ . Here we sum all the values of  $f(x, y)$  that are not zero for the corresponding  $y$ . The distribution function of this marginal distribution is

$$(12) \quad F_2(y) = P(X \text{ arbitrary}, Y \leq y) = \sum_{y^* \leq y} f_2(y^*).$$

### EXAMPLE 3

#### Marginal Distributions of a Discrete Two-Dimensional Random Variable

In drawing 3 cards with replacement from a bridge deck let us consider

$$(X, Y), \quad X = \text{Number of queens}, \quad Y = \text{Number of kings or aces}.$$

The deck has 52 cards. These include 4 queens, 4 kings, and 4 aces. Hence in a single trial a queen has probability  $\frac{4}{52} = \frac{1}{13}$  and a king or ace  $\frac{8}{52} = \frac{2}{13}$ . This gives the probability function of  $(X, Y)$ ,

$$f(x, y) = \frac{3!}{x!y!(3-x-y)!} \left(\frac{1}{13}\right)^x \left(\frac{2}{13}\right)^y \left(\frac{10}{13}\right)^{3-x-y} \quad (x + y \leq 3)$$

and  $f(x, y) = 0$  otherwise. Table 24.1 shows in the center the values of  $f(x, y)$  and on the right and lower margins the values of the probability functions  $f_1(x)$  and  $f_2(y)$  of the marginal distributions of  $X$  and  $Y$ , respectively. ■

**Table 24.1** Values of the Probability Functions  $f(x, y)$ ,  $f_1(x)$ ,  $f_2(y)$  in Drawing Three Cards with Replacement from a Bridge Deck, where  $X$  is the Number of Queens Drawn and  $Y$  is the Number of Kings or Aces Drawn

$x \quad y$	0	1	2	3	$f_1(x)$
0	$\frac{1000}{2197}$	$\frac{600}{2197}$	$\frac{120}{2197}$	$\frac{8}{2197}$	$\frac{1728}{2197}$
1	$\frac{300}{2197}$	$\frac{120}{2197}$	$\frac{12}{2197}$	0	$\frac{432}{2197}$
2	$\frac{30}{2197}$	$\frac{6}{2197}$	0	0	$\frac{36}{2197}$
3	$\frac{1}{2197}$	0	0	0	$\frac{1}{2197}$
$f_2(y)$	$\frac{1331}{2197}$	$\frac{726}{2197}$	$\frac{132}{2197}$	$\frac{8}{2197}$	

## Marginal Distributions of a Continuous Distribution

This is conceptually the same as for discrete distributions, with probability functions and sums replaced by densities and integrals. For a continuous random variable  $(X, Y)$  with density  $f(x, y)$  we now have the **marginal distribution** of  $X$  in  $(X, Y)$ , defined by the distribution function

$$(13) \quad F_1(x) = P(X \leq x, -\infty < Y < \infty) = \int_{-\infty}^x f_1(x^*) dx^*$$

with the density  $f_1$  of  $X$  obtained from  $f(x, y)$  by integration over  $y$ ,

$$(14) \quad f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

Interchanging the roles of  $X$  and  $Y$ , we obtain the **marginal distribution** of  $Y$  in  $(X, Y)$  with the distribution function

$$(15) \quad F_2(y) = P(-\infty < X < \infty, Y \leq y) = \int_{-\infty}^y f_2(y^*) dy^*$$

and density

$$(16) \quad f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

## Independence of Random Variables

$X$  and  $Y$  in a (discrete or continuous) random variable  $(X, Y)$  are said to be **independent** if

$$(17) \quad F(x, y) = F_1(x)F_2(y)$$

holds for all  $(x, y)$ . Otherwise these random variables are said to be **dependent**. These definitions are suggested by the corresponding definitions for events in Sec. 24.3.

Necessary and sufficient for independence is

$$(18) \quad f(x, y) = f_1(x)f_2(y)$$

for all  $x$  and  $y$ . Here the  $f$ 's are the above probability functions if  $(X, Y)$  is discrete or those densities if  $(X, Y)$  is continuous. (See Prob. 20.)

### EXAMPLE 4 Independence and Dependence

In tossing a dime and a nickel,  $X = \text{Number of heads on the dime}$ ,  $Y = \text{Number of heads on the nickel}$  may assume the values 0 or 1 and are independent. The random variables in Table 24.1 are dependent. ■

**Extension of Independence to  $n$ -Dimensional Random Variables.** This will be needed throughout Chap. 25. The distribution of such a random variable  $\mathbf{X} = (X_1, \dots, X_n)$  is determined by a **distribution function** of the form

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

The random variables  $X_1, \dots, X_n$  are said to be **independent** if

$$(19) \quad F(x_1, \dots, x_n) = F_1(x_1)F_2(x_2) \cdots F_n(x_n)$$

for all  $(x_1, \dots, x_n)$ . Here  $F_j(x_j)$  is the distribution function of the marginal distribution of  $X_j$  in  $\mathbf{X}$ , that is,

$$F_j(x_j) = P(X_j \leq x_j, X_k \text{ arbitrary}, k = 1, \dots, n, k \neq j).$$

Otherwise these random variables are said to be **dependent**.

## Functions of Random Variables

When  $n = 2$ , we write  $X_1 = X, X_2 = Y, x_1 = x, x_2 = y$ . Taking a nonconstant continuous function  $g(x, y)$  defined for all  $x, y$ , we obtain a random variable  $Z = g(X, Y)$ . For example, if we roll two dice and  $X$  and  $Y$  are the numbers the dice turn up in a trial, then  $Z = X + Y$  is the sum of those two numbers (see Fig. 514 in Sec. 24.5).

In the case of a **discrete** random variable  $(X, Y)$  we may obtain the probability function  $f(z)$  of  $Z = g(X, Y)$  by summing all  $f(x, y)$  for which  $g(x, y)$  equals the value of  $z$  considered; thus

$$(20) \quad f(z) = P(Z = z) = \sum_{g(x,y)=z} \sum f(x, y).$$

Hence the distribution function of  $Z$  is

$$(21) \quad F(z) = P(Z \leq z) = \sum_{g(x,y) \leq z} \sum f(x, y)$$

where we sum all values of  $f(x, y)$  for which  $g(x, y) \leq z$ .

In the case of a **continuous** random variable  $(X, Y)$  we similarly have

$$(22) \quad F(z) = P(Z \leq z) = \iint_{g(x,y) \leq z} f(x, y) dx dy$$

where for each  $z$  we integrate the density  $f(x, y)$  of  $(X, Y)$  over the region  $g(x, y) \leq z$  in the  $xy$ -plane, the boundary curve of this region being  $g(x, y) = z$ .

## Addition of Means

The number

$$(23) \quad E(g(X, Y)) = \begin{cases} \sum_x \sum_y g(x, y)f(x, y) & [(X, Y) \text{ discrete}] \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y) dx dy & [(X, Y) \text{ continuous}] \end{cases}$$

is called the *mathematical expectation* or, briefly, the **expectation** of  $g(X, Y)$ . Here it is assumed that the double series converges absolutely and the integral of  $|g(x, y)|f(x, y)$  over the  $xy$ -plane exists (is finite). Since summation and integration are linear processes, we have from (23)

$$(24) \quad E(ag(X, Y) + bh(X, Y)) = aE(g(X, Y)) + bE(h(X, Y)).$$

An important special case is

$$E(X + Y) = E(X) + E(Y),$$

and by induction we have the following result.

### THEOREM 1

#### Addition of Means

The mean (expectation) of a sum of random variables equals the sum of the means (expectations), that is,

$$(25) \quad E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n).$$

Furthermore, we readily obtain

### THEOREM 2

#### Multiplication of Means

The mean (expectation) of the product of **independent** random variables equals the product of the means (expectations), that is,

$$(26) \quad E(X_1 X_2 \cdots X_n) = E(X_1)E(X_2) \cdots E(X_n).$$

**PROOF** If  $X$  and  $Y$  are independent random variables (both discrete or both continuous), then  $E(XY) = E(X)E(Y)$ . In fact, in the discrete case we have

$$E(XY) = \sum_x \sum_y xyf(x, y) = \sum_x xf_1(x) \sum_y yf_2(y) = E(X)E(Y),$$

and in the continuous case the proof of the relation is similar. Extension to  $n$  independent random variables gives (26), and Theorem 2 is proved. ■

## Addition of Variances

This is another matter of practical importance that we shall need. As before, let  $Z = X + Y$  and denote the mean and variance of  $Z$  by  $\mu$  and  $\sigma^2$ . Then we first have (see Team Project 20(a) in Problem Set 24.6)

$$\sigma^2 = E([Z - \mu]^2) = E(Z^2) - [E(Z)]^2.$$

From (24) we see that the first term on the right equals

$$E(Z^2) = E(X^2 + 2XY + Y^2) = E(X^2) + 2E(XY) + E(Y^2).$$

For the second term on the right we obtain from Theorem 1

$$[E(Z)]^2 = [E(X) + E(Y)]^2 = [E(X)]^2 + 2E(X)E(Y) + [E(Y)]^2.$$

By substituting these expressions into the formula for  $\sigma^2$  we have

$$\begin{aligned} \sigma^2 &= E(X^2) - [E(X)]^2 + E(Y^2) - [E(Y)]^2 \\ &\quad + 2[E(XY) - E(X)E(Y)]. \end{aligned}$$

From Team Project 20, Sec. 24.6, we see that the expression in the first line on the right is the sum of the variances of  $X$  and  $Y$ , which we denote by  $\sigma_1^2$  and  $\sigma_2^2$ , respectively. The quantity in the second line (except for the factor 2) is

$$(27) \quad \sigma_{XY} = E(XY) - E(X)E(Y)$$

and is called the **covariance** of  $X$  and  $Y$ . Consequently, our result is

$$(28) \quad \sigma^2 = \sigma_1^2 + \sigma_2^2 + 2\sigma_{XY}.$$

If  $X$  and  $Y$  are independent, then

$$E(XY) = E(X)E(Y);$$

hence  $\sigma_{XY} = 0$ , and

$$(29) \quad \sigma^2 = \sigma_1^2 + \sigma_2^2.$$

Extension to more than two variables gives the basic

### THEOREM 3

#### Addition of Variances

*The variance of the sum of **independent** random variables equals the sum of the variances of these variables.*



**CAUTION!** In the numerous applications of Theorems 1 and 3 we must always remember that Theorem 3 holds only for *independent* variables.

This is the end of Chap. 24 on probability theory. Most of the concepts, methods, and special distributions discussed in this chapter will play a fundamental role in the next chapter, which deals with methods of **statistical inference**, that is, conclusions from samples to populations, whose unknown properties we want to know and try to discover by looking at suitable properties of samples that we have obtained.

## PROBLEM SET 24.9

- Let  $f(x, y) = k$  when  $8 \leq x \leq 12$  and  $0 \leq y \leq 2$  and zero elsewhere. Find  $k$ . Find  $P(X \leq 11, 1 \leq Y \leq 1.5)$  and  $P(9 \leq X \leq 13, Y \leq 1)$ .
- Find  $P(X > 4, Y > 4)$  and  $P(X \leq 1, Y \leq 1)$  if  $(X, Y)$  has the density  $f(x, y) = \frac{1}{32}$  if  $x \geq 0, y \geq 0, x + y \leq 8$ .
- Let  $f(x, y) = k$  if  $x > 0, y > 0, x + y < 3$  and 0 otherwise. Find  $k$ . Sketch  $f(x, y)$ . Find  $P(X + Y \leq 1), P(Y > X)$ .
- Find the density of the marginal distribution of  $X$  in Prob. 2.
- Find the density of the marginal distribution of  $Y$  in Fig. 524.
- If certain sheets of wrapping paper have a mean weight of 10 g each, with a standard deviation of 0.05 g, what are the mean weight and standard deviation of a pack of 10,000 sheets?
- What are the mean thickness and the standard deviation of transformer cores each consisting of 50 layers of sheet metal and 49 insulating paper layers if the metal sheets have mean thickness 0.5 mm each with a standard deviation of 0.05 mm and the paper layers have mean 0.05 mm each with a standard deviation of 0.02 mm?
- Let  $X$  [cm] and  $Y$  [cm] be the diameters of a pin and hole, respectively. Suppose that  $(X, Y)$  has the density  $f(x, y) = 625$  if  $0.98 < x < 1.02, 1.00 < y < 1.04$  and 0 otherwise. (a) Find the marginal distributions. (b) What is the probability that a pin chosen at random will fit a hole whose diameter is 1.00?
- Using Theorems 1 and 3, obtain the formulas for the mean and the variance of the binomial distribution.
- Using Theorem 1, obtain the formula for the mean of the hypergeometric distribution. Can you use Theorem 3 to obtain the variance of that distribution?
- A 5-gear assembly is put together with spacers between the gears. The mean thickness of the gears is 5.020 cm with a standard deviation of 0.003 cm. The mean thickness of the spacers is 0.040 cm with a standard deviation of 0.002 cm. Find the mean and standard deviation of the assembled units consisting of 5 randomly selected gears and 4 randomly selected spacers.
- If the mean weight of certain (empty) containers is 5 lb the standard deviation is 0.2 lb, and if the filling of the containers has mean weight 100 lb and standard deviation 0.5 lb, what are the mean weight and the standard deviation of filled containers?
- Find  $P(X > Y)$  when  $(X, Y)$  has the density  $f(x, y) = 0.25e^{-0.5(x+y)}$  if  $x \geq 0, y \geq 0$  and 0 otherwise.
- An electronic device consists of two components. Let  $X$  and  $Y$  [years] be the times to failure of the first and second components, respectively. Assume that  $(X, Y)$  has the density  $f(x, y) = 4e^{-2(x+y)}$  if  $x > 0$  and  $y > 0$  and 0 otherwise. (a) Are  $X$  and  $Y$  dependent or independent? (b) Find the densities of the marginal distributions. (c) What is the probability that the first component will have a lifetime of 2 years or longer?
- Give an example of two different discrete distributions that have the same marginal distributions.
- Prove (2).
- Let  $(X, Y)$  have the probability function  $f(0, 0) = f(1, 1) = \frac{1}{8},$   
 $f(0, 1) = f(1, 0) = \frac{3}{8}.$   
Are  $X$  and  $Y$  independent?
- Let  $(X, Y)$  have the density  $f(x, y) = k$  if  $x^2 + y^2 < 1$  and 0 otherwise. Determine  $k$ . Find the densities of the marginal distributions. Find the probability  $P(X^2 + Y^2 < \frac{1}{4})$ .
- Show that the random variables with the densities  $f(x, y) = x + y$  and  $g(x, y) = (x + \frac{1}{2})(y + \frac{1}{2})$  if  $0 \leq x \leq 1, 0 \leq y \leq 1$  and  $f(x, y) = 0$  and  $g(x, y) = 0$  elsewhere, have the same marginal distribution.
- Prove the statement involving (18).

## CHAPTER 24 REVIEW QUESTIONS AND PROBLEMS

1. What are stem-and-leaf plots? Boxplots? Histograms? Compare their advantages.
2. What properties of data are measured by the mean? The median? The standard deviation? The variance?
3. What do we mean by an experiment? An outcome? An event? Give examples.
4. What is a random variable? Its distribution function? Its probability function or density?
5. State the definition of probability from memory. Give simple examples.
6. What is sampling with and without replacement? What distributions are involved?
7. When is the Poisson distribution a good approximation of the binomial distribution? The normal distribution?
8. Explain the use of the tables of the normal distribution. If you have a CAS, how would you proceed without the tables?
9. State the main theorems on probability. Illustrate them by simple examples.
10. State the most important facts about distributions of two random variables and their marginal distributions.
11. Make a stem-and-leaf plot, histogram, and boxplot of the data 110, 113, 109, 118, 110, 115, 104, 111, 116, 113.
12. Same task as in Prob. 11. for the data 13.5, 13.2, 12.1, 13.6, 13.3.
13. Find the mean, standard deviation, and variance in Prob. 11.
14. Find the mean, standard deviation, and variance in Prob. 12.
15. Show that the mean always lies between the smallest and the largest data value.
16. What are the outcomes in the sample space of the experiment of simultaneously tossing three coins?
17. Plot a histogram of the data 8, 2, 4, 10 and guess  $\bar{x}$  and  $s$  by inspecting the histogram. Then calculate  $\bar{x}$ ,  $s^2$ , and  $s$ .
18. Using a Venn diagram, show that  $A \subseteq B$  if and only if  $A \cap B = A$ .
19. Suppose that 3% of bolts made by a machine are defective, the defectives occurring at random during production. If the bolts are packaged 50 per box, what is the binomial approximation of the probability that a given box will contain  $x = 0, 1, \dots, 5$  defectives?
20. Of a lot of 12 items, 3 are defective. (a) Find the number of different samples of 3 items. Find the number of samples of 3 items containing (b) no defectives, (c) 1 defective, (d) 2 defectives, (e) 3 defectives.
21. Find the probability function of  $X = \text{Number of times of tossing a fair coin until the first head appears}$ .
22. If the life of ball bearings has the density  $f(x) = ke^{-x}$  if  $0 \leq x \leq 2$  and 0 otherwise, what is  $k$ ? What is the probability  $P(X \geq 1)$ ?
23. Find the mean and variance of a discrete random variable  $X$  having the probability function  $f(0) = \frac{1}{4}$ ,  $f(1) = \frac{1}{2}$ ,  $f(2) = \frac{1}{4}$ .
24. Let  $X$  be normal with mean 14 and variance 4. Determine  $c$  such that  $P(X \leq c) = 95\%$ ,  $P(X \leq c) = 5\%$ ,  $P(X \leq c) = 99.5\%$ .
25. Let  $X$  be normal with mean 80 and variance 9. Find  $P(X > 83)$ ,  $P(X < 81)$ ,  $P(X < 80)$ , and  $P(78 < X < 82)$ .

## SUMMARY OF CHAPTER 24

### Data Analysis. Probability Theory

A *random experiment*, briefly called **experiment**, is a process in which the result (“**outcome**”) depends on “chance” (effects of factors unknown to us). Examples are games of chance with dice or cards, measuring the hardness of steel, observing weather conditions, or recording the number of accidents in a city. (Thus the word “experiment” is used here in a much wider sense than in common language.) The outcomes are regarded as points (elements) of a set  $S$ , called the **sample space**, whose subsets are called **events**. For events  $E$  we define a **probability**  $P(E)$  by the axioms (Sec. 24.3)

$$\begin{aligned}
 (1) \quad & 0 \leq P(E) \leq 1 \\
 & P(S) = 1 \\
 & P(E_1 \cup E_2 \cup \cdots) = P(E_1) + P(E_2) + \cdots \quad (E_j \cap E_k = \emptyset).
 \end{aligned}$$

These axioms are motivated by properties of frequency distributions of data (Sec. 24.1).

The complement  $E^c$  of  $E$  has the probability

$$(2) \quad P(E^c) = 1 - P(E).$$

The **conditional probability** of an event  $B$  under the condition that an event  $A$  happens is (Sec. 24.3)

$$(3) \quad P(B|A) = \frac{P(A \cap B)}{P(A)} \quad [P(A) > 0].$$

Two events  $A$  and  $B$  are called **independent** if the probability of their simultaneous appearance in a trial equals the product of their probabilities, that is, if

$$(4) \quad P(A \cap B) = P(A)P(B).$$

With an experiment we associate a **random variable**  $X$ . This is a function defined on  $S$  whose values are real numbers; furthermore,  $X$  is such that the probability  $P(X = a)$  with which  $X$  assumes any value  $a$ , and the probability  $P(a < X \leq b)$  with which  $X$  assumes any value in an interval  $a < X \leq b$  are defined (Sec. 24.5). The **probability distribution** of  $X$  is determined by the distribution function

$$(5) \quad F(x) = P(X \leq x).$$

In applications there are two important kinds of random variables: those of the **discrete** type, which appear if we count (defective items, customers in a bank, etc.) and those of the **continuous** type, which appear if we measure (length, speed, temperature, weight, etc.).

A discrete random variable has a **probability function**

$$(6) \quad f(x) = P(X = x).$$

Its **mean**  $\mu$  and **variance**  $\sigma^2$  are (Sec. 24.6)

$$(7) \quad \mu = \sum_j x_j f(x_j) \quad \text{and} \quad \sigma^2 = \sum_j (x_j - \mu)^2 f(x_j)$$

where the  $x_j$  are the values for which  $X$  has a positive probability. Important discrete random variables and distributions are the **binomial**, **Poisson**, and **hypergeometric distributions** discussed in Sec. 24.7.

A continuous random variable has a **density**

$$(8) \quad f(x) = F'(x) \quad [\text{see (5)}].$$

Its mean and variance are (Sec. 24.6)

$$(9) \quad \mu = \int_{-\infty}^{\infty} x f(x) dx \quad \text{and} \quad \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

Very important is the **normal distribution** (Sec. 24.8), whose density is

$$(10) \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right]$$

and whose distribution function is (Sec. 24.8; Tables A7, A8 in App. 5)

$$(11) \quad F(x) = \Phi \left( \frac{x - \mu}{\sigma} \right).$$

A **two-dimensional random variable**  $(X, Y)$  occurs if we simultaneously observe two quantities (for example, height  $X$  and weight  $Y$  of adults). Its distribution function is (Sec. 24.9)

$$(12) \quad F(x, y) = P(X \leq x, Y \leq y).$$

$X$  and  $Y$  have the distribution functions (Sec. 24.9)

$$(13) \quad F_1(x) = P(X \leq x, Y \text{ arbitrary}) \quad \text{and} \quad F_2(y) = P(x \text{ arbitrary}, Y \leq y)$$

respectively; their distributions are called **marginal distributions**. If both  $X$  and  $Y$  are discrete, then  $(X, Y)$  has a probability function

$$f(x, y) = P(X = x, Y = y).$$

If both  $X$  and  $Y$  are continuous, then  $(X, Y)$  has a density  $f(x, y)$ .



## CHAPTER 25

# Mathematical Statistics

In probability theory we set up mathematical models of processes that are affected by “chance.” In mathematical statistics or, briefly, **statistics**, we check these models against the observable reality. This is called **statistical inference**. It is done by **sampling**, that is, by drawing random samples, briefly called **samples**. These are sets of values from a much larger set of values that could be studied, called the **population**. An example is 10 diameters of screws drawn from a large lot of screws. Sampling is done in order to see whether a model of the population is accurate enough for practical purposes. If this is the case, the model can be used for predictions, decisions, and actions, for instance, in planning productions, buying equipment, investing in business projects, and so on.

Most important methods of statistical inference are **estimation of parameters** (Secs. 25.2), determination of **confidence intervals** (Sec. 25.3), and **hypothesis testing** (Sec. 25.4, 25.7, 25.8), with application to *quality control* (Sec. 25.5) and *acceptance sampling* (Sec. 25.6).

In the last section (25.9) we give an introduction to **regression** and **correlation analysis**, which concern experiments involving two variables.

*Prerequisite:* Chap. 24.

*Sections that may be omitted in a shorter course:* 25.5, 25.6, 25.8.

*References, Answers to Problems, and Statistical Tables:* App. 1 Part G, App. 2, App. 5.

## 25.1 Introduction. Random Sampling

**Mathematical statistics** consists of methods for designing and evaluating random experiments to obtain information about practical problems, such as exploring the relation between iron content and density of iron ore, the quality of raw material or manufactured products, the efficiency of air-conditioning systems, the performance of certain cars, the effect of advertising, the reactions of consumers to a new product, etc.

**Random variables** occur more frequently in engineering (and elsewhere) than one would think. For example, properties of mass-produced articles (screws, lightbulbs, etc.) always show **random variation**, due to small (uncontrollable!) differences in raw material or manufacturing processes. Thus the diameter of screws is a random variable  $X$  and we have *nondefective screws*, with diameter between given tolerance limits, and *defective screws*, with diameter outside those limits. We can ask for the distribution of  $X$ , for the percentage of defective screws to be expected, and for necessary improvements of the production process.

**Samples** are selected from populations—20 screws from a lot of 1000, 100 of 5000 voters, 8 beavers in a wildlife conservation project—because inspecting the entire population would be too expensive, time-consuming, impossible or even senseless (think

of destructive testing of lightbulbs or dynamite). To obtain meaningful conclusions, samples must be **random selections**. Each of the 1000 screws must have the same chance of being sampled (of being drawn when we sample), at least approximately. Only then will the sample mean  $\bar{x} = (x_1 + \cdots + x_{20})/20$  (Sec. 24.1) of a sample of size  $n = 20$  (or any other  $n$ ) be a good approximation of the population mean  $\mu$  (Sec. 24.6); and the accuracy of the approximation will generally improve with increasing  $n$ , as we shall see. Similarly for other parameters (standard deviation, variance, etc.).

**Independent sample values** will be obtained in experiments with an infinite sample space  $S$  (Sec. 24.2), certainly for the normal distribution. This is also true in sampling with replacement. It is approximately true in drawing *small* samples from a large finite population (for instance, 5 or 10 of 1000 items). However, if we sample without replacement from a small population, the effect of dependence of sample values may be considerable.

**Random numbers** help in obtaining samples that are in fact random selections. This is sometimes not easy to accomplish because there are many subtle factors that can bias sampling (by personal interviews, by poorly working machines, by the choice of nontypical observation conditions, etc.). Random numbers can be obtained from a **random number generator** in Maple, Mathematica, or other systems listed on p. 789. (The numbers are not truly random, as they would be produced in flipping coins or rolling dice, but are calculated by a tricky formula that produces numbers that do have practically all the essential features of true randomness. Because these numbers eventually repeat, they must not be used in cryptography, for example, where true randomness is required.)

### EXAMPLE 1

#### Random Numbers from a Random Number Generator

To select a sample of size  $n = 10$  from 80 given ball bearings, we number the bearings from 1 to 80. We then let the generator randomly produce 10 of the integers from 1 to 80 and include the bearings with the numbers obtained in our sample, for example.

44 55 53 03 52 61 67 78 39 54

or whatever.

Random numbers are also contained in (older) statistical tables. ■

**Representing and processing data** were considered in Sec. 24.1 in connection with frequency distributions. These are the empirical counterparts of probability distributions and helped motivating axioms and properties in probability theory. The new aspect in this chapter is **randomness**: the data are samples selected **randomly** from a population. Accordingly, we can immediately make the connection to Sec. 24.1, using stem-and-leaf plots, box plots, and histograms for representing samples graphically.

Also, we now call the mean  $\bar{x}$  in (5), Sec. 24.1, the **sample mean**

$$(1) \quad \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j = \frac{1}{n} (x_1 + x_2 + \cdots + x_n).$$

We call  $n$  the **sample size**, the variance  $s^2$  in (6), Sec. 24.1, the **sample variance**

$$(2) \quad s^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2 = \frac{1}{n-1} [(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2],$$

and its positive square root  $s$  the **sample standard deviation**.  $\bar{x}$ ,  $s^2$ , and  $s$  are called **parameters of a sample**; they will be needed throughout this chapter.

## 25.2 Point Estimation of Parameters

Beginning in this section, we shall discuss the most basic practical tasks in statistics and corresponding statistical methods to accomplish them. The first of them is point estimation of **parameters**, that is, of quantities appearing in distributions, such as  $p$  in the binomial distribution and  $\mu$  and  $\sigma$  in the normal distribution.

A **point estimate** of a parameter is a number (point on the real line), which is computed from a given sample and serves as an approximation of the unknown exact value of the parameter of the population. An **interval estimate** is an interval (“*confidence interval*”) obtained from a sample; such estimates will be considered in the next section. Estimation of parameters is of great practical importance in many applications.

As an approximation of the mean  $\mu$  of a population we may take the mean  $\bar{x}$  of a corresponding sample. This gives the estimate  $\hat{\mu} = \bar{x}$  for  $\mu$ , that is,

$$(1) \quad \hat{\mu} = \bar{x} = \frac{1}{n}(x_1 + \cdots + x_n)$$

where  $n$  is the sample size. Similarly, an estimate  $\hat{\sigma}^2$  for the variance of a population is the variance  $s^2$  of a corresponding sample, that is,

$$(2) \quad \hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2.$$

Clearly, (1) and (2) are estimates of parameters for distributions in which  $\mu$  or  $\sigma^2$  appear explicitly as parameters, such as the normal and Poisson distributions. For the binomial distribution,  $p = \mu/n$  [see (3) in Sec. 24.7]. From (1) we thus obtain for  $p$  the estimate

$$(3) \quad \hat{p} = \frac{\bar{x}}{n}.$$

We mention that (1) is a special case of the so-called **method of moments**. In this method the parameters to be estimated are expressed in terms of the moments of the distribution (see Sec. 24.6). In the resulting formulas, those moments of the distribution are replaced by the corresponding moments of the sample. This gives the estimates. Here the  **$k$ th moment of a sample**  $x_1, \dots, x_n$  is

$$m_k = \frac{1}{n} \sum_{j=1}^n x_j^k.$$

## Maximum Likelihood Method

Another method for obtaining estimates is the so-called **maximum likelihood method** of R. A. Fisher [*Messenger Math.* **41** (1912), 155–160]. To explain it, we consider a discrete (or continuous) random variable  $X$  whose probability function (or density)  $f(x)$  depends on a single parameter  $\theta$ . We take a corresponding sample of  $n$  **independent** values  $x_1, \dots, x_n$ . Then in the discrete case the probability that a sample of size  $n$  consists precisely of those  $n$  values is

$$(4) \quad l = f(x_1)f(x_2) \cdots f(x_n).$$

In the continuous case the probability that the sample consists of values in the small intervals  $x_j \leq x \leq x_j + \Delta x$  ( $j = 1, 2, \dots, n$ ) is

$$(5) \quad f(x_1)\Delta x f(x_2)\Delta x \cdots f(x_n)\Delta x = l(\Delta x)^n.$$

Since  $f(x_j)$  depends on  $\theta$ , the function  $l$  in (5) given by (4) depends on  $x_1, \dots, x_n$  and  $\theta$ . We imagine  $x_1, \dots, x_n$  to be given and fixed. Then  $l$  is a function of  $\theta$ , which is called the **likelihood function**. The basic idea of the maximum likelihood method is quite simple, as follows. We choose *that* approximation for the unknown value of  $\theta$  for which  $l$  is as large as possible. If  $l$  is a differentiable function of  $\theta$ , a necessary condition for  $l$  to have a maximum in an interval (not at the boundary) is

$$(6) \quad \frac{\partial l}{\partial \theta} = 0.$$

(We write a *partial* derivative, because  $l$  depends also on  $x_1, \dots, x_n$ .) A solution of (6) depending on  $x_1, \dots, x_n$  is called a **maximum likelihood estimate** for  $\theta$ . We may replace (6) by

$$(7) \quad \frac{\partial \ln l}{\partial \theta} = 0,$$

because  $f(x_j) > 0$ , a maximum of  $l$  is in general positive, and  $\ln l$  is a monotone increasing function of  $l$ . This often simplifies calculations.

**Several Parameters.** If the distribution of  $X$  involves  $r$  parameters  $\theta_1, \dots, \theta_r$ , then instead of (6) we have the  $r$  conditions  $\partial l / \partial \theta_1 = 0, \dots, \partial l / \partial \theta_r = 0$ , and instead of (7) we have

$$(8) \quad \frac{\partial \ln l}{\partial \theta_1} = 0, \quad \dots, \quad \frac{\partial \ln l}{\partial \theta_r} = 0.$$

### EXAMPLE 1 Normal Distribution

Find maximum likelihood estimates for  $\theta_1 = \mu$  and  $\theta_2 = \sigma$  in the case of the normal distribution.

**Solution.** From (1), Sec. 24.8, and (4) we obtain the likelihood function

$$l = \left( \frac{1}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\sigma} \right)^n e^{-h} \quad \text{where} \quad h = \frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2.$$



Taking logarithms, we have

$$\ln l = -n \ln \sqrt{2\pi} - n \ln \sigma - h.$$

The first equation in (8) is  $\partial(\ln l)/\partial\mu = 0$ , written out

$$\frac{\partial \ln l}{\partial \mu} = -\frac{\partial h}{\partial \mu} = \frac{1}{\sigma^2} \sum_{j=1}^n (x_j - \mu) = 0. \quad \text{hence} \quad \sum_{j=1}^n x_j - n\mu = 0.$$

The solution is the desired estimate  $\hat{\mu}$  for  $\mu$ : we find

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^n x_j = \bar{x}.$$

The second equation in (8) is  $\partial(\ln l)/\partial\sigma = 0$ , written out

$$\frac{\partial \ln l}{\partial \sigma} = -\frac{n}{\sigma} - \frac{\partial h}{\partial \sigma} = -\frac{1}{\sigma} + \frac{1}{\sigma^3} \sum_{j=1}^n (x_j - \mu)^2 = 0.$$

Replacing  $\mu$  by  $\hat{\mu}$  and solving for  $\sigma^2$ , we obtain the estimate

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2$$

which we shall use in Sec. 25.7. Note that this differs from (2). We cannot discuss criteria for the goodness of estimates but want to mention that for small  $n$ , formula (2) is preferable. ■

## PROBLEM SET 25.2

- Normal distribution.** Apply the maximum likelihood method to the normal distribution with  $\mu = 0$ .
- Find the maximum likelihood estimate for the parameter  $\mu$  of a normal distribution with known variance  $\sigma^2 = \sigma_0^2 = 16$ .
- Poisson distribution.** Derive the maximum likelihood estimator for  $\mu$ . Apply it to the sample (10, 25, 26, 17, 10, 4), giving numbers of minutes with 0–10, 11–20, 21–30, 31–40, 41–50, more than 50 fliers per minute, respectively, checking in at some airport check-in.
- Uniform distribution.** Show that, in the case of the parameters  $a$  and  $b$  of the uniform distribution (see Sec. 24.6), the maximum likelihood estimate cannot be obtained by equating the first derivative to zero. How can we obtain maximum likelihood estimates in this case, more or less by using common sense?
- Binomial distribution.** Derive a maximum likelihood estimate for  $p$ .
- Extend Prob. 5 as follows. Suppose that  $m$  times  $n$  trials were made and in the first  $n$  trials  $A$  happened  $k_1$  times, in the second  $n$  trials  $A$  happened  $k_2$  times,  $\dots$ , in the  $m$ th  $n$  trials  $A$  happened  $k_m$  times. Find a maximum likelihood estimate of  $p$  based on this information.
- Suppose that in Prob. 6 we made 3 times 4 trials and  $A$  happened 2, 3, 2 times, respectively. Estimate  $p$ .
- Geometric distribution.** Let  $X = \text{Number of independent trials until an event } A \text{ occurs}$ . Show that  $X$  has a geometric distribution, defined by the probability function  $f(x) = pq^{x-1}$ ,  $x = 1, 2, \dots$ , where  $p$  is the probability of  $A$  in a single trial and  $q = 1 - p$ . Find the maximum likelihood estimate of  $p$  corresponding to a sample  $x_1, x_2, \dots, x_n$  of observed values of  $X$ .
- In Prob. 8, show that  $f(1) + f(2) + \dots = 1$  (as it should be!). Calculate independently of Prob. 8 the maximum likelihood of  $p$  in Prob. 8 corresponding to a single observed value of  $X$ .
- In rolling a die, suppose that we get the first “Six” in the 7th trial and in doing it again we get it in the 6th trial. Estimate the probability  $p$  of getting a “Six” in rolling that die once.
- Find the maximum likelihood estimate of  $\theta$  in the density  $f(x) = \theta e^{-\theta x}$  if  $x \geq 0$  and  $f(x) = 0$  if  $x < 0$ .
- In Prob. 11, find the mean  $\mu$ , substitute it in  $f(x)$ , find the maximum likelihood estimate of  $\mu$ , and show that it is identical with the estimate for  $\mu$  which can be obtained from that for  $\theta$  in Prob. 11.

13. Compute  $\hat{\theta}$  in Prob. 11 from the sample 1.9, 0.4, 0.7, 0.6, 1.4. Graph the sample distribution function  $\hat{F}(x)$  and the distribution function  $F(x)$  of the random variable, with  $\theta = \hat{\theta}$ , on the same axes. Do they agree reasonably well? (We consider goodness of fit systematically in Sec. 25.7.)
14. Do the same task as in Prob. 13 if the given sample is 0.4, 0.7, 0.2, 1.1, 0.1.
15. **CAS EXPERIMENT. Maximum Likelihood Estimates. (MLEs).** Find experimentally how much MLEs can differ depending on the sample size. *Hint.* Generate many samples of the same size  $n$ , e.g., of the standardized normal distribution, and record  $\bar{x}$  and  $s^2$ . Then increase  $n$ .

## 25.3 Confidence Intervals

**Confidence intervals**<sup>1</sup> for an unknown parameter  $\theta$  of some distribution (e.g.,  $\theta = \mu$ ) are intervals  $\theta_1 \leq \theta \leq \theta_2$  that contain  $\theta$ , not with certainty but with a high probability  $\gamma$ , which we can choose (95% and 99% are popular). Such an interval is calculated from a sample.  $\gamma = 95\%$  means probability  $1 - \gamma = 5\% = \frac{1}{20}$  of being wrong—one of about 20 such intervals will not contain  $\theta$ . Instead of writing  $\theta_1 \leq \theta \leq \theta_2$ , we denote this more distinctly by writing

$$(1) \quad \text{CONF}_\gamma \{ \theta_1 \leq \theta \leq \theta_2 \}.$$

Such a special symbol, CONF, seems worthwhile in order to avoid the misunderstanding that  $\theta$  *must* lie between  $\theta_1$  and  $\theta_2$ .

$\gamma$  is called the **confidence level**, and  $\theta_1$  and  $\theta_2$  are called the **lower** and **upper confidence limits**. They depend on  $\gamma$ . The larger we choose  $\gamma$ , the smaller is the error probability  $1 - \gamma$ , but the longer is the confidence interval. If  $\gamma \rightarrow 1$ , then its length goes to infinity. The choice of  $\gamma$  depends on the kind of application. In taking no umbrella, a 5% chance of getting wet is not tragic. In a medical decision of life or death, a 5% chance of being wrong may be too large and a 1% chance of being wrong ( $\gamma = 99\%$ ) may be more desirable.

Confidence intervals are more valuable than point estimates (Sec. 25.2). Indeed, we can take the midpoint of (1) as an approximation of  $\theta$  and half the length of (1) as an “error bound” (not in the strict sense of numerics, but except for an error whose probability we know).

$\theta_1$  and  $\theta_2$  in (1) are calculated from a sample  $x_1, \dots, x_n$ . These are  $n$  observations of a random variable  $X$ . Now comes a **standard trick**. We regard  $x_1, \dots, x_n$  as *single observations of  $n$  random variables  $X_1, \dots, X_n$  (with the same distribution, namely, that of  $X$ )*. Then  $\theta_1 = \theta_1(x_1, \dots, x_n)$  and  $\theta_2 = \theta_2(x_1, \dots, x_n)$  in (1) are observed values of two random variables  $\Theta_1 = \Theta_1(X_1, \dots, X_n)$  and  $\Theta_2 = \Theta_2(X_1, \dots, X_n)$ . The condition (1) involving  $\gamma$  can now be written

$$(2) \quad P(\Theta_1 \leq \theta \leq \Theta_2) = \gamma.$$

Let us see what all this means in concrete practical cases.

In each case in this section we shall first state the steps of obtaining a confidence interval in the form of a table, then consider a typical example, and finally justify those steps theoretically.

<sup>1</sup>JERZY NEYMAN (1894–1981), American statistician, developed the theory of confidence intervals (*Annals of Mathematical Statistics* 6 (1935), 111–116).

## Confidence Interval for $\mu$ of the Normal Distribution with Known $\sigma^2$

**Table 25.1** Determination of a Confidence Interval for the Mean  $\mu$  of a Normal Distribution with Known Variance  $\sigma^2$

**Step 1.** Choose a confidence level  $\gamma$  (95%, 99%, or the like).

**Step 2.** Determine the corresponding  $c$ :

$\gamma$	0.90	0.95	0.99	0.999
$c$	1.645	1.960	2.576	3.291

**Step 3.** Compute the mean  $\bar{x}$  of the sample  $x_1, \dots, x_n$ .

**Step 4.** Compute  $k = c\sigma/\sqrt{n}$ . The confidence interval for  $\mu$  is

$$(3) \quad \text{CONF}_{\gamma} \{ \bar{x} - k \leq \mu \leq \bar{x} + k \}.$$

### EXAMPLE 1 Confidence Interval for $\mu$ of the Normal Distribution with Known $\sigma^2$

Determine a 95% confidence interval for the mean of a normal distribution with variance  $\sigma^2 = 9$ , using a sample of  $n = 100$  values with mean  $\bar{x} = 5$ .

**Solution.** **Step 1.**  $\gamma = 0.95$  is required. **Step 2.** The corresponding  $c$  equals 1.960; see Table 25.1. **Step 3.**  $\bar{x} = 5$  is given. **Step 4.** We need  $k = 1.960 \cdot 3/\sqrt{100} = 0.588$ . Hence  $\bar{x} - k = 4.412$ ,  $\bar{x} + k = 5.588$  and the confidence interval is  $\text{CONF}_{0.95} \{4.412 \leq \mu \leq 5.588\}$ .

This is sometimes written  $\mu = 5 \pm 0.588$ , but we shall not use this notation, which can be misleading.

With your CAS you can determine this interval more directly. Similarly for the other examples in this section. ■

**Theory for Table 25.1.** The method in Table 25.1 follows from the basic

### THEOREM 1

#### Sum of Independent Normal Random Variables

Let  $X_1, \dots, X_n$  be **independent** normal random variables each of which has mean  $\mu$  and variance  $\sigma^2$ . Then the following holds.

- (a) The sum  $X_1 + \dots + X_n$  is normal with mean  $n\mu$  and variance  $n\sigma^2$ .
- (b) The following random variable  $\bar{X}$  is normal with mean  $\mu$  and variance  $\sigma^2/n$ .

$$(4) \quad \bar{X} = \frac{1}{n} (X_1 + \dots + X_n)$$

- (c) The following random variable  $Z$  is normal with mean 0 and variance 1.

$$(5) \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

**PROOF** The statements about the mean and variance in (a) follow from Theorems 1 and 3 in Sec. 24.9. From this, and Theorem 2 in Sec. 24.6, we see that  $\bar{X}$  has the mean  $(1/n)n\mu = \mu$  and the variance  $(1/n)^2 n\sigma^2 = \sigma^2/n$ . This implies that  $Z$  has the mean 0 and variance 1, by Theorem 2(b) in Sec. 24.6. The normality of  $X_1 + \cdots + X_n$  is proved in Ref. [G3] listed in App. 1. This implies the normality of (4) and (5). ■

**Derivation of (3) in Table 25.1.** Sampling from a normal distribution gives independent sample values (see Sec. 25.1), so that Theorem 1 applies. Hence we can choose  $\gamma$  and then determine  $c$  such that

$$(6) \quad P(-c \leq Z \leq c) = P\left(-c \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq c\right) = \Phi(c) - \Phi(-c) = \gamma.$$

For the value  $\gamma = 0.95$  we obtain  $z(D) = 1.960$  from Table A8 in App. 5, as used in Example 1. For  $\gamma = 0.9, 0.99, 0.999$  we get the other values of  $c$  listed in Table 25.1. Finally, all we have to do is to convert the inequality in (6) into one for  $\mu$  and insert observed values obtained from the sample. We multiply  $-c \leq Z \leq c$  by  $-1$  and then by  $\sigma/\sqrt{n}$ , writing  $c\sigma/\sqrt{n} = k$  (as in Table 25.1),

$$\begin{aligned} P(-c \leq Z \leq c) &= P(c \geq -Z \geq -c) = P\left(c \geq \frac{\mu - \bar{X}}{\sigma/\sqrt{n}} \geq -c\right) \\ &= P(k \geq \mu - \bar{X} \geq -k) = \gamma. \end{aligned}$$

Adding  $\bar{X}$  gives  $P(\bar{X} + k \geq \mu \geq \bar{X} - k) = \gamma$  or

$$(7) \quad P(\bar{X} - k \leq \mu \leq \bar{X} + k) = \gamma.$$

Inserting the observed value  $\bar{x}$  of  $\bar{X}$  gives (3). Here we have regarded  $x_1, \dots, x_n$  as single observations of  $X_1, \dots, X_n$  (the standard trick!), so that  $x_1 + \cdots + x_n$  is an observed value of  $X_1 + \cdots + X_n$  and  $\bar{x}$  is an observed value of  $\bar{X}$ . Note further that (7) is of the form (2) with  $\Theta_1 = \bar{X} - k$  and  $\Theta_2 = \bar{X} + k$ . ■

### EXAMPLE 2 Sample Size Needed for a Confidence Interval of Prescribed Length

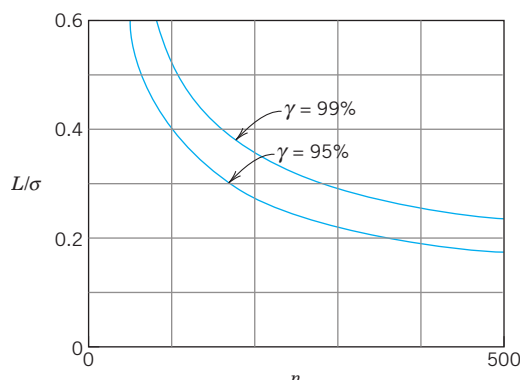
How large must  $n$  be in Example 1 if we want to obtain a 95% confidence interval of length  $L = 0.4$ ?

**Solution.** The interval (3) has the length  $L = 2k = 2c\sigma/\sqrt{n}$ . Solving for  $n$ , we obtain

$$n = (2c\sigma/L)^2.$$

In the present case the answer is  $n = (2 \cdot 1.960 \cdot 3/0.4)^2 \approx 870$ .

Figure 526 shows how  $L$  decreases as  $n$  increases and that for  $\gamma = 99\%$  the confidence interval is substantially longer than for  $\gamma = 95\%$  (and the same sample size  $n$ ). ■



**Fig. 526.** Length of the confidence interval (3) (measured in multiples of  $\sigma$ ) as a function of the sample size  $n$  for  $\gamma = 95\%$  and  $\gamma = 99\%$

## Confidence Interval for $\mu$ of the Normal Distribution with Unknown $\sigma^2$

In practice  $\sigma^2$  is frequently unknown. Then the method in Table 25.1 does not help and the whole theory changes, although the steps of determining a confidence interval for  $\mu$  remain quite similar. They are shown in Table 25.2. We see that  $k$  differs from that in Table 25.1, namely, the sample standard deviation  $s$  has taken the place of the unknown standard deviation  $\sigma$  of the population. And  $c$  now depends on the sample size  $n$  and must be determined from Table A9 in App. 5 or from your CAS. That table lists values  $z$  for given values of the distribution function (Fig. 527)

$$(8) \quad F(z) = K_m \int_{-\infty}^x \left(1 + \frac{u^2}{m}\right)^{-(m+1)/2} du$$

of the ***t*-distribution**. Here,  $m (= 1, 2, \dots)$  is a parameter, called the **number of degrees of freedom** of the distribution (**abbreviated d.f.**). In the present case,  $m = n - 1$ ; see Table 25.2. The constant  $K_m$  is such that  $F(\infty) = 1$ . By integration it turns out that  $K_m = \Gamma(\frac{1}{2}m + \frac{1}{2}) / [\sqrt{m\pi} \Gamma(\frac{1}{2}m)]$ , where  $\Gamma$  is the gamma function (see (24) in App. A3.1).

**Table 25.2** Determination of a Confidence Interval for the Mean  $\mu$  of a Normal Distribution with Unknown Variance  $\sigma^2$

**Step 1.** Choose a confidence level  $\gamma$  (95%, 99%, or the like).

**Step 2.** Determine the solution  $c$  of the equation

$$(9) \quad F(c) = \frac{1}{2}(1 + \gamma)$$

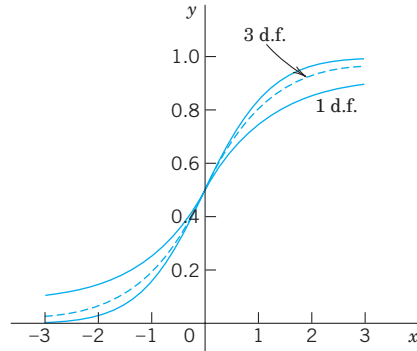
from the table of the *t*-distribution with  $n - 1$  degrees of freedom (Table A9 in App. 5; or use a CAS;  $n$  = sample size).

**Step 3.** Compute the mean  $\bar{x}$  and the variance  $s^2$  of the sample  $x_1, \dots, x_n$ .

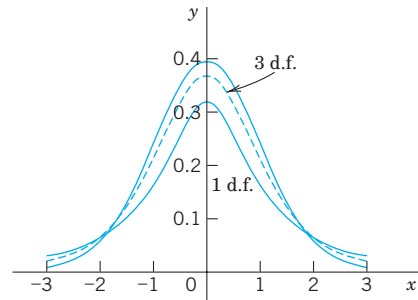
**Step 4.** Compute  $k = cs/\sqrt{n}$ . The confidence interval is

$$(10) \quad \text{CONF}_\gamma \{ \bar{x} - k \leq \mu \leq \bar{x} + k \}.$$

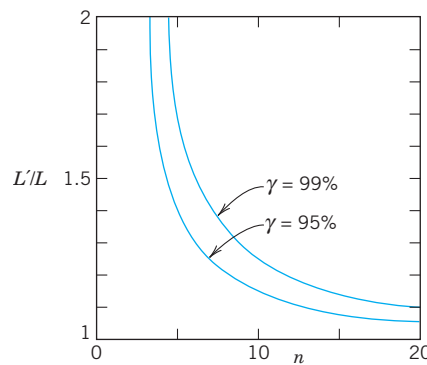
Figure 528 compares the curve of the density of the  $t$ -distribution with that of the normal distribution. The latter is steeper. This illustrates that Table 25.1 (which uses more information, namely, the known value of  $\sigma^2$ ) yields shorter confidence intervals than Table 25.2. This is confirmed in Fig. 529, which also gives an idea of the gain by increasing the sample size.



**Fig. 527.** Distribution functions of the  $t$ -distribution with 1 and 3 d.f. and of the standardized normal distribution (steepest curve)



**Fig. 528.** Densities of the  $t$ -distribution with 1 and 3 d.f. and of the standardized normal distribution



**Fig. 529.** Ratio of the lengths  $L'$  and  $L$  of the confidence intervals (10) and (3) with  $\gamma = 95\%$  and  $\gamma = 99\%$  as a function of the sample size  $n$  for equal  $s$  and  $\sigma$

### EXAMPLE 3 Confidence Interval for $\mu$ of the Normal Distribution with Unknown $\sigma^2$

Five independent measurements of the point of inflammation (flash point) of Diesel oil (D-2) gave the values (in  $^{\circ}\text{F}$ ) 144 147 146 142 144. Assuming normality, determine a 99% confidence interval for the mean.

**Solution.** *Step 1.*  $\gamma = 0.99$  is required.

*Step 2.*  $F(c) = \frac{1}{2}(1 + \gamma) = 0.995$ , and Table A9 in App. 5 with  $n - 1 = 4$  d.f. gives  $c = 4.60$ .

*Step 3.*  $\bar{x} = 144.6$ ,  $s^2 = 3.8$ .

*Step 4.*  $k = \sqrt{3.8} \cdot 4.60 / \sqrt{5} = 4.01$ . The confidence interval is  $\text{CONF}_{0.99} \{140.5 \leq \mu \leq 148.7\}$ .

If the variance  $\sigma^2$  were known and equal to the sample variance  $s^2$ , thus  $\sigma^2 = 3.8$ , then Table 25.1 would give  $k = c\sigma/\sqrt{n} = 2.576\sqrt{3.8}/\sqrt{5} = 2.25$  and  $\text{CONF}_{0.99} \{142.35 \leq \mu \leq 146.85\}$ . We see that the present interval is almost twice as long as that obtained from Table 25.1 (with  $\sigma^2 = 3.8$ ). Hence for small samples the difference is considerable! See also Fig. 529. ■

**Theory for Table 25.2.** For deriving (10) in Table 25.2 we need from Ref. [G3]

### THEOREM 2

#### Student's $t$ -Distribution

Let  $X_1, \dots, X_n$  be independent normal random variables with the same mean  $\mu$  and the same variance  $\sigma^2$ . Then the random variable

$$(11) \quad T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has a  $t$ -distribution [see (8)] with  $n - 1$  degrees of freedom (d.f.); here  $\bar{X}$  is given by (4) and

$$(12) \quad S^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2.$$

**Derivation of (10).** This is similar to the derivation of (3). We choose a number  $\gamma$  between 0 and 1 and determine a number  $c$  from Table A9 in App. 5 with  $n - 1$  d.f. (or from a CAS) such that

$$(13) \quad P(-c \leq T \leq c) = F(c) - F(-c) = \gamma.$$

Since the  $t$ -distribution is symmetric, we have

$$F(-c) = 1 - F(c),$$

and (13) assumes the form (9). Substituting (11) into (13) and transforming the result as before, we obtain

$$(14) \quad P(\bar{X} - K \leq \mu \leq \bar{X} + K) = \gamma$$

where

$$K = cS/\sqrt{n}.$$

By inserting the observed values  $\bar{x}$  of  $\bar{X}$  and  $s^2$  of  $S^2$  into (14) we finally obtain (10). ■

## Confidence Interval for the Variance $\sigma^2$ of the Normal Distribution

Table 25.3 shows the steps, which are similar to those in Tables 25.1 and 25.2.

**Table 25.3** Determination of a Confidence Interval for the Variance  $\sigma^2$  of a Normal Distribution, Whose Mean Need Not Be Known**Step 1.** Choose a confidence level  $\gamma$  (95%, 99%, or the like).**Step 2.** Determine solutions  $c_1$  and  $c_2$  of the equations

$$(15) \quad F(c_1) = \frac{1}{2}(1 - \gamma), \quad F(c_2) = \frac{1}{2}(1 + \gamma)$$

from the table of the chi-square distribution with  $n - 1$  degrees of freedom (Table A10 in App. 5; or use a CAS;  $n$  = sample size).**Step 3.** Compute  $(n - 1)s^2$ , where  $s^2$  is the variance of the sample  $x_1, \dots, x_n$ .**Step 4.** Compute  $k_1 = (n - 1)s^2/c_1$  and  $k_2 = (n - 1)s^2/c_2$ . The confidence interval is

$$(16) \quad \text{CONF}_\gamma \{k_2 \leq \sigma^2 \leq k_1\}.$$

**EXAMPLE 4** Confidence Interval for the Variance of the Normal DistributionDetermine a 95% confidence interval (16) for the variance, using Table 25.3 and a sample (tensile strength of sheet steel in  $\text{kg/mm}^2$ , rounded to integer values)

89 84 87 81 89 86 91 90 78 89 87 99 83 89.

**Solution.** **Step 1.**  $\gamma = 0.95$  is required.**Step 2.** For  $n - 1 = 13$  we find

$$c_1 = 5.01 \quad \text{and} \quad c_2 = 24.74.$$

**Step 3.**  $13s^2 = 326.9$ .**Step 4.**  $13s^2/c_1 = 65.25$ ,  $13s^2/c_2 = 13.21$ .

The confidence interval is

$$\text{CONF}_{0.95} \{13.21 \leq \sigma^2 \leq 65.25\}.$$

This is rather large, and for obtaining a more precise result, one would need a much larger sample. ■

**Theory for Table 25.3.** In Table 25.1 we used the normal distribution, in Table 25.2 the  $t$ -distribution, and now we shall use the  $\chi^2$ -distribution (*chi-square distribution*), whose distribution function is  $F(z) = 0$  if  $z < 0$  and

$$F(z) = C_m \int_0^z e^{-u/2} u^{(m-2)/2} du \quad \text{if } z \geq 0 \quad (\text{Fig. 530}).$$

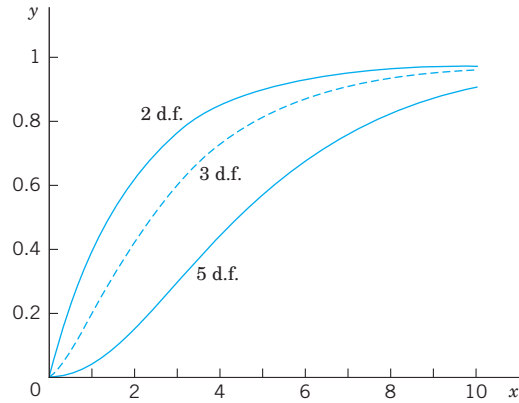
The parameter  $m (= 1, 2, \dots)$  is called the **number of degrees of freedom** (d.f.), and

$$C_m = 1/[2^{m/2} \Gamma(\frac{1}{2}m)].$$

Note that the distribution is not symmetric (see also Fig. 531).



For deriving (16) in Table 25.3 we need the following theorem.



**Fig. 530.** Distribution function of the chi-square distribution with 2, 3, 5 d.f.

### THEOREM 3

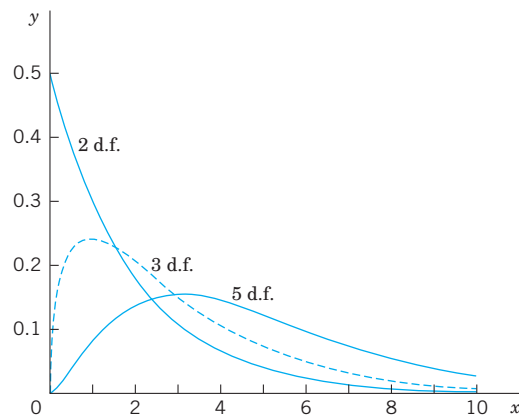
#### Chi-Square Distribution

*Under the assumptions in Theorem 2 the random variable*

$$(17) \quad Y = (n - 1) \frac{S^2}{\sigma^2}$$

*with  $S^2$  given by (12) has a chi-square distribution with  $n - 1$  degrees of freedom.*

Proof in Ref. [G3], listed in App. 1.



**Fig. 531.** Density of the chi-square distribution with 2, 3, 5 d.f.

**Derivation of (16).** This is similar to the derivation of (3) and (10). We choose a number  $\gamma$  between 0 and 1 and determine  $c_1$  and  $c_2$  from Table A10, App. 5, such that [see (15)]

$$P(Y \leq c_1) = F(c_1) = \frac{1}{2}(1 - \gamma), \quad P(Y \leq c_2) = F(c_2) = \frac{1}{2}(1 + \gamma).$$

Subtraction yields

$$P(c_1 \leq Y \leq c_2) = P(Y \leq c_2) - P(Y \leq c_1) = F(c_2) - F(c_1) = \gamma.$$

Transforming  $c_1 \leq Y \leq c_2$  with  $Y$  given by (17) into an inequality for  $\sigma^2$ , we obtain

$$\frac{n-1}{c_2} S^2 \leq \sigma^2 \leq \frac{n-1}{c_1} S^2.$$

By inserting the observed value  $s^2$  of  $S^2$  we obtain (16). ■

## Confidence Intervals for Parameters of Other Distributions

The methods in Tables 25.1–25.3 for confidence intervals for  $\mu$  and  $\sigma^2$  are designed for the normal distribution. We now show that they can also be applied to other distributions if we use large samples.

We know that if  $X_1, \dots, X_n$  are independent random variables with the same mean  $\mu$  and the same variance  $\sigma^2$ , then their sum  $Y_n = X_1 + \dots + X_n$  has the following properties.

(A)  $Y_n$  has the mean  $n\mu$  and the variance  $n\sigma^2$  (by Theorems 1 and 3 in Sec. 24.9).

(B) If those variables are normal, then  $Y_n$  is normal (by Theorem 1).

If those random variables are not normal, then (B) is not applicable. However, for large  $n$  the random variable  $Y_n$  is still *approximately* normal. This follows from the central limit theorem, which is one of the most fundamental results in probability theory.

### THEOREM 4

#### Central Limit Theorem

Let  $X_1, \dots, X_n, \dots$  be independent random variables that have the same distribution function and therefore the same mean  $\mu$  and the same variance  $\sigma^2$ . Let  $Y_n = X_1 + \dots + X_n$ . Then the random variable

$$(18) \quad Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}}$$

is **asymptotically normal** with mean 0 and variance 1; that is, the distribution function  $F_n(x)$  of  $Z_n$  satisfies

$$\lim_{n \rightarrow \infty} F_n(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du.$$

A proof can be found in Ref. [G3] listed in App. 1.

Hence, when applying Tables 25.1–25.3 to a nonnormal distribution, we must use **sufficiently large samples**. As a rule of thumb, if the sample indicates that the skewness of the distribution (the asymmetry; see Team Project 20(d), Problem Set 24.6) is small, use at least  $n = 20$  for the mean and at least  $n = 50$  for the variance.

### PROBLEM SET 25.3

- Why are interval estimates generally more useful than point estimates?

#### 2–6 MEAN (VARIANCE KNOWN)

- Find a 95% confidence interval for the mean of a normal population with standard deviation 4.00 from the sample 39, 51, 49, 43, 57, 59. Does that interval get longer or shorter if we take  $\gamma = 0.99$  instead of 0.95? By what factor?
- By what factor does the length of the interval in Prob. 2 change if we double the sample size?
- Determine a 95% confidence interval for the mean  $\mu$  of a normal population with variance  $\sigma^2 = 16$ , using a sample of size 200 with mean 74.81.
- What sample size would be needed for obtaining a 95% confidence interval (3) of length  $2\sigma$ ? Of length  $\sigma$ ?
- What sample size is needed to obtain a 99% confidence interval of length 2.0 for the mean of a normal population with variance 25? Use Fig. 526. Check by calculation.

#### MEAN (VARIANCE UNKNOWN)

- Find a 95% confidence interval for the percentage of cars on a certain highway that have poorly adjusted brakes, using a random sample of 800 cars stopped at a roadblock on that highway, 126 of which had poorly adjusted brakes.
- K. Pearson result.** Find a 99% confidence interval for  $p$  in the binomial distribution from a classical result by K. Pearson, who in 24,000 trials of tossing a coin obtained 12,012 Heads. Do you think that the coin was fair?

- 9–11** Find a 99% confidence interval for the mean of a normal population from the sample:

- Copper content (%) of brass 66, 66, 65, 64, 66, 67, 64, 65, 63, 64
- Melting point ( $^{\circ}\text{C}$ ) of aluminum 660, 667, 654, 663, 662
- Knoop hardness of diamond 9500, 9800, 9750, 9200, 9400, 9550

- CAS EXPERIMENT. Confidence Intervals.** Obtain 100 samples of size 10 of the standardized normal distribution. Calculate from them and graph the corresponding 95% confidence intervals for the mean and count how many of them do not contain 0. Does the result support the theory? Repeat the whole experiment, compare and comment.

#### 13–17 VARIANCE

Find a 95% confidence interval for the variance of a normal population from the sample:

- Length of 20 bolts with sample mean 20.2 cm and sample variance  $0.04 \text{ cm}^2$
- Carbon monoxide emission (grams per mile) of a certain type of passenger car (cruising at 55 mph): 17.3, 17.8, 18.0, 17.7, 18.2, 17.4, 17.6, 18.1
- Mean energy (keV) of delayed neutron group (Group 3, half-life 6.2 s) for uranium  $\text{U}^{235}$  fission: a sample of 100 values with mean 442.5 and variance 9.3
- Ultimate tensile strength (k psi) of alloy steel (Maraging H) at room temperature: 251, 255, 258, 253, 253, 252, 250, 252, 255, 256
- The sample in Prob. 9
- If  $X_1$  and  $X_2$  are independent normal random variables with mean 14 and 8 and variance 2 and 5, respectively, what distribution does  $3X_1 - X_2$  have? *Hint.* Use Team Project 14(g) in Sec. 24.8.
- A machine fills boxes weighing  $Y$  lb with  $X$  lb of salt, where  $X$  and  $Y$  are normal with mean 100 lb and 5 lb and standard deviation 1 lb and 0.5 lb, respectively. What percent of filled boxes weighing between 104 lb and 106 lb are to be expected?
- If the weight  $X$  of bags of cement is normally distributed with a mean of 40 kg and a standard deviation of 2 kg, how many bags can a delivery truck carry so that the probability of the total load exceeding 2000 kg will be 5%?

## 25.4 Testing of Hypotheses. Decisions

The ideas of confidence intervals and of tests<sup>2</sup> are the two most important ideas in modern statistics. In a statistical **test** we make inference from sample to population through testing a **hypothesis**, resulting from experience or observations, from a theory or a quality requirement, and so on. In many cases the result of a test is used as a basis for a **decision**, for instance, to

<sup>2</sup>Beginning around 1930, a systematic theory of tests was developed by NEYMAN (see Sec. 25.3) and EGON SHARPE PEARSON (1895–1980), English statistician, the son of Karl Pearson (see the footnote on p. 1086).

buy (or not to buy) a certain model of car, depending on a test of the fuel efficiency (miles/gal) (and other tests, of course), to apply some medication, depending on a test of its effect; to proceed with a marketing strategy, depending on a test of consumer reactions, etc.

Let us explain such a test in terms of a typical example and introduce the corresponding standard notions of statistical testing.

### EXAMPLE 1 Test of a Hypothesis. Alternative. Significance Level $\alpha$

We want to buy 100 coils of a certain kind of wire, provided we can verify the manufacturer's claim that the wire has a breaking limit  $\mu = \mu_0 = 200$  lb (or more). This is a test of the **hypothesis** (also called *null hypothesis*)  $\mu = \mu_0 = 200$ . We shall not buy the wire if the (statistical) test shows that actually  $\mu = \mu_1 < \mu_0$ , the wire is weaker, the claim does not hold.  $\mu_1$  is called the **alternative** (or *alternative hypothesis*) of the test. We shall **accept** the hypothesis if the test suggests that it is true, except for a small error probability  $\alpha$ , called the **significance level** of the test. Otherwise we **reject** the hypothesis. Hence  $\alpha$  is the probability of rejecting a hypothesis although it is true. The choice of  $\alpha$  is up to us. 5% and 1% are popular values.

For the test we need a sample. We randomly select 25 coils of the wire, cut a piece from each coil, and determine the breaking limit experimentally. Suppose that this sample of  $n = 25$  values of the breaking limit has the mean  $\bar{x} = 197$  lb (somewhat less than the claim!) and the standard deviation  $s = 6$  lb.

At this point we could only speculate whether this difference  $197 - 200 = -3$  is due to randomness, is a chance effect, or whether it is **significant**, due to the actually inferior quality of the wire. To continue beyond speculation requires probability theory, as follows.

We assume that the breaking limit is normally distributed. (This assumption could be tested by the method in Sec. 25.7. Or we could remember the central limit theorem (Sec. 25.3) and take a still larger sample.) Then

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

in (11), Sec. 25.3, with  $\mu = \mu_0$  has a **t-distribution** with  $n - 1$  degrees of freedom ( $n - 1 = 24$  for our sample). Also  $\bar{x} = 197$  and  $s = 6$  are observed values of  $\bar{X}$  and  $S$  to be used later. We can now choose a significance level, say,  $\alpha = 5\%$ . From Table A9 in App. 5 or from a CAS we then obtain a critical value  $c$  such that  $P(T \leq c) = \alpha = 5\%$ . For  $P(T \leq \tilde{c}) = 1 - \alpha = 95\%$  the table gives  $\tilde{c} = 1.71$ , so that  $c = -\tilde{c} = -1.71$  because of the symmetry of the distribution (Fig. 532).

We now reason as follows—this is the **crucial idea** of the test. If the hypothesis is true, we have a chance of only  $\alpha (= 5\%)$  that we observe a value  $t$  of  $T$  (calculated from a sample) that will fall between  $-\infty$  and  $-1.71$ . Hence, if we nevertheless do observe such a  $t$ , we assert that the hypothesis cannot be true and we reject it. Then we accept the alternative. If, however,  $t \geq c$ , we accept the hypothesis.

A simple calculation finally gives  $t = (197 - 200)/(6/\sqrt{25}) = -2.5$  as an observed value of  $T$ . Since  $-2.5 < -1.71$ , we reject the hypothesis (the manufacturer's claim) and accept the alternative  $\mu = \mu_1 < 200$ , the wire seems to be weaker than claimed. ■

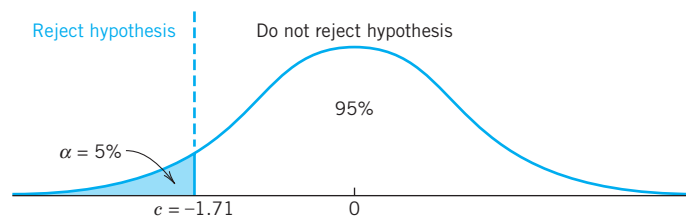


Fig. 532. t-distribution in Example 1

This example illustrates the **steps of a test**:

1. Formulate the **hypothesis**  $\theta = \theta_0$  to be tested. ( $\theta_0 = \mu_0$  in the example.)
2. Formulate an **alternative**  $\theta = \theta_1$ . ( $\theta_1 = \mu_1$  in the example.)
3. Choose a **significance level**  $\alpha$  (5%, 1%, 0.1%).
4. Use a random variable  $\hat{\Theta} = g(X_1, \dots, X_n)$  whose distribution depends on the hypothesis and on the alternative, and this distribution is known in both cases. Determine

a critical value  $c$  from the distribution of  $\hat{\Theta}$ , assuming the hypothesis to be true. (In the example,  $\hat{\Theta} = T$ , and  $c$  is, obtained from  $P(T \leq c) = \alpha$ .)

5. Use a sample  $x_1, \dots, x_n$  to determine an observed value  $\hat{\theta} = g(x_1, \dots, x_n)$  of  $\hat{\Theta}$ . ( $t$  in the example.)

6. Accept or reject the hypothesis, depending on the size of  $\hat{\theta}$  relative to  $c$ . ( $t < c$  in the example, rejection of the hypothesis.)

Two important facts require further discussion and careful attention. The first is the choice of an alternative. In the example,  $\mu_1 < \mu_0$ , but other applications may require  $\mu_1 > \mu_0$  or  $\mu_1 \neq \mu_0$ . The second fact has to do with errors. We know that  $\alpha$  (the significance level of the test) is the probability of *rejecting a true* hypothesis. And we shall discuss the probability  $\beta$  of *accepting a false* hypothesis.

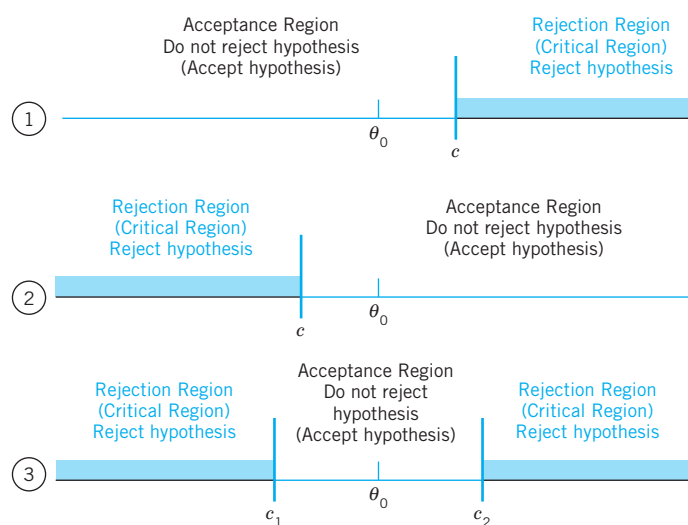
### One-Sided and Two-Sided Alternatives (Fig. 533)

Let  $\theta$  be an unknown parameter in a distribution, and suppose that we want to test the hypothesis  $\theta = \theta_0$ . Then there are three main kinds of alternatives, namely,

- (1)  $\theta > \theta_0$
- (2)  $\theta < \theta_0$
- (3)  $\theta \neq \theta_0$ .

(1) and (2) are **one-sided alternatives**, and (3) is a **two-sided alternative**.

We call **rejection region** (or **critical region**) the region such that we reject the hypothesis if the observed value in the test falls in this region. In ① the critical  $c$  lies to the right of  $\theta_0$  because so does the alternative. Hence the rejection region extends to the right. This is called a **right-sided test**. In ② the critical  $c$  lies to the left of  $\theta_0$  (as in Example 1), the rejection region extends to the left, and we have a **left-sided test** (Fig. 533, middle part). These are **one-sided tests**. In ③ we have two rejection regions. This is called a **two-sided test** (Fig. 533, lower part).



**Fig. 533.** Test in the case of alternative (1) (upper part of the figure), alternative (2) (middle part), and alternative (3)

All three kinds of alternatives occur in practical problems. For example, (1) may arise if  $\theta_0$  is the maximum tolerable inaccuracy of a voltmeter or some other instrument. Alternative (2) may occur in testing strength of material, as in Example 1. Finally,  $\theta_0$  in (3) may be the diameter of axle-shafts, and shafts that are too thin or too thick are equally undesirable, so that we have to watch for deviations in both directions.

## Errors in Tests

Tests always involve **risks of making false decisions**:

- (I) Rejecting a true hypothesis (**Type I error**).  
 $\alpha$  = Probability of making a Type I error.
- (II) Accepting a false hypothesis (**Type II error**).  
 $\beta$  = Probability of making a Type II error.

Clearly, we cannot avoid these errors because no absolutely certain conclusions about populations can be drawn from samples. But we show that there are ways and means of choosing suitable levels of risks, that is, of values  $\alpha$  and  $\beta$ . The choice of  $\alpha$  depends on the nature of the problem (e.g., a small risk  $\alpha = 1\%$  is used if it is a matter of life or death).

Let us discuss this systematically for a test of a hypothesis  $\theta = \theta_0$  against an alternative that is a single number  $\theta_1$ , for simplicity. We let  $\theta_1 > \theta_0$ , so that we have a right-sided test. For a left-sided or a two-sided test the discussion is quite similar.

We choose a critical  $c > \theta_0$  (as in the upper part of Fig. 533, by methods discussed below). From a given sample  $x_1, \dots, x_n$  we then compute a value

$$\hat{\theta} = g(x_1, \dots, x_n)$$

with a suitable  $g$  (whose choice will be a main point of our further discussion; for instance, take  $g = (x_1 + \dots + x_n)/n$  in the case in which  $\theta$  is the mean). If  $\hat{\theta} > c$ , we reject the hypothesis. If  $\hat{\theta} \leq c$ , we accept it. Here, the value  $\hat{\theta}$  can be regarded as an observed value of the random variable

$$(4) \quad \hat{\Theta} = g(X_1, \dots, X_n)$$

because  $x_j$  may be regarded as an observed value of  $X_j$ ,  $j = 1, \dots, n$ . In this test there are two possibilities of making an error, as follows.

**Type I Error** (see Table 25.4). The hypothesis is true but is rejected (hence the alternative is accepted) because  $\Theta$  assumes a value  $\hat{\theta} > c$ . Obviously, the probability of making such an error equals

$$(5) \quad P(\hat{\Theta} > c)_{\theta=\theta_0} = \alpha.$$

$\alpha$  is called the **significance level** of the test, as mentioned before.

**Type II Error** (see Table 25.4). The hypothesis is false but is accepted because  $\hat{\Theta}$  assumes a value  $\hat{\theta} \leq c$ . The probability of making such an error is denoted by  $\beta$ ; thus

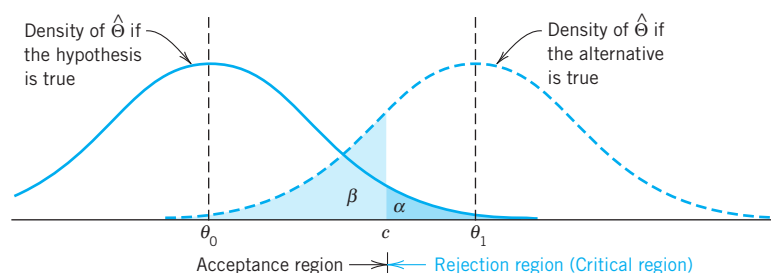
$$(6) \quad P(\hat{\Theta} \leq c)_{\theta=\theta_1} = \beta.$$

$\eta = 1 - \beta$  is called the **power** of the test. Obviously, the power  $\eta$  is the probability of avoiding a Type II error.

**Table 25.4** Type I and Type II Errors in Testing a Hypothesis  $\theta = \theta_0$  Against an Alternative  $\theta = \theta_1$

		Unknown Truth	
		$\theta = \theta_0$	$\theta = \theta_1$
Accepted	$\theta = \theta_0$	True decision $P = 1 - \alpha$	Type II error $P = \beta$
	$\theta = \theta_1$	Type I error $P = \alpha$	True decision $P = 1 - \beta$

Formulas (5) and (6) show that both  $\alpha$  and  $\beta$  depend on  $c$ , and we would like to choose  $c$  so that these probabilities of making errors are as small as possible. But the important Figure 534 shows that these are conflicting requirements because to let  $\alpha$  decrease we must shift  $c$  to the right, but then  $\beta$  increases. In practice we first choose  $\alpha$  (5%, sometimes 1%), then determine  $c$ , and finally compute  $\beta$ . If  $\beta$  is large so that the power  $\eta = 1 - \beta$  is small, we should repeat the test, choosing a larger sample, for reasons that will appear shortly.



**Fig. 534.** Illustration of Type I and II errors in testing a hypothesis  $\theta = \theta_0$  against an alternative  $\theta = \theta_1$  ( $> \theta_0$ , right-sided test)

If the alternative is not a single number but is of the form (1)–(3), then  $\beta$  becomes a function of  $\theta$ . This function  $\beta(\theta)$  is called the **operating characteristic** (OC) of the test and its curve the **OC curve**. Clearly, in this case  $\eta = 1 - \beta$  also depends on  $\theta$ . This function  $\eta(\theta)$  is called the **power function** of the test. (Examples will follow.)

Of course, from a test that leads to the acceptance of a certain hypothesis  $\theta_0$ , it does **not** follow that this is the only possible hypothesis or the best possible hypothesis. Hence the terms “**not reject**” or “**fail to reject**” are perhaps better than the term “**accept**.”

## Test for $\mu$ of the Normal Distribution with Known $\sigma^2$

The following example explains the three kinds of hypotheses.

### EXAMPLE 2

#### Test for the Mean of the Normal Distribution with Known Variance

Let  $X$  be a normal random variable with variance  $\sigma^2 = 9$ . Using a sample of size  $n = 10$  with mean  $\bar{x}$ , test the hypothesis  $\mu = \mu_0 = 24$  against the three kinds of alternatives, namely,

- (a)  $\mu > \mu_0$       (b)  $\mu < \mu_0$       (c)  $\mu \neq \mu_0$ .

**Solution.** We choose the significance level  $\alpha = 0.05$ . An estimate of the mean will be obtained from

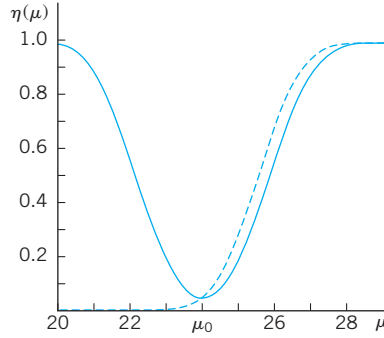
$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n).$$

If the hypothesis is true,  $\bar{X}$  is normal with mean  $\mu = 24$  and variance  $\sigma^2/n = 0.9$ , see Theorem 1, Sec. 25.3. Hence we may obtain the critical value  $c$  from Table A8 in App. 5.

**Case (a). Right-Sided Test.** We determine  $c$  from  $P(\bar{X} > c)_{\mu=24} = \alpha = 0.05$ , that is,

$$P(\bar{X} \leq c)_{\mu=24} = \Phi\left(\frac{c - 24}{\sqrt{0.9}}\right) = 1 - \alpha = 0.95.$$

Table A8 in App. 5 gives  $(c - 24)/\sqrt{0.9} = 1.645$ , and  $c = 25.56$ , which is greater than  $\mu_0$ , as in the upper part of Fig. 533. If  $\bar{x} \leq 25.56$ , the hypothesis is accepted. If  $\bar{x} > 25.56$ , it is rejected. The power function of the test is (Fig. 535)



**Fig. 535.** Power function  $\eta(\mu)$  in Example 2, case (a) (dashed) and case (c)

$$\begin{aligned} \eta(\mu) &= P(\bar{X} > 25.56)_{\mu} = 1 - P(\bar{X} \leq 25.56)_{\mu} \\ (7) \quad &= 1 - \Phi\left(\frac{25.56 - \mu}{\sqrt{0.9}}\right) = 1 - \Phi(26.94 - 1.05\mu) \end{aligned}$$

**Case (b). Left-Sided Test.** The critical value  $c$  is obtained from the equation

$$P(\bar{X} \leq c)_{\mu=24} = \Phi\left(\frac{c - 24}{\sqrt{0.9}}\right) = \alpha = 0.05.$$

Table A8 in App. 5 yields  $c = 24 - 1.56 = 22.44$ . If  $\bar{x} \geq 22.44$ , we accept the hypothesis. If  $\bar{x} < 22.44$ , we reject it. The power function of the test is

$$(8) \quad \eta(\mu) = P(\bar{X} \geq 22.44)_{\mu} = \Phi\left(\frac{22.44 - \mu}{\sqrt{0.9}}\right) = \Phi(23.65 - 1.05\mu).$$

**Case (c). Two-Sided Test.** Since the normal distribution is symmetric, we choose  $c_1$  and  $c_2$  equidistant from  $\mu = 24$ , say,  $c_1 = 24 - k$  and  $c_2 = 24 + k$ , and determine  $k$  from

$$P(24 - k \leq \bar{X} \leq 24 + k)_{\mu=24} = \Phi\left(\frac{k}{\sqrt{0.9}}\right) - \Phi\left(-\frac{k}{\sqrt{0.9}}\right) = 1 - \alpha = 0.95.$$



Table A8 in App. 5 gives  $k/\sqrt{0.9} = 1.960$ , hence  $k = 1.86$ . This gives the values  $c_1 = 24 - 1.86 = 22.14$  and  $c_2 = 24 + 1.86 = 25.86$ . If  $\bar{x}$  is not smaller than  $c_1$  and not greater than  $c_2$ , we accept the hypothesis. Otherwise we reject it. The power function of the test is (Fig. 535)

$$\begin{aligned}
 \eta(\mu) &= P(\bar{X} < 22.14)_\mu + P(\bar{X} > 25.86)_\mu = P(\bar{X} < 22.14)_\mu + 1 - P(\bar{X} \leq 25.86)_\mu \\
 (9) \quad &= 1 + \Phi\left(\frac{22.14 - \mu}{\sqrt{0.9}}\right) - \Phi\left(\frac{25.86 - \mu}{\sqrt{0.9}}\right) \\
 &= 1 + \Phi(23.34 - 1.05\mu) - \Phi(27.26 - 1.05\mu).
 \end{aligned}$$

Consequently, the operating characteristic  $\beta(\mu) = 1 - \eta(\mu)$  (see before) is (Fig. 536)

$$\beta(\mu) = \Phi(27.26 - 1.05\mu) - \Phi(23.34 - 1.05\mu).$$

If we take a larger sample, say, of size  $n = 100$  (instead of 10), then  $\sigma^2/n = 0.09$  (instead of 0.9) and the critical values are  $c_1 = 23.41$  and  $c_2 = 24.59$ , as can be readily verified. Then the operating characteristic of the test is

$$\begin{aligned}
 \beta(\mu) &= \Phi\left(\frac{24.59 - \mu}{\sqrt{0.09}}\right) - \Phi\left(\frac{23.41 - \mu}{\sqrt{0.09}}\right) \\
 &= \Phi(81.97 - 3.33\mu) - \Phi(78.03 - 3.33\mu).
 \end{aligned}$$

Figure 536 shows that the corresponding OC curve is steeper than that for  $n = 10$ . This means that the increase of  $n$  has led to an improvement of the test. In any practical case,  $n$  is chosen as small as possible but so large that the test brings out deviations between  $\mu$  and  $\mu_0$  that are of practical interest. For instance, if deviations of  $\pm 2$  units are of interest, we see from Fig. 536 that  $n = 10$  is much too small because when  $\mu = 24 - 2 = 22$  or  $\mu = 24 + 2 = 26$   $\beta$  is almost 50%. On the other hand, we see that  $n = 100$  is sufficient for that purpose. ■

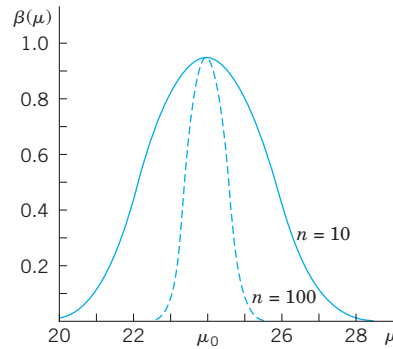


Fig. 536. Curves of the operating characteristic (OC curves) in Example 2, case (c), for two different sample sizes  $n$

## Test for $\mu$ When $\sigma^2$ Is Unknown, and for $\sigma^2$

### EXAMPLE 3 Test for the Mean of the Normal Distribution with Unknown Variance

The tensile strength of a sample of  $n = 16$  manila ropes (diameter 3 in.) was measured. The sample mean was  $\bar{x} = 4482$  kg, and the sample standard deviation was  $s = 115$  kg (N. C. Wiley, 41st Annual Meeting of the American Society for Testing Materials). Assuming that the tensile strength is a normal random variable, test the hypothesis  $\mu_0 = 4500$  kg against the alternative  $\mu_1 = 4400$  kg. Here  $\mu_0$  may be a value given by the manufacturer, while  $\mu_1$  may result from previous experience.

**Solution.** We choose the significance level  $\alpha = 5\%$ . If the hypothesis is true, it follows from Theorem 2 in Sec. 25.3, that the random variable

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{\bar{X} - 4500}{S/4}$$

has a  $t$ -distribution with  $n - 1 = 15$  d.f. The test is left-sided. The critical value  $c$  is obtained from  $P(T < c)_{\mu_0} = \alpha = 0.05$ . Table A9 in App. 5 gives  $c = -1.75$ . As an observed value of  $T$  we obtain from the sample  $t = (4482 - 4500)/(115/4) = -0.626$ . We see that  $t > c$  and accept the hypothesis. For obtaining numeric values of the power of the test, we would need tables called noncentral Student  $t$ -tables; we shall not discuss this question here. ■

#### EXAMPLE 4 Test for the Variance of the Normal Distribution

Using a sample of size  $n = 15$  and sample variance  $s^2 = 13$  from a normal population, test the hypothesis  $\sigma^2 = \sigma_0^2 = 10$  against the alternative  $\sigma^2 = \sigma_1^2 = 20$ .

**Solution.** We choose the significance level  $\alpha = 5\%$ . If the hypothesis is true, then

$$Y = (n - 1) \frac{S^2}{\sigma_0^2} = 14 \frac{S^2}{10} = 1.4S^2$$

has a chi-square distribution with  $n - 1 = 14$  d.f. by Theorem 3, Sec. 25.3. From

$$P(Y > c) = \alpha = 0.05, \quad \text{that is,} \quad P(Y \leq c) = 0.95,$$

and Table A10 in App. 5 with 14 degrees of freedom we obtain  $c = 23.68$ . This is the critical value of  $Y$ . Hence to  $S^2 = \sigma_0^2 Y / (n - 1) = 0.714Y$  there corresponds the critical value  $c^* = 0.714 \cdot 23.68 = 16.91$ . Since  $s^2 < c^*$ , we accept the hypothesis.

If the alternative is true, the random variable  $Y_1 = 14S^2/\sigma_1^2 = 0.7S^2$  has a chi-square distribution with 14 d.f. Hence our test has the power

$$\eta = P(S^2 > c^*)_{\sigma^2=20} = P(Y_1 > 0.7c^*)_{\sigma^2=20} = 1 - P(Y_1 \leq 11.84)_{\sigma^2=20}.$$

From a more extensive table of the chi-square distribution (e.g. in Ref. [G3] or [G8]) or from your CAS, you see that  $\eta \approx 62\%$ . Hence the Type II risk is very large, namely, 38%. To make this risk smaller, we would have to increase the sample size. ■

## Comparison of Means and Variances

#### EXAMPLE 5 Comparison of the Means of Two Normal Distributions

Using a sample  $x_1, \dots, x_{n_1}$  from a normal distribution with unknown mean  $\mu_x$  and a sample  $y_1, \dots, y_{n_2}$  from another normal distribution with unknown mean  $\mu_y$ , we want to test the hypothesis that the means are equal,  $\mu_x = \mu_y$ , against an alternative, say,  $\mu_x > \mu_y$ . The variances need not be known but are assumed to be equal.<sup>3</sup>

Two cases of comparing means are of practical importance:

**Case A.** *The samples have the same size. Furthermore, each value of the first sample corresponds to precisely one value of the other, because corresponding values result from the same person or thing (paired comparison)—* for example, two measurements of the same thing by two different methods or two measurements from the two eyes of the same person. More generally, they may result from pairs of *similar* individuals or things, for example, identical twins, pairs of used front tires from the same car, etc. Then we should form the differences of corresponding values and test the hypothesis that the population corresponding to the differences has mean 0, using the method in Example 3. If we have a choice, this method is better than the following.

<sup>3</sup>This assumption of equality of variances can be tested, as shown in the next example. If the test shows that they differ significantly, choose two samples of the same size  $n_1 = n_2 = n$  (not too small,  $> 30$ , say), use the test in Example 2 together with the fact that (12) is an observed value of an approximately standardized normal random variable.

**Case B.** The two samples are independent and not necessarily of the same size. Then we may proceed as follows. Suppose that the alternative is  $\mu_x > \mu_y$ . We choose a significance level  $\alpha$ . Then we compute the sample means  $\bar{x}$  and  $\bar{y}$  as well as  $(n_1 - 1)s_x^2$  and  $(n_2 - 1)s_y^2$ , where  $s_x^2$  and  $s_y^2$  are the sample variances. Using Table A9 in App. 5 with  $n_1 + n_2 - 2$  degrees of freedom, we now determine  $c$  from

$$(10) \quad P(T \leq c) = 1 - \alpha.$$

We finally compute

$$(11) \quad t_0 = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \frac{\bar{x} - \bar{y}}{\sqrt{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}}.$$

It can be shown that this is an observed value of a random variable that has a  $t$ -distribution with  $n_1 + n_2 - 2$  degrees of freedom, provided the hypothesis is true. If  $t_0 \leq c$ , the hypothesis is accepted. If  $t_0 > c$ , it is rejected.

If the alternative is  $\mu_x \neq \mu_y$ , then (10) must be replaced by

$$(10^*) \quad P(T \leq c_1) = 0.5\alpha, \quad P(T \leq c_2) = 1 - 0.5\alpha.$$

Note that for samples of equal size  $n_1 = n_2 = n$ , formula (11) reduces to

$$(12) \quad t_0 = \sqrt{n} \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2 + s_y^2}}.$$

To illustrate the computations, let us consider the two samples  $(x_1, \dots, x_{n_1})$  and  $(y_1, \dots, y_{n_2})$  given by

	105	108	86	103	103	107	124	105
and								
	89	92	84	97	103	107	111	97

showing the relative output of tin plate workers under two different working conditions [J. J. B. Worth, *Journal of Industrial Engineering* **9**, 249–253). Assuming that the corresponding populations are normal and have the same variance, let us test the hypothesis  $\mu_x = \mu_y$  against the alternative  $\mu_x \neq \mu_y$ . (Equality of variances will be tested in the next example.)

**Solution.** We find

$$\bar{x} = 105.125, \quad \bar{y} = 97.500, \quad s_x^2 = 106.125, \quad s_y^2 = 84.000.$$

We choose the significance level  $\alpha = 5\%$ . From (10\*) with  $0.5\alpha = 2.5\%$ ,  $1 - 0.5\alpha = 97.5\%$  and Table A9 in App. 5 with 14 degrees of freedom we obtain  $c_1 = -2.14$  and  $c_2 = 2.14$ . Formula (12) with  $n = 8$  gives the value

$$t_0 = \sqrt{8} \cdot 7.625 / \sqrt{190.125} = 1.56.$$

Since  $c_1 \leq t_0 \leq c_2$ , we **accept the hypothesis**  $\mu_x = \mu_y$  that under both conditions the mean output is the same.

Case A applies to the example because the two first sample values correspond to a certain type of work, the next two were obtained in another kind of work, etc. So we may use the differences

$$16 \quad 16 \quad 2 \quad 6 \quad 0 \quad 0 \quad 13 \quad 8$$

of corresponding sample values and the method in Example 3 to test the hypothesis  $\mu = 0$ , where  $\mu$  is the mean of the population corresponding to the differences. As a logical alternative we take  $\mu \neq 0$ . The sample mean is  $\bar{d} = 7.625$ , and the sample variance is  $s^2 = 45.696$ . Hence

$$t = \sqrt{8} (7.625 - 0) / \sqrt{45.696} = 3.19.$$

From  $P(T \leq c_1) = 2.5\%$ ,  $P(T \leq c_2) = 97.5\%$  and Table A9 in App. 5 with  $n - 1 = 7$  degrees of freedom we obtain  $c_1 = -2.36$ ,  $c_2 = 2.36$  and **reject the hypothesis** because  $t = 3.19$  does not lie between  $c_1$  and  $c_2$ . Hence our present test, in which we used more information (but the same samples), shows that the difference in output is significant. ■

**EXAMPLE 6** Comparison of the Variance of Two Normal Distributions

Using the two samples in the last example, test the hypothesis  $\sigma_x^2 = \sigma_y^2$ ; assume that the corresponding populations are normal and the nature of the experiment suggests the alternative  $\sigma_x^2 > \sigma_y^2$ .

**Solution.** We find  $s_x^2 = 106.125$ ,  $s_y^2 = 84.000$ . We choose the significance level  $\alpha = 5\%$ . Using  $P(V \leq c) = 1 - \alpha = 95\%$  and Table A11 in App. 5, with  $(n_1 - 1, n_2 - 1) = (7, 7)$  degrees of freedom, we determine  $c = 3.79$ . We finally compute  $v_0 = s_x^2/s_y^2 = 1.26$ . Since  $v_0 \leq c$ , we accept the hypothesis. If  $v_0 < c$ , we would reject it.

This test is justified by the fact that  $v_0$  is an observed value of a random variable that has a so-called **F-distribution** with  $(n_1 - 1, n_2 - 1)$  degrees of freedom, provided the hypothesis is true. (Proof in Ref. [G3] listed in App. 1.) The *F*-distribution with  $(m, n)$  degrees of freedom was introduced by R. A. Fisher<sup>4</sup> and has the distribution function  $F(z) = 0$  if  $z < 0$  and

$$(13) \quad F(z) = K_{mn} \int_0^z t^{(m-2)/2} (mt + n)^{-(m+n)/2} dt \quad (z \geq 0),$$

where  $K_{mn} = m^{m/2} n^{n/2} \Gamma(\frac{1}{2}m + \frac{1}{2}n) / \Gamma(\frac{1}{2}m) \Gamma(\frac{1}{2}n)$ . (For  $\Gamma$  see App. A3.1.)

This long section contained the basic ideas and concepts of testing, along with typical applications and you may perhaps want to review it quickly before going on, because the next sections concern an adaptation of these ideas to tasks of great practical importance and resulting tests in connection with quality control, acceptance (or rejection) of goods produced, and so on.

**PROBLEM SET 25.4**

- From memory: Make a list of the three types of alternatives, each with a typical example of your own.
- Make a list of methods in this section, each with the distribution needed in testing.
- Test  $\mu = 0$  against  $\mu > 0$ , assuming normality and using the sample 0, 1, -1, 3, -8, 6, 1 (deviations of the azimuth [multiples of 0.01 radian] in some revolution of a satellite). Choose  $\alpha = 5\%$ .
- In one of his classical experiments Buffon obtained 2048 heads in tossing a coin 4040 times. Was the coin fair?
- Do the same test as in Prob. 4, using a result by K. Pearson, who obtained 6019 heads in 12,000 trials.
- Assuming normality and known variance  $\sigma^2 = 9$ , test the hypothesis  $\mu = 60.0$  against the alternative  $\mu = 57.0$  using a sample of size 20 with mean  $\bar{x} = 58.50$  and choosing  $\alpha = 5\%$ .
- How does the result in Prob. 6 change if we use a smaller sample, say, of size 5, the other data ( $\bar{x} = 58.05$ ,  $\alpha = 5\%$ , etc.) remaining as before?
- Determine the power of the test in Prob. 6.
- What is the rejection region in Prob. 6 in the case of a two-sided test with  $\alpha = 5\%$ ?
- CAS EXPERIMENT. Tests of Means and Variances.**
  - Obtain 100 samples of size 10 each from the normal distribution with mean 100 and variance 25. For each sample, test the hypothesis  $\mu_0 = 100$  against the alternative  $\mu_1 > 100$  at the level of  $\alpha = 10\%$ . Record the number of rejections of the hypothesis. Do the whole experiment once more and compare.
  - Set up a similar experiment for the variance of a normal distribution and perform it 100 times.
- A firm sells oil in cans containing 5000 g oil per can and is interested to know whether the mean weight differs significantly from 5000 g at the 5% level, in which case the filling machine has to be adjusted. Set up a hypothesis and an alternative and perform the test, assuming normality and using a sample of 50 fillings with mean 4990 g and standard deviation 20 g.

<sup>4</sup>After the pioneering work of the English statistician and biologist, KARL PEARSON (1857–1936), the founder of the English school of statistics, and WILLIAM SEALY GOSSET (1876–1937), who discovered the *t*-distribution (and published under the name “Student”), the English statistician Sir RONALD AYLMER FISHER (1890–1962), professor of eugenics in London (1933–1943) and professor of genetics in Cambridge, England (1943–1957) and Adelaide, Australia (1957–1962), had great influence on the further development of modern statistics.

12. If a sample of 25 tires of a certain kind has a mean life of 37,000 miles and a standard deviation of 5000 miles, can the manufacturer claim that the true mean life of such tires is greater than 35,000 miles? Set up and test a corresponding hypothesis at the 5% level, assuming normality.
13. If simultaneous measurements of electric voltage by two different types of voltmeter yield the differences (in volts) 0.4, -0.6, 0.2, 0.0, 1.0, 1.4, 0.4, 1.6, can we assert at the 5% level that there is no significant difference in the calibration of the two types of instruments? Assume normality.
14. If a standard medication cures about 75% of patients with a certain disease and a new medication cured 310 of the first 400 patients on whom it was tried, can we conclude that the new medication is better? Choose  $\alpha = 5\%$ . First guess. Then calculate.
15. Suppose that in the past the standard deviation of weights of certain 100.0-oz packages filled by a machine was 0.8 oz. Test the hypothesis  $H_0: \sigma = 0.8$  against the alternative  $H_1: \sigma > 0.8$  (an undesirable increase), using a sample of 20 packages with standard deviation 1.0 oz and assuming normality. Choose  $\alpha = 5\%$ .
16. Suppose that in operating battery-powered electrical equipment, it is less expensive to replace all batteries at fixed intervals than to replace each battery individually when it breaks down, provided the standard deviation of the lifetime is less than a certain limit, say, less than 5 hours. Set up and apply a suitable test, using a sample of 28 values of lifetimes with standard deviation  $s = 3.5$  hours and assuming normality: choose  $\alpha = 5\%$ .
17. Brand A gasoline was used in 16 similar automobiles under identical conditions. The corresponding sample of 16 values (miles per gallon) had mean 19.6 and standard deviation 0.4. Under the same conditions, high-power brand B gasoline gave a sample of 16 values with mean 20.2 and standard deviation 0.6. Is the mileage of B significantly better than that of A? Test at the 5% level; assume normality. First guess. Then calculate.
18. The two samples 70, 80, 30, 70, 60, 80 and 140, 120, 130, 120, 120, 130, 120 are values of the differences of temperatures ( $^{\circ}\text{C}$ ) of iron at two stages of casting, taken from two different crucibles. Is the variance of the first population larger than that of the second? Assume normality. Choose  $\alpha = 5\%$ .
19. Show that for a normal distribution the two types of errors in a test of a hypothesis  $H_0: \mu = \mu_0$  against an alternative  $H_1: \mu = \mu_1$  can be made as small as one pleases (not zero!) by taking the sample sufficiently large.
20. Test for equality of population means against the alternative that the means are different assuming normality, choosing  $\alpha = 5\%$  and using two samples of sizes 12 and 18, with mean 10 and 14, respectively, and equal standard deviation 3.

## 25.5 Quality Control

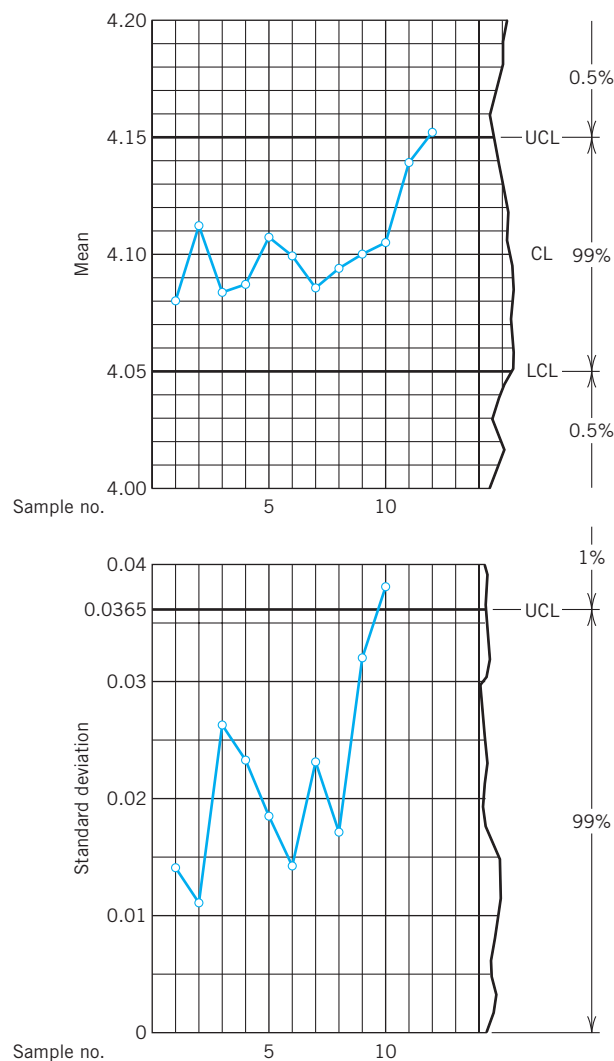
The ideas on testing can be adapted and extended in various ways to serve basic practical needs in engineering and other fields. We show this in the remaining sections for some of the most important tasks solvable by statistical methods. As a first such area of problems, we discuss industrial quality control, a highly successful method used in various industries.

No production process is so perfect that all the products are completely alike. There is always a small variation that is caused by a great number of small, uncontrollable factors and must therefore be regarded as a chance variation. It is important to make sure that the products have required values (for example, length, strength, or whatever property may be essential in a particular case). For this purpose one makes a test of the hypothesis that the products have the required property, say,  $\mu = \mu_0$ , where  $\mu_0$  is a required value. If this is done after an entire lot has been produced (for example, a lot of 100,000 screws), the test will tell us how good or how bad the products are, but it is obviously too late to alter undesirable results. It is much better to test during the production run. This is done at regular intervals of time (for example, every hour or half-hour) and is called **quality control**. Each time a sample of the same size is taken, in practice 3 to 10 times. If the hypothesis is rejected, we stop the production and look for the cause of the trouble.

If we stop the production process even though it is progressing properly, we make a Type I error. If we do not stop the process even though something is not in order, we make a Type II error (see Sec. 25.4). The result of each test is marked in graphical form on what is called a **control chart**. This was proposed by W. A. Shewhart in 1924 and makes quality control particularly effective.

## Control Chart for the Mean

An illustration and example of a control chart is given in the upper part of Fig. 537. This control chart for the mean shows the **lower control limit** LCL, the **center control line** CL, and the **upper control limit** UCL. The two **control limits** correspond to the critical values  $c_1$  and  $c_2$  in case (c) of Example 2 in Sec. 25.4. As soon as a sample mean falls outside the range between the control limits, we reject the hypothesis and assert that the



**Fig. 537.** Control charts for the mean (upper part of figure) and the standard deviation in the case of the samples on p. 1089

production process is “out of control”; that is, we assert that there has been a shift in process level. Action is called for whenever a point exceeds the limits.

If we choose control limits that are too loose, we shall not detect process shifts. On the other hand, if we choose control limits that are too tight, we shall be unable to run the process because of frequent searches for nonexistent trouble. The usual significance level is  $\alpha = 1\%$ . From Theorem 1 in Sec. 25.3 and Table A8 in App. 5 we see that in the case of the normal distribution the corresponding control limits for the mean are

$$(1) \quad \text{LCL} = \mu_0 - 2.58 \frac{\sigma}{\sqrt{n}}, \quad \text{UCL} = \mu_0 + 2.58 \frac{\sigma}{\sqrt{n}}.$$

Here  $\sigma$  is assumed to be known. If  $\sigma$  is unknown, we may compute the standard deviations of the first 20 or 30 samples and take their arithmetic mean as an approximation of  $\sigma$ . The broken line connecting the means in Fig. 537 is merely to display the results.

Additional, more subtle controls are often used in industry. For instance, one observes the motions of the sample means above and below the centerline, which should happen frequently. Accordingly, long runs (conventionally of length 7 or more) of means all above (or all below) the centerline could indicate trouble.

**Table 25.5 Twelve Samples of Five Values Each  
(Diameter of Small Cylinders, Measured in Millimeters)**

Sample Number	Sample Values					$\bar{x}$	$s$	$R$
1	4.06	4.08	4.08	4.08	4.10	4.080	0.014	0.04
2	4.10	4.10	4.12	4.12	4.12	4.112	0.011	0.02
3	4.06	4.06	4.08	4.10	4.12	4.084	0.026	0.06
4	4.06	4.08	4.08	4.10	4.12	4.088	0.023	0.06
5	4.08	4.10	4.12	4.12	4.12	4.108	0.018	0.04
6	4.08	4.10	4.10	4.10	4.12	4.100	0.014	0.04
7	4.06	4.08	4.08	4.10	4.12	4.088	0.023	0.06
8	4.08	4.08	4.10	4.10	4.12	4.096	0.017	0.04
9	4.06	4.08	4.10	4.12	4.14	4.100	0.032	0.08
10	4.06	4.08	4.10	4.12	4.16	4.104	0.038	0.10
11	4.12	4.14	4.14	4.14	4.16	4.140	0.014	0.04
12	4.14	4.14	4.16	4.16	4.16	4.152	0.011	0.02

## Control Chart for the Variance

In addition to the mean, one often controls the variance, the standard deviation, or the range. To set up a control chart for the variance in the case of a normal distribution, we may employ the method in Example 4 of Sec. 25.4 for determining control limits. It is customary to use only one control limit, namely, an upper control limit. Now from Example 4 of Sec. 25.4 we have  $S^2 = \sigma_0^2 Y / (n - 1)$ , where, because of our normality assumption, the random variable  $Y$  has a chi-square distribution with  $n - 1$  degrees of freedom. Hence the desired control limit is

$$(2) \quad \text{UCL} = \frac{\sigma^2 c}{n - 1}$$

where  $c$  is obtained from the equation

$$P(Y > c) = \alpha, \quad \text{that is,} \quad P(Y \leq c) = 1 - \alpha$$

and the table of the chi-square distribution (Table A10 in App. 5) with  $n - 1$  degrees of freedom (or from your CAS); here  $\alpha$  (5% or 1%, say) is the probability that in a properly running process an observed value  $s^2$  of  $S^2$  is greater than the upper control limit.

If we wanted a control chart for the variance with both an upper control limit UCL and a lower control limit LCL, these limits would be

$$(3) \quad \text{LCL} = \frac{\sigma^2 c_1}{n - 1} \quad \text{and} \quad \text{UCL} = \frac{\sigma^2 c_2}{n - 1},$$

where  $c_1$  and  $c_2$  are obtained from Table A10 with  $n - 1$  d.f. and the equations

$$(4) \quad P(Y \leq c_1) = \frac{\alpha}{2} \quad \text{and} \quad P(Y \leq c_2) = 1 - \frac{\alpha}{2}.$$

## Control Chart for the Standard Deviation

To set up a control chart for the standard deviation, we need an upper control limit

$$(5) \quad \text{UCL} = \frac{\sigma \sqrt{c}}{\sqrt{n - 1}}$$

obtained from (2). For example, in Table 25.5 we have  $n = 5$ . Assuming that the corresponding population is normal with standard deviation  $\sigma = 0.02$  and choosing  $\alpha = 1\%$ , we obtain from the equation

$$P(Y \leq c) = 1 - \alpha = 99\%$$

and Table A10 in App. 5 with 4 degrees of freedom the critical value  $c = 13.28$  and from (5) the corresponding value

$$\text{UCL} = \frac{0.02 \sqrt{13.28}}{\sqrt{4}} = 0.0365,$$

which is shown in the lower part of Fig. 537.

A control chart for the standard deviation with both an upper and a lower control limit is obtained from (3).

## Control Chart for the Range

Instead of the variance or standard deviation, one often controls the **range**  $R$  (= largest sample value minus smallest sample value). It can be shown that in the case of the normal distribution, the standard deviation  $\sigma$  is proportional to the expectation of the random



variable  $R^*$  for which  $R$  is an observed value, say,  $\sigma = \lambda_n E(R^*)$  where the factor of proportionality  $\lambda_n$  depends on the sample size  $n$  and has the values

$n$	2	3	4	5	6	7	8	9	10
$\lambda_n = \sigma/E(R^*)$	0.89	0.59	0.49	0.43	0.40	0.37	0.35	0.34	0.32

$n$	12	14	16	18	20	30	40	50
$\lambda_n = \sigma/E(R^*)$	0.31	0.29	0.28	0.28	0.27	0.25	0.23	0.22

Since  $R$  depends on two sample values only, it gives less information about a sample than  $s$  does. Clearly, the larger the sample size  $n$  is, the more information we lose in using  $R$  instead of  $s$ . A practical rule is to use  $s$  when  $n$  is larger than 10.

## PROBLEM SET 25.5

- Suppose a machine for filling cans with lubricating oil is set so that it will generate fillings which form a normal population with mean 1 gal and standard deviation 0.02 gal. Set up a control chart of the type shown in Fig. 537 for controlling the mean, that is, find LCL and UCL, assuming that the sample size is 4.
- Three-sigma control chart.** Show that in Prob. 1, the requirement of the significance level  $\alpha = 0.3\%$  leads to  $LCL = \mu - 3\sigma/\sqrt{n}$  and  $UCL = \mu + 3\sigma/\sqrt{n}$ , and find the corresponding numeric values.
- What sample size should we choose in Prob. 1 if we want LCL and UCL somewhat closer together, say,  $UCL - LCL = 0.02$ , without changing the significance level?
- What effect on  $UCL - LCL$  does it have if we double the sample size? If we switch from  $\alpha = 1\%$  to  $\alpha = 5\%$ ?
- How should we change the sample size in controlling the mean of a normal population if we want  $UCL - LCL$  to decrease to half its original value?
- Graph the means of the following 10 samples (thickness of gaskets, coded values) on a control chart for means, assuming that the population is normal with mean 5 and standard deviation 1.16.

Time	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00
Sample	5	7	7	4	5	6	5	5	3	3
values	2	5	3	4	6	4	5	2	4	6
	5	4	6	3	4	6	6	5	8	6
	6	4	5	6	6	4	4	3	4	8

- Graph the ranges of the samples in Prob. 6 on a control chart for ranges.
- Graph  $\lambda_n = \sigma/E(R^*)$  as a function of  $n$ . Why is  $\lambda_n$  a monotone decreasing function of  $n$ ?
- Eight samples of size 2 were taken from a lot of screws. The values (length in inches) are

Sample No.	1	2	3	4	5	6	7	8
Length	3.50	3.51	3.49	3.52	3.53	3.49	3.48	3.52
	3.51	3.48	3.50	3.50	3.49	3.50	3.47	3.49

Assuming that the population is normal with mean 3.500 and variance 0.0004 and using (1), set up a control chart for the mean and graph the sample means on the chart.

- Attribute control charts.** Fifteen samples of size 100 were taken from a production of containers. The numbers of defectives (leaking containers) in those samples (in the order observed) were  
1 4 5 4 9 7 0 5 6 13 0 2 1 12 8  
From previous experience it was known that the average fraction defective is  $p = 4\%$  provided that the process of production is running properly. Using the binomial distribution, set up a *fraction defective chart* (also called a *p-chart*), that is, choose the

LCL = 0 and determine the UCL for the fraction defective (in percent) by the use of 3-sigma limits, where  $\sigma^2$  is the variance of the random variable

$$\bar{X} = \text{Fraction defective in a sample of size } 100.$$

Is the process under control?

11. **Number of defectives.** Find formulas for the UCL, CL, and LCL (corresponding to  $3\sigma$ -limits) in the case of a control chart for the number of defectives, assuming that, in a state of statistical control, the fraction of defectives is  $p$ .
12. **CAS PROJECT. Control Charts.** (a) Obtain 100 samples of 4 values each from the normal distribution with mean 8.0 and variance 0.16 and their means, variances, and ranges.
  - (b) Use these samples for making up a control chart for the mean.
  - (c) Use them on a control chart for the standard deviation.
  - (d) Make up a control chart for the range.
  - (e) Describe quantitative properties of the samples that you can see from those charts (e.g., whether the

corresponding process is under control, whether the quantities observed vary randomly, etc.).

13. Since the presence of a point outside control limits for the mean indicates trouble, how often would we be making the mistake of looking for nonexistent trouble if we used (a) 1-sigma limits, (b) 2-sigma limits? Assume normality.
14. What LCL and UCL should we use instead of (1) if, instead of  $\bar{x}$ , we use the sum  $x_1 + \cdots + x_n$  of the sample values? Determine these limits in the case of Fig. 537.
15. **Number of defects per unit.** A so-called *c-chart* or *defects-per-unit chart* is used for the control of the number  $X$  of defects per unit (for instance, the number of defects per 100 meters of paper, the number of missing rivets in an airplane wing, etc.). (a) Set up formulas for CL and LCL, UCL corresponding to  $\mu \pm 3\sigma$ , assuming that  $X$  has a Poisson distribution. (b) Compute CL, LCL, and UCL in a control process of the number of imperfections in sheet glass; assume that this number is 3.6 per sheet on the average when the process is in control.

## 25.6 Acceptance Sampling

**Acceptance sampling** is usually done when products leave the factory (or in some cases even within the factory). The standard situation in acceptance sampling is that a **producer** supplies to a **consumer** (a buyer or wholesaler) a lot of  $N$  items (a carton of screws, for instance). The decision to **accept** or **reject** the lot is made by determining the number  $x$  of **defectives** (= defective items) in a sample of size  $n$  from the lot. The lot is accepted if  $x \leq c$ , where  $c$  is called the **acceptance number**, giving the allowable number of defectives. If  $x > c$ , the consumer rejects the lot. Clearly, producer and consumer must agree on a certain **sampling plan** giving  $n$  and  $c$ .

From the hypergeometric distribution we see that the event  $A$ : “Accept the lot” has probability (see Sec. 24.7)

$$(1) \quad P(A) = P(X \leq c) = \sum_{x=0}^c \binom{M}{x} \binom{N-M}{n-x} / \binom{N}{n}$$

where  $M$  is the number of defectives in a lot of  $N$  items. In terms of the **fraction defective**  $\theta = M/N$  we can write (1) as

$$(2) \quad P(A; \theta) = \sum_{x=0}^c \binom{N\theta}{x} \binom{N-N\theta}{n-x} / \binom{N}{n}.$$

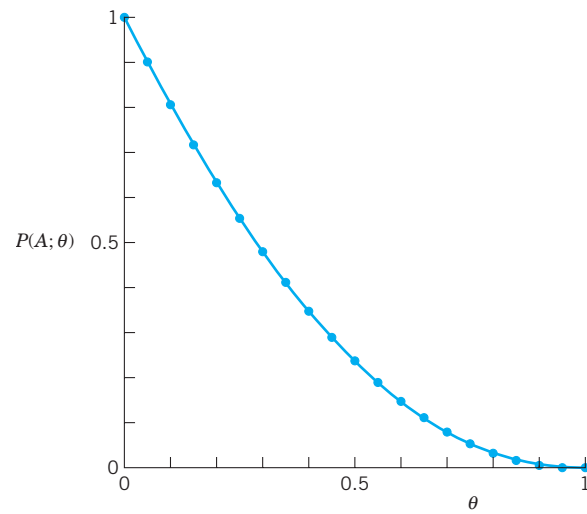
$P(A; \theta)$  can assume  $n+1$  values corresponding to  $\theta = 0, 1/N, 2/N, \dots, N/N$ ; here,  $n$  and  $c$  are fixed. A monotone smooth curve through these points is called the **operating characteristic curve (OC curve)** of the sampling plan considered.

**EXAMPLE 1** Sampling Plan

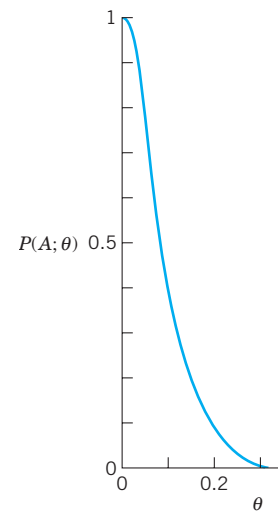
Suppose that certain tool bits are packaged 20 to a box, and the following sampling plan is used. A sample of two tool bits is drawn, and the corresponding box is accepted if and only if both bits in the sample are good. In this case,  $N = 20$ ,  $n = 2$ ,  $c = 0$ , and (2) takes the form (a factor 2 drops out)

$$\begin{aligned} P(A; \theta) &= \binom{20\theta}{0} \binom{20-20\theta}{2} / \binom{20}{2} \\ &= \frac{(20-20\theta)(19-20\theta)}{380}. \end{aligned}$$

The values of  $P(A, \theta)$  for  $\theta = 0, 1/20, 2/20, \dots, 20/20$  and the resulting OC curve are shown in Fig. 538. (Verify!) ■



**Fig. 538.** OC curve of the sampling plan with  $n = 2$  and  $c = 0$  for lots of size  $N = 20$



**Fig. 539.** OC curve in Example 2

In most practical cases  $\theta$  will be small (less than 10%). Then if we take small samples compared to  $N$ , we can approximate (2) by the Poisson distribution (Sec. 24.7); thus

$$(3) \quad P(A; \theta) \sim e^{-\mu} \sum_{x=0}^c \frac{\mu^x}{x!} \quad (\mu = n\theta).$$

**EXAMPLE 2** Sampling Plan. Poisson Distribution

Suppose that for large lots the following sampling plan is used. A sample of size  $n = 20$  is taken. If it contains not more than one defective, the lot is accepted. If the sample contains two or more defectives, the lot is rejected. In this plan, we obtain from (3)

$$P(A; \theta) \sim e^{-20\theta}(1 + 20\theta),$$

The corresponding OC curve is shown in Fig. 539. ■

## Errors in Acceptance Sampling

We show how acceptance sampling fits into general test theory (Sec. 25.4) and what this means from a practical point of view. The producer wants the probability  $\alpha$  of rejecting

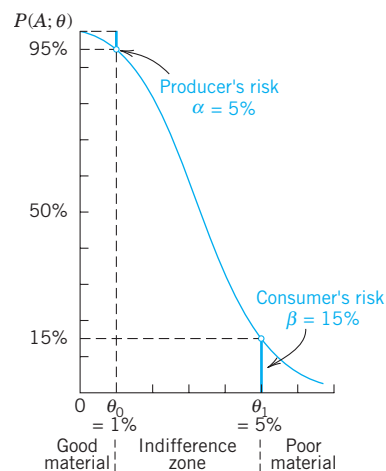


Fig. 540. OC curve, producer's and consumer's risks

an **acceptable lot** (a lot for which  $\theta$  does not exceed a certain number  $\theta_0$  on which the two parties agree) to be small.  $\theta_0$  is called the **acceptable quality level (AQL)**. Similarly, the consumer (the buyer) wants the probability  $\beta$  of accepting an **unacceptable lot** (a lot for which  $\theta$  is greater than or equal to some  $\theta_1$ ) to be small.  $\theta_1$  is called the **lot tolerance percent defective (LTPD)** or the **rejectable quality level (RQL)**.  $\alpha$  is called **producer's risk**. It corresponds to a Type I error in Sec. 25.4.  $\beta$  is called **consumer's risk** and corresponds to a Type II error. Figure 540 shows an example. We see that the points  $(\theta_0, 1 - \alpha)$  and  $(\theta_1, \beta)$  lie on the OC curve. It can be shown that for large lots we can choose  $\theta_0, \theta_1 (> \theta_0), \alpha, \beta$  and then determine  $n$  and  $c$  such that the OC curve runs very close to those prescribed points. Table 25.6 shows the analogy between acceptance sampling and hypothesis testing in Sec. 25.4.

Table 25.6 Acceptance Sampling and Hypothesis Testing

Acceptance Sampling	Hypothesis Testing
Acceptable quality level (AQL) $\theta = \theta_0$	Hypothesis $\theta = \theta_0$
Lot tolerance percent defectives (LTPD) $\theta = \theta_1$	Alternative $\theta = \theta_1$
Allowable number of defectives $c$	Critical value $c$
Producer's risk $\alpha$ of rejecting a lot with $\theta \leq \theta_0$	Probability $\alpha$ of making a Type I error (significance level)
Consumer's risk $\beta$ of accepting a lot with $\theta \geq \theta_1$	Probability $\beta$ of making a Type II error

## Rectification

**Rectification** of a *rejected* lot means that the lot is inspected item by item and all defectives are removed and replaced by nondefective items. (This may be too expensive if the lot is cheap; in this case the lot may be sold at a cut-rate price or scrapped.) If a production turns out  $100\theta\%$  defectives, then in  $K$  lots of size  $N$  each,  $KN\theta$  of the  $KN$  items are

defectives. Now  $KP(A; \theta)$  of these lots are accepted. These contain  $KPN\theta$  defectives, whereas the rejected and rectified lots contain no defectives, because of the rectification. Hence after the rectification the fraction defective in all  $K$  lots equals  $KPN\theta/KN$ . This is called the **average outgoing quality** (AOQ); thus

$$(4) \quad \text{AOQ}(\theta) = \theta P(A; \theta).$$

Figure 541 shows an example. Since  $\text{AOQ}(0) = 0$  and  $P(A; 1) = 0$ , the AOQ curve has a maximum at some  $\theta = \theta^*$ , giving the **average outgoing quality limit** (AOQL). This is the worst average quality that may be expected to be accepted under rectification.

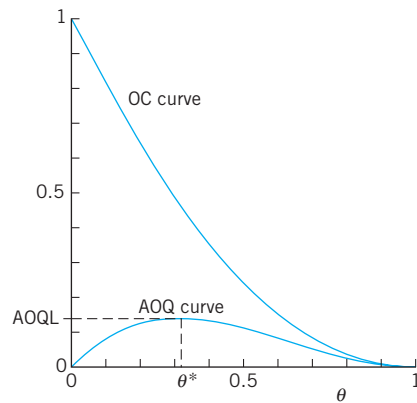


Fig. 541. OC curve and AOQ curve for the sampling plan in Fig. 538

## PROBLEM SET 25.6

1. Lots of kitchen knives are inspected by a sampling plan that uses a sample of size 20 and the acceptance number  $c = 1$ . What is the probability of accepting a lot with 1%, 2%, 10% defectives (knives with dull blades)? Use Table A6 of the Poisson distribution in App. 5. Graph the OC curve.
2. What happens in Prob. 1 if the sample size is increased to 50? First guess. Then calculate. Graph the OC curve and compare.
3. How will the probabilities in Prob. 1 with  $n = 20$  change (up or down) if we decrease  $c$  to zero? First guess.
4. What are the producer's and consumer's risks in Prob. 1 if the AQL is 2% and the RQL is 15%?
5. Lots of copper pipes are inspected according to a sample plan that uses sample size 25 and acceptance number 1. Graph the OC curve of the plan, using the Poisson approximation. Find the producer's risk if the AQL is 1.5%.
6. Graph the AOQ curve in Prob. 5. Determine the AOQL, assuming that rectification is applied.
7. In Example 1 in the text, what are the producer's and consumer's risks if the AQL is 0.1 and the RQL is 0.6?
8. What happens in Example 1 in the text if we increase the sample size to  $n = 3$ , leaving the other data as before? Compute  $P(A; 0.1)$  and  $P(A; 0.2)$  and compare with Example 1.
9. Graph and compare sampling plans with  $c = 1$  and increasing values of  $n$ , say,  $n = 2, 3, 4$ . (Use the binomial distribution.)
10. Find the binomial approximation of the hypergeometric distribution in Example 1 in the text and compare the approximate and the accurate values.

11. Samples of 3 fuses are drawn from lots and a lot is accepted if in the corresponding sample we find no more than 1 defective fuse. Criticize this sampling plan. In particular, find the probability of accepting a lot that is 50% defective. (Use the binomial distribution (7), Sec. 24.7.)
12. If in a sampling plan for large lots of spark plugs, the sample size is 100 and we want the AQL to be 5% and the producer's risk 2%, what acceptance number  $c$  should we choose? (Use the normal approximation of the binomial distribution in Sec. 24.8.)
13. What is the consumer's risk in Prob. 12 if we want the RQL to be 12%? Use  $c = 9$  from the answer of Prob. 12.
14. A lot of batteries for wrist watches is accepted if and only if a sample of 20 contains at most 1 defective. Graph the OC and AOQ curves. Find AOQL. [Use (3).]
15. Graph the OC curve and the AOQ curve for the single sampling plan for large lots with  $n = 5$  and  $c = 0$ , and find the AOQL.

## 25.7 Goodness of Fit. $\chi^2$ -Test

To test for **goodness of fit** means that we wish to test that a certain function  $F(x)$  is the distribution function of a distribution from which we have a sample  $x_1, \dots, x_n$ . Then we test whether the **sample distribution function**  $\tilde{F}(x)$  defined by

$$\tilde{F}(x) = \text{Sum of the relative frequencies of all sample values } x_j \text{ not exceeding } x$$

fits  $F(x)$  "sufficiently well." If this is so, we shall accept the hypothesis that  $F(x)$  is the distribution function of the population; if not, we shall reject the hypothesis.

This test is of considerable practical importance, and it differs in character from the tests for parameters ( $\mu$ ,  $\sigma^2$ , etc.) considered so far.

To test in that fashion, we have to know how much  $\tilde{F}(x)$  can differ from  $F(x)$  if the hypothesis is true. Hence we must first introduce a quantity that measures the deviation of  $\tilde{F}(x)$  from  $F(x)$ , and we must know the probability distribution of this quantity under the assumption that the hypothesis is true. Then we proceed as follows. We determine a number  $c$  such that, if the hypothesis is true, a deviation greater than  $c$  has a small preassigned probability. If, nevertheless, a deviation greater than  $c$  occurs, we have reason to doubt that the hypothesis is true and we reject it. On the other hand, if the deviation does not exceed  $c$ , so that  $\tilde{F}(x)$  approximates  $F(x)$  sufficiently well, we accept the hypothesis. Of course, if we accept the hypothesis, this means that we have insufficient evidence to reject it, and this does not exclude the possibility that there are other functions that would not be rejected in the test. In this respect the situation is quite similar to that in Sec. 25.4.

Table 25.7 shows a test of that type, which was introduced by R. A. Fisher. This test is justified by the fact that if the hypothesis is true, then  $\chi_0^2$  is an observed value of a random variable whose distribution function approaches that of the chi-square distribution with  $K - 1$  degrees of freedom (or  $K - r - 1$  degrees of freedom if  $r$  parameters are estimated) as  $n$  approaches infinity. The requirement that at least five sample values lie in each interval in Table 25.7 results from the fact that for finite  $n$  that random variable has only *approximately* a chi-square distribution. A proof can be found in Ref. [G3] listed in App. 1. If the sample is so small that the requirement cannot be satisfied, one may continue with the test, but then use the result with caution.

**Table 25.7** Chi-square Test for the Hypothesis That  $F(x)$  is the Distribution Function of a Population from Which a Sample  $x_1, \dots, x_n$  is Taken

**Step 1.** Subdivide the  $x$ -axis into  $K$  intervals  $I_1, I_2, \dots, I_K$  such that each interval contains at least 5 values of the given sample  $x_1, \dots, x_n$ . Determine the number  $b_j$  of sample values in the interval  $I_j$ , where  $j = 1, \dots, K$ . If a sample value lies at a common boundary point of two intervals, add 0.5 to each of the two corresponding  $b_j$ .

**Step 2.** Using  $F(x)$ , compute the probability  $p_j$  that the random variable  $X$  under consideration assumes any value in the interval  $I_j$ , where  $j = 1, \dots, K$ . Compute

$$e_j = np_j.$$

(This is the number of sample values theoretically expected in  $I_j$  if the hypothesis is true.)

**Step 3.** Compute the deviation

$$(1) \quad \chi_0^2 = \sum_{j=1}^K \frac{(b_j - e_j)^2}{e_j}.$$

**Step 4.** Choose a significance level (5%, 1%, or the like).

**Step 5.** Determine the solution  $c$  of the equation

$$P(\chi^2 \leq c) = 1 - \alpha$$

from the table of the chi-square distribution with  $K - 1$  degrees of freedom (Table A10 in App. 5). If  $r$  parameters of  $F(x)$  are unknown and their maximum likelihood estimates (Sec. 25.2) are used, then use  $K - r - 1$  degrees of freedom (instead of  $K - 1$ ). If  $\chi_0^2 \leq c$ , accept the hypothesis. If  $\chi_0^2 > c$ , reject the hypothesis.

**Table 25.8** Sample of 100 Values of the Splitting Tensile Strength (lb/in.<sup>2</sup>) of Concrete Cylinders

320	380	340	410	380	340	360	350	320	370
350	340	350	360	370	350	380	370	300	420
370	390	390	440	330	390	330	360	400	370
320	350	360	340	340	350	350	390	380	340
400	360	350	390	400	350	360	340	370	420
420	400	350	370	330	320	390	380	400	370
390	330	360	380	350	330	360	300	360	360
360	390	350	370	370	350	390	370	370	340
370	400	360	350	380	380	360	340	330	370
340	360	390	400	370	410	360	400	340	360

D. L. IVEY, Splitting tensile tests on structural lightweight aggregate concrete. Texas Transportation Institute, College Station, Texas.

**EXAMPLE 1** Test of Normality

Test whether the population from which the sample in Table 25.8 was taken is normal.

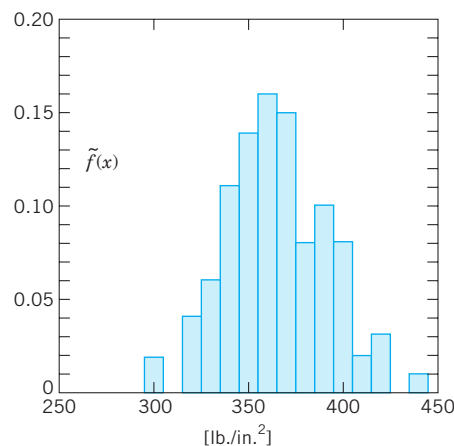
**Solution.** Table 25.8 shows the values (column by column) in the order obtained in the experiment. Table 25.9 gives the frequency distribution and Fig. 542 the histogram. It is hard to guess the outcome of the test—does the histogram resemble a normal density curve sufficiently well or not?

The maximum likelihood estimates for  $\mu$  and  $\sigma^2$  are  $\hat{\mu} = \bar{x} = 364.7$  and  $\hat{\sigma}^2 = 712.9$ . The computation in Table 25.10 yields  $\chi_0^2 = 2.688$ . It is very interesting that the interval  $375 \cdots 385$  contributes over 50% of  $\chi_0^2$ . From the histogram we see that the corresponding frequency looks much too small. The second largest contribution comes from  $395 \cdots 405$ , and the histogram shows that the frequency seems somewhat too large, which is perhaps not obvious from inspection.

**Table 25.9** Frequency Table of the Sample in Table 25.8

1 Tensile Strength $x$ [lb/in. <sup>2</sup> ]	2 Absolute Frequency	3 Relative Frequency $\tilde{f}(x)$	4 Cumulative Absolute Frequency	5 Cumulative Relative Frequency $\tilde{F}(x)$
300	2	0.02	2	0.02
310	0	0.00	2	0.02
320	4	0.04	6	0.06
330	6	0.06	12	0.12
340	11	0.11	23	0.23
350	14	0.14	37	0.37
360	16	0.16	53	0.53
370	15	0.15	68	0.68
380	8	0.08	76	0.76
390	10	0.10	86	0.86
400	8	0.08	94	0.94
410	2	0.02	96	0.96
420	3	0.03	99	0.99
430	0	0.00	99	0.99
440	1	0.01	100	1.00

We choose  $\alpha = 5\%$ . Since  $K = 10$  and we estimated  $r = 2$  parameters we have to use Table A10 in App. 5 with  $K - r - 1 = 7$  degrees of freedom. We find  $c = 14.07$  as the solution of  $P(\chi^2 \leq c) = 95\%$ . Since  $\chi_0^2 < c$ , we accept the hypothesis that the population is normal. ■



**Fig. 542.** Frequency histogram of the sample in Table 25.8



**Table 25.10** Computations in Example 1

$x_j$	$\frac{x_j - 364.7}{26.7}$	$\Phi\left(\frac{x_j - 364.7}{26.7}\right)$	$e_j$	$b_j$	Term in (1)
$-\infty \cdots 325$	$-\infty \cdots -1.49$	$0.0000 \cdots 0.0681$	6.81	6	0.096
$325 \cdots 335$	$-1.49 \cdots -1.11$	$0.0681 \cdots 0.1335$	6.54	6	0.045
$335 \cdots 345$	$-1.11 \cdots -0.74$	$0.1335 \cdots 0.2296$	9.61	11	0.201
$345 \cdots 355$	$-0.74 \cdots -0.36$	$0.2296 \cdots 0.3594$	12.98	14	0.080
$355 \cdots 365$	$-0.36 \cdots 0.01$	$0.3594 \cdots 0.5040$	14.46	16	0.164
$365 \cdots 375$	$0.01 \cdots 0.39$	$0.5040 \cdots 0.6517$	14.77	15	0.0004
$375 \cdots 385$	$0.39 \cdots 0.76$	$0.6517 \cdots 0.7764$	12.47	8	1.602
$385 \cdots 395$	$0.76 \cdots 1.13$	$0.7764 \cdots 0.8708$	9.44	10	0.033
$395 \cdots 405$	$1.13 \cdots 1.51$	$0.8708 \cdots 0.9345$	6.37	8	0.417
$405 \cdots \infty$	$1.51 \cdots \infty$	$0.9345 \cdots 1.0000$	6.55	6	0.046

$$\chi_0^2 = 2.688$$

## PROBLEM SET 25.7

- Verify the calculations in Example 1 of the text.
- If it is known that 25% of certain steel rods produced by a standard process will break when subjected to a load of 5000 lb, can we claim that a new, less expensive process yields the same breakage rate if we find that in a sample of 80 rods produced by the new process, 27 rods broke when subjected to that load? (Use  $\alpha = 5\%$ .)
- If 100 flips of a coin result in 40 heads and 60 tails, can we assert on the 5% level that the coin is fair?
- If in 10 flips of a coin we get the same ratio as in Prob. 3 (4 heads and 6 tails), is the conclusion the same as in Prob. 3? First conjecture, then compute.
- Can you claim, on a 5% level, that a die is fair if 60 trials give 1,  $\dots$ , 6 with absolute frequencies 10, 13, 9, 11, 9, 8?
- Solve Prob. 5 if rolling a die 180 times gives 33, 27, 29, 35, 25, 31.
- If a service station had served 60, 49, 56, 46, 68, 39 cars from Monday through Friday between 1 P.M. and 2 P.M., can one claim on a 5% level that the differences are due to randomness? First guess. Then calculate.
- A manufacturer claims that in a process of producing drill bits, only 2.5% of the bits are dull. Test the claim against the alternative that more than 2.5% of the bits are dull, using a sample of 400 bits containing 17 dull ones. Use  $\alpha = 5\%$ .
- In a table of properly rounded function values, even and odd last decimals should appear about equally often. Test this for the 90 values of  $J_1(x)$  in Table A1 in App. 5.
- TEAM PROJECT. Difficulty with Random Selection.** 77 students were asked to choose 3 of the integers 11, 12, 13,  $\dots$ , 30 completely arbitrarily. The amazing result was as follows.
 

Number	11	12	13	14	15	16	17	18	19	20
Frequ.	11	10	20	8	13	9	21	9	16	8

Number	21	22	23	24	25	26	27	28	29	30
Frequ.	12	8	15	10	10	9	12	8	13	9

If the selection were completely random, the following hypotheses should be true.

  - The 20 numbers are equally likely.
  - The 10 even numbers together are as likely as the 10 odd numbers together.
  - The 6 prime numbers together have probability 0.3 and the 14 other numbers together have probability 0.7.

Test these hypotheses, using  $\alpha = 5\%$ . Design further experiments that illustrate the difficulties of random selection.
- CAS EXPERIMENT. Random Number Generator.** Check your generator experimentally by imitating results of  $n$  trials of rolling a fair die, with a convenient  $n$  (e.g., 60 or 300 or the like). Do this many times and see whether you can notice any “nonrandomness” features, for example, too few Sixes, too many even numbers, etc., or whether your generator seems to work properly. Design and perform other kinds of checks.
- Test for normality at the 1% level using a sample of  $n = 79$  (rounded) values  $x$  (tensile strength [ $\text{kg}/\text{mm}^2$ ])

of steel sheets of 0.3 mm thickness).  $a = a(x)$  = absolute frequency. (Take the first two values together, also the last three, to get  $K = 5$ .)

$x$	57	58	59	60	61	62	63	64
$a$	4	10	17	27	8	9	3	1

- 13. Mendel's pathbreaking experiments.** In a famous plant-crossing experiment, the Austrian Augustinian father Gregor Mendel (1822–1884) obtained 355 yellow and 123 green peas. Test whether this agrees with Mendel's theory according to which the ratio should be 3:1.

- 14. Accidents in a foundry.** Does the random variable  $X$  = Number of accidents per week have a Poisson distribution if, within 50 weeks, 33 were accident-free, 1 accident occurred in 11 of the 50 weeks, 2 in 6 of

the weeks, and more than 2 accidents in no week? Choose  $\alpha = 5\%$ .

- 15. Radioactivity. Rutherford-Geiger experiments.**

Using the given sample, test that the corresponding population has a Poisson distribution.  $x$  is the number of alpha particles per 7.5-s intervals observed by E. Rutherford and H. Geiger in one of their classical experiments in 1910, and  $a(x)$  is the absolute frequency (= number of time periods during which exactly  $x$  particles were observed). Use  $\alpha = 5\%$ .

$x$	0	1	2	3	4	5	6
$a$	57	203	383	525	532	408	273
$x$	7	8	9	10	11	12	$\geq 13$
$a$	139	45	27	10	4	2	0

## 25.8 Nonparametric Tests

**Nonparametric tests**, also called **distribution-free tests**, are valid for any distribution. Hence they are used in cases when the kind of distribution is unknown, or is known but such that no tests specifically designed for it are available. In this section we shall explain the basic idea of these tests, which are based on “**order statistics**” and are rather simple. If there is a choice, then tests designed for a specific distribution generally give better results than do nonparametric tests. For instance, this applies to the tests in Sec. 25.4 for the normal distribution.

We shall discuss two tests in terms of typical examples. In deriving the distributions used in the test, it is essential that the distributions, from which we sample, are continuous. (Nonparametric tests can also be derived for discrete distributions, but this is slightly more complicated.)

### EXAMPLE 1 Sign Test for the Median

A **median** of the population is a solution  $x = \tilde{\mu}$  of the equation  $F(x) = 0.5$ , where  $F$  is the distribution function of the population.

Suppose that eight radio operators were tested, first in rooms without air-conditioning and then in air-conditioned rooms over the same period of time, and the difference of errors (unconditioned minus conditioned) were

$$9 \quad 4 \quad 0 \quad 6 \quad 4 \quad 0 \quad 7 \quad 11.$$

Test the hypothesis  $\tilde{\mu} = 0$  (that is, air-conditioning has no effect) against the alternative  $\tilde{\mu} > 0$  (that is, inferior performance in unconditioned rooms).

**Solution.** We choose the significance level  $\alpha = 5\%$ . If the hypothesis is true, the probability  $p$  of a positive difference is the same as that of a negative difference. Hence in this case,  $p = 0.5$ , and the random variable

$$X = \text{Number of positive values among } n \text{ values}$$

has a binomial distribution with  $p = 0.5$ . Our sample has eight values. We omit the values 0, which do not contribute to the decision. Then six values are left, all of which are positive. Since

$$\begin{aligned} P(X = 6) &= \binom{6}{6} (0.5)^6 (0.5)^0 \\ &= 0.0156 \\ &= 1.56\% \end{aligned}$$

we have observed an event whose probability is very small if the hypothesis is true; in fact  $1.56\% < \alpha = 5\%$ . Hence we assert that the alternative  $\tilde{\mu} > 0$  is true. That is, the number of errors made in unconditioned rooms is significantly higher, so that installation of air conditioning should be considered. ■

### EXAMPLE 2 Test for Arbitrary Trend

A certain machine is used for cutting lengths of wire. Five successive pieces had the lengths

$$29 \quad 31 \quad 28 \quad 30 \quad 32.$$

Using this sample, test the hypothesis that there is **no trend**, that is, the machine does not have the tendency to produce longer and longer pieces or shorter and shorter pieces. Assume that the type of machine suggests the alternative that there is *positive trend*, that is, there is the tendency of successive pieces to get longer.

**Solution.** We count the number of **transpositions** in the sample, that is, the number of times a larger value precedes a smaller value:

29 precedes 28 (1 transposition),

31 precedes 28 and 30 (2 transpositions).

The remaining three sample values follow in ascending order. Hence in the sample there are  $1 + 2 = 3$  transpositions. We now consider the random variable

$$T = \text{Number of transpositions.}$$

If the hypothesis is true (no trend), then each of the  $5! = 120$  permutations of five elements 1 2 3 4 5 has the same probability ( $1/120$ ). We arrange these permutations according to their number of transpositions:

$T = 0$	$T = 1$	$T = 2$	$T = 3$	
1 2 3 4 5	1 2 3 5 4	1 2 4 5 3	1 2 5 4 3	
	1 2 4 3 5	1 2 5 3 4	1 3 4 5 2	
	1 3 2 4 5	1 3 2 5 4	1 3 5 2 4	
	2 1 3 4 5	1 3 4 2 5	1 4 2 5 3	
		1 4 2 3 5	1 4 3 2 5	
		2 1 3 5 4	1 5 2 3 4	
		2 1 4 3 5	2 1 4 5 3	
		2 3 1 4 5	2 1 5 3 4	etc.
		3 1 2 4 5	2 3 1 5 4	
			2 3 4 1 5	
			2 4 1 3 5	
			3 1 2 5 4	
			3 1 4 2 5	
			3 2 1 4 5	
			4 1 2 3 5	

From this we obtain

$$P(T \leq 3) = \frac{1}{120} + \frac{4}{120} + \frac{9}{120} + \frac{15}{120} = \frac{29}{120} = 24\%.$$

We accept the hypothesis because we have observed an event that has a relatively large probability (certainly much more than 5%) if the hypothesis is true.

Values of the distribution function of  $T$  in the case of no trend are shown in Table A12, App. 5. For instance, if  $n = 3$ , then  $F(0) = 0.167$ ,  $F(1) = 0.500$ ,  $F(2) = 1 - 0.167$ . If  $n = 4$ , then  $F(0) = 0.042$ ,  $F(1) = 0.167$ ,  $F(2) = 0.375$ ,  $F(3) = 1 - 0.375$ ,  $F(4) = 1 - 0.167$ , and so on.

Our method and those values refer to *continuous* distributions. Theoretically, we may then expect that all the values of a sample are different. Practically, some sample values may still be equal, because of rounding: If  $m$  values are equal, add  $m(m-1)/4$  (= mean value of the transpositions in the case of the permutations of  $m$  elements), that is,  $\frac{1}{2}$  for each pair of equal values,  $\frac{3}{2}$  for each triple, etc. ■

### PROBLEM SET 25.8

- What would change in Example 1 had we observed only 5 positive values? Only 4?
- Test  $\tilde{\mu} = 0$  against  $\tilde{\mu} > 0$ , using 1, -1, 1, 3, -8, 6, 0 (deviations of the azimuth [multiples of 0.01 radian] in some revolution of a satellite).
- Are oil filters of type  $A$  better than type  $B$  filters if in 11 trials,  $A$  gave cleaner oil than  $B$  in 7 cases,  $B$  gave cleaner oil than  $A$  in 1 case, whereas in 3 of the trials the results for  $A$  and  $B$  were practically the same?
- Does a process of producing stainless steel pipes of length 20 ft for nuclear reactors need adjustment if, in a sample, 4 pipes have the exact length and 15 are shorter and 3 longer than 20 ft? Use the normal approximation of the binomial distribution.
- Do the computations in Prob. 4 without the use of the DeMoivre–Laplace limit theorem in Sec. 24.8.
- Thirty new employees were grouped into 15 pairs of similar intelligence and experience and were then instructed in data processing by an old method ( $A$ ) applied to one (randomly selected) person of each pair, and by a new presumably better method ( $B$ ) applied to the other person of each pair. Test for equality of methods against the alternative that ( $B$ ) is better than ( $A$ ), using the following scores obtained after the end of the training period.

$A$	60	70	80	85	75	40	70	45	95	80	90	60	80	75	65
$B$	65	85	85	80	95	65	100	60	90	85	100	75	90	60	80

- Assuming normality, solve Prob. 6 by a suitable test from Sec. 25.4.
- In a clinical experiment, each of 10 patients were given two different sedatives  $A$  and  $B$ . The following table shows the effect (increase of sleeping time, measured in hours). Using the sign test, find out whether the difference is significant.

$A$	1.9	0.8	1.1	0.1	-0.1	4.4	5.5	1.6	4.6	3.4
$B$	0.7	-1.6	-0.2	-1.2	-0.1	3.4	3.7	0.8	0.0	2.0
Difference	1.2	2.4	1.3	1.3	0.0	1.0	1.8	0.8	4.6	1.4

- Assuming that the populations corresponding to the samples in Prob. 8 are normal, apply a suitable test for the normal distribution.
- Test whether a thermostatic switch is properly set to 50°C against the alternative that its setting is too low. Use a sample of 9 values, 8 of which are less than 50°C and 1 is greater.
- How would you proceed in the sign test if the hypothesis is  $\tilde{\mu} = \tilde{\mu}_0$  (any number) instead of  $\tilde{\mu} = 0$ ?
- Test the hypothesis that, for a certain type of voltmeter, readings are independent of temperature  $T$  [°C] against the alternative that they tend to increase with  $T$ . Use a sample of values obtained by applying a constant voltage:

Temperature $T$ [°C]	10	20	30	40	50
Reading $V$ [volts]	99.5	101.1	100.4	100.8	101.6

- Does the amount of fertilizer increase the yield of wheat  $X$  [kg/plot]? Use a sample of values ordered according to increasing amounts of fertilizer:

33.4 35.3 31.6 35.0 36.1 37.6 36.5 38.7.

- Apply the test explained in Example 2 to the following data ( $x$  = diastolic blood pressure [mm Hg],  $y$  = weight of heart [in grams] of 10 patients who died of cerebral hemorrhage).

$x$	121	120	95	123	140	112	92	100	102	91
$y$	521	465	352	455	490	388	301	395	375	418

- Does an increase in temperature cause an increase of the yield of a chemical reaction from which the following sample was taken?

Temperature [°C]	10	20	30	40	60	80
Yield [kg/min]	0.6	1.1	0.9	1.6	1.2	2.0

## 25.9 Regression. Fitting Straight Lines. Correlation

So far we were concerned with random experiments in which we observed a single quantity (random variable) and got samples whose values were single numbers. In this section we discuss experiments in which we observe or measure two quantities simultaneously, so that we get samples of *pairs* of values  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ . Most applications involve one of two kinds of experiments, as follows.

1. In **regression analysis** one of the two variables, call it  $x$ , can be regarded as an ordinary variable because we can measure it without substantial error or we can even give it values we want.  $x$  is called the **independent variable**, or sometimes the **controlled variable** because we can control it (set it at values we choose). The other variable,  $Y$ , is a random variable, and we are interested in the dependence of  $Y$  on  $x$ . Typical examples are the dependence of the blood pressure  $Y$  on the age  $x$  of a person or, as we shall now say, the regression of  $Y$  on  $x$ , the regression of the gain of weight  $Y$  of certain animals on the daily ration of food  $x$ , the regression of the heat conductivity  $Y$  of cork on the specific weight  $x$  of the cork, etc.
2. In **correlation analysis** both quantities are random variables and we are interested in relations between them. Examples are the relation (one says “correlation”) between wear  $X$  and wear  $Y$  of the front tires of cars, between grades  $X$  and  $Y$  of students in mathematics and in physics, respectively, between the hardness  $X$  of steel plates in the center and the hardness  $Y$  near the edges of the plates, etc.

### Regression Analysis

In regression analysis the dependence of  $Y$  on  $x$  is a dependence of the mean  $\mu$  of  $Y$  on  $x$ , so that  $\mu = \mu(x)$  is a function in the ordinary sense. The curve of  $\mu(x)$  is called the **regression curve** of  $Y$  on  $x$ .

In this section we discuss the simplest case, namely, that of a straight **regression line**

$$(1) \quad \mu(x) = \kappa_0 + \kappa_1 x.$$

Then we may want to graph the sample values as  $n$  points in the  $xY$ -plane, fit a straight line through them, and use it for estimating  $\mu(x)$  at values of  $x$  that interest us, so that we know what values of  $Y$  we can expect for those  $x$ . Fitting that line by eye would not be good because it would be subjective; that is, different persons' results would come out differently, particularly if the points are scattered. So we need a mathematical method that gives a unique result depending only on the  $n$  points. A widely used procedure is the method of least squares by Gauss and Legendre. For our task we may formulate it as follows.

#### Least Squares Principle

*The straight line should be fitted through the given points so that the sum of the squares of the distances of those points from the straight line is minimum, where the distance is measured in the vertical direction (the  $y$ -direction). (Formulas below.)*

To get uniqueness of the straight line, we need some extra condition. To see this, take the sample  $(0, 1), (0, -1)$ . Then all the lines  $y = k_1x$  with any  $k_1$  satisfy the principle. (Can you see it?) The following assumption will imply uniqueness, as we shall find out.

**General Assumption (A1)**

*The  $x$ -values  $x_1, \dots, x_n$  in our sample  $(x_1, y_1), \dots, (x_n, y_n)$  are not all equal.*

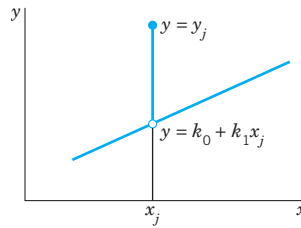
From a given sample  $(x_1, y_1), \dots, (x_n, y_n)$  we shall now determine a straight line by least squares. We write the line as

$$(2) \quad y = k_0 + k_1x$$

and call it the **sample regression line** because it will be the counterpart of the population regression line (1).

Now a sample point  $(x_j, y_j)$  has the vertical distance (distance measured in the  $y$ -direction) from (2) given by

$$|y_j - (k_0 + k_1x_j)| \quad (\text{see Fig. 543}).$$



**Fig. 543.** Vertical distance of a point  $(x_j, y_j)$  from a straight line  $y = k_0 + k_1x$

Hence the sum of the squares of these distances is

$$(3) \quad q = \sum_{j=1}^n (y_j - k_0 - k_1x_j)^2.$$

In the method of least squares we now have to determine  $k_0$  and  $k_1$  such that  $q$  is minimum. From calculus we know that a necessary condition for this is

$$(4) \quad \frac{\partial q}{\partial k_0} = 0 \quad \text{and} \quad \frac{\partial q}{\partial k_1} = 0.$$

We shall see that from this condition we obtain for the sample regression line the formula

$$(5) \quad y - \bar{y} = k_1(x - \bar{x}).$$

Here  $\bar{x}$  and  $\bar{y}$  are the means of the  $x$ - and the  $y$ -values in our sample, that is,

$$(6) \quad \begin{aligned} (a) \quad \bar{x} &= \frac{1}{n} (x_1 + \cdots + x_n) \\ (b) \quad \bar{y} &= \frac{1}{n} (y_1 + \cdots + y_n). \end{aligned}$$

The slope  $k_1$  in (5) is called the **regression coefficient** of the sample and is given by

$$(7) \quad k_1 = \frac{s_{xy}}{s_x^2}.$$

Here the “**sample covariance**”  $s_{xy}$  is

$$(8) \quad s_{xy} = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y}) = \frac{1}{n-1} \left[ \sum_{j=1}^n x_j y_j - \frac{1}{n} \left( \sum_{i=1}^n x_i \right) \left( \sum_{j=1}^n y_j \right) \right]$$

and  $s_x^2$  is given by

$$(9a) \quad s_x^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2 = \frac{1}{n-1} \left[ \sum_{j=1}^n x_j^2 - \frac{1}{n} \left( \sum_{j=1}^n x_j \right)^2 \right].$$

From (5) we see that the sample regression line passes through the point  $(\bar{x}, \bar{y})$ , by which it is determined, together with the regression coefficient (7). We may call  $s_x^2$  the *variance* of the  $x$ -values, but we should keep in mind that  $x$  is an ordinary variable, not a random variable.

We shall soon also need

$$(9b) \quad s_y^2 = \frac{1}{n-1} \sum_{j=1}^n (y_j - \bar{y})^2 = \frac{1}{n-1} \left[ \sum_{j=1}^n y_j^2 - \frac{1}{n} \left( \sum_{j=1}^n y_j \right)^2 \right].$$

**Derivation of (5) and (7).** Differentiating (3) and using (4), we first obtain

$$\begin{aligned} \frac{\partial q}{\partial k_0} &= -2 \sum (y_j - k_0 - k_1 x_j) = 0, \\ \frac{\partial q}{\partial k_1} &= -2 \sum x_j (y_j - k_0 - k_1 x_j) = 0 \end{aligned}$$

where we sum over  $j$  from 1 to  $n$ . We now divide by 2, write each of the two sums as three sums, and take the sums containing  $y_j$  and  $x_j y_j$  over to the right. Then we get the “**normal equations**”

$$(10) \quad \begin{aligned} k_0 n + k_1 \sum x_j &= \sum y_j \\ k_0 \sum x_j + k_1 \sum x_j^2 &= \sum x_j y_j. \end{aligned}$$

This is a linear system of two equations in the two unknowns  $k_0$  and  $k_1$ . Its coefficient determinant is [see (9)]

$$\begin{vmatrix} n & \sum x_j \\ \sum x_j & \sum x_j^2 \end{vmatrix} = n \sum x_j^2 - \left( \sum x_j \right)^2 = n(n-1)s_x^2 = n \sum (x_j - \bar{x})^2$$

and is not zero because of Assumption (A1). Hence the system has a unique solution. Dividing the first equation of (10) by  $n$  and using (6), we get  $k_0 = \bar{y} - k_1\bar{x}$ . Together with  $y = k_0 + k_1x$  in (2) this gives (5). To get (7), we solve the system (10) by Cramer's rule (Sec. 7.6) or elimination, finding

$$(11) \quad k_1 = \frac{n \sum x_j y_j - \sum x_j \sum y_j}{n(n-1)s_x^2}.$$

This gives (7)–(9) and completes the derivation. [The equality of the two expressions in (8) and in (9) may be shown by the student]. ■

### EXAMPLE 1 Regression Line

The decrease of volume  $y$  [%] of leather for certain fixed values of high pressure  $x$  [atmospheres] was measured. The results are shown in the first two columns of Table 25.11. Find the regression line of  $y$  on  $x$ .

**Solution.** We see that  $n = 4$  and obtain the values  $\bar{x} = 28000/4 = 7000$ ,  $\bar{y} = 19.0/4 = 4.75$ , and from (9) and (8)

**Table 25.11** Regression of the Decrease of Volume  $y$  [%] of Leather on the Pressure  $x$  [Atmospheres]

Given Values		Auxiliary Values	
$x_j$	$y_j$	$x_j^2$	$x_j y_j$
4000	2.3	16,000,000	9200
6000	4.1	36,000,000	24,600
8000	5.7	64,000,000	45,600
10,000	6.9	100,000,000	69,000
28,000	19.0	216,000,000	148,400

$$s_x^2 = \frac{1}{3} \left( 216,000,000 - \frac{28,000^2}{4} \right) = \frac{20,000,000}{3}$$

$$s_{xy} = \frac{1}{3} \left( 148,400 - \frac{28,000 \cdot 19}{4} \right) = \frac{15,400}{3}.$$

Hence  $k_1 = 15,400/20,000,000 = 0.00077$  from (7), and the regression line is

$$y - 4.75 = 0.00077(x - 7000) \quad \text{or} \quad y = 0.00077x - 0.64.$$

Note that  $y(0) = -0.64$ , which is physically meaningless, but typically indicates that a linear relation is merely an approximation valid on some restricted interval. ■



## Confidence Intervals in Regression Analysis

If we want to get confidence intervals, we have to make assumptions about the distribution of  $Y$  (which we have not made so far; least squares is a “geometric principle,” nowhere involving probabilities!). We assume normality and independence in sampling:

### Assumption (A2)

For each fixed  $x$  the random variable  $Y$  is normal with mean (1), that is,

$$(12) \quad \mu(x) = \kappa_0 + \kappa_1 x$$

and variance  $\sigma^2$  independent of  $x$ .

### Assumption (A3)

The  $n$  performances of the experiment by which we obtain a sample

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

are independent.

$\kappa_1$  in (12) is called the **regression coefficient** of the population because it can be shown that, under Assumptions (A1)–(A3), the maximum likelihood estimate of  $\kappa_1$  is the sample regression coefficient  $k_1$  given by (11).

Under Assumptions (A1)–(A3), we may now obtain a confidence interval for  $\kappa_1$ , as shown in Table 25.12.

**Table 25.12** Determination of a Confidence Interval for  $\kappa_1$  in (1) under Assumptions (A1)–(A3)

**Step 1.** Choose a confidence level  $\gamma$  (95%, 99%, or the like).

**Step 2.** Determine the solution  $c$  of the equation

$$(13) \quad F(c) = \frac{1}{2}(1 + \gamma)$$

from the table of the  $t$ -distribution with  $n - 2$  degrees of freedom (Table A9 in App. 5;  $n$  = sample size).

**Step 3.** Using a sample  $(x_1, y_1), \dots, (x_n, y_n)$ , compute  $(n - 1)s_x^2$  from (9a),  $(n - 1)s_{xy}$  from (8),  $k_1$  from (7),

$$(14) \quad (n - 1)s_y^2 = \sum_{j=1}^n y_j^2 - \frac{1}{n} \left( \sum_{j=1}^n y_j \right)^2$$

[as in (9b)], and

$$(15) \quad q_0 = (n - 1)(s_y^2 - k_1^2 s_x^2).$$

**Step 4.** Compute

$$K = c \sqrt{\frac{q_0}{(n - 2)(n - 1)s_x^2}}.$$

The confidence interval is

$$(16) \quad \text{CONF}_\gamma \{k_1 - K \leq \kappa_1 \leq k_1 + K\}.$$

**EXAMPLE 2** Confidence Interval for the Regression Coefficient

Using the sample in Table 25.11, determine a confidence interval for  $\kappa_1$  by the method in Table 25.12.

**Solution.** *Step 1.* We choose  $\gamma = 0.95$ .

*Step 2.* Equation (13) takes the form  $F(c) = 0.975$ , and Table A9 in App. 5 with  $n - 2 = 2$  degrees of freedom gives  $c = 4.30$ .

*Step 3.* From Example 1 we have  $3s_x^2 = 20,000,000$  and  $k_1 = 0.00077$ . From Table 25.11 we compute

$$\begin{aligned} 3s_y^2 &= 102.0 - \frac{19^2}{4} \\ &= 11.95. \\ q_0 &= 11.95 - 20,000,000 \cdot 0.00077^2 \\ &= 0.092. \end{aligned}$$

*Step 4.* We thus obtain

$$\begin{aligned} K &= 4.30 \sqrt{0.092 / (2 \cdot 20,000,000)} \\ &= 0.000206 \end{aligned}$$

and

$$\text{CONF}_{0.95} \{0.00056 \leq \kappa_1 \leq 0.00098\}.$$

**Correlation Analysis**

We shall now give an introduction to the basic facts in correlation analysis; for proofs see Ref. [G2] or [G8] in App. 1.

**Correlation analysis** is concerned with the relation between  $X$  and  $Y$  in a two-dimensional random variable  $(X, Y)$  (Sec. 24.9). A sample consists of  $n$  ordered pairs of values  $(x_1, y_1), \dots, (x_n, y_n)$ , as before. The interrelation between the  $x$  and  $y$  values in the sample is measured by the sample covariance  $s_{xy}$  in (8) or by the sample **correlation coefficient**

$$(17) \quad r = \frac{s_{xy}}{s_x s_y}$$

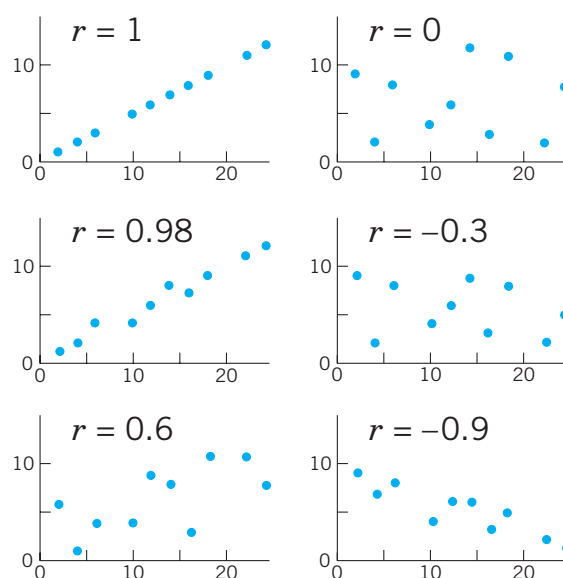
with  $s_x$  and  $s_y$  given in (9). Here  $r$  has the advantage that it does not change under a multiplication of the  $x$  and  $y$  values by a factor (in going from feet to inches, etc.).

**THEOREM 1****Sample Correlation Coefficient**

*The sample correlation coefficient  $r$  satisfies  $-1 \leq r \leq 1$ . In particular,  $r = \pm 1$  if and only if the sample values lie on a straight line. (See Fig. 544.)*

The theoretical counterpart of  $r$  is the **correlation coefficient**  $\rho$  of  $X$  and  $Y$ ,

$$(18) \quad \rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Fig. 544. Samples with various values of the correlation coefficient  $r$ 

where  $\mu_X = E(X)$ ,  $\mu_Y = E(Y)$ ,  $\sigma_X^2 = E([X - \mu_X]^2)$ ,  $\sigma_Y^2 = E([Y - \mu_Y]^2)$  (the means and variances of the marginal distributions of  $X$  and  $Y$ ; see Sec. 24.9), and  $\sigma_{XY}$  is the **covariance** of  $X$  and  $Y$  given by (see Sec. 24.9)

$$(19) \quad \sigma_{XY} = E([X - \mu_X][Y - \mu_Y]) = E(XY) - E(X)E(Y).$$

The analog of Theorem 1 is

### THEOREM 2

#### Correlation Coefficient

The correlation coefficient  $\rho$  satisfies  $-1 \leq \rho \leq 1$ . In particular,  $\rho = \pm 1$  if and only if  $X$  and  $Y$  are **linearly related**, that is,  $Y = \gamma X + \delta$ ,  $X = \gamma^* Y + \delta^*$ .

$X$  and  $Y$  are called **uncorrelated** if  $\rho = 0$ .

### THEOREM 3

#### Independence. Normal Distribution

- (a) Independent  $X$  and  $Y$  (see Sec. 24.9) are uncorrelated.
- (b) If  $(X, Y)$  is normal (see below), then uncorrelated  $X$  and  $Y$  are independent.

Here the two-dimensional normal distribution can be introduced by taking two independent standardized normal random variables  $X^*$ ,  $Y^*$ , whose joint distribution thus has the density

$$(20) \quad f^*(x^*, y^*) = \frac{1}{2\pi} e^{-(x^{*2} + y^{*2})/2}$$

(representing a surface of revolution over the  $x^*y^*$ -plane with a bell-shaped curve as cross section) and setting

$$\begin{aligned} X &= \mu_X + \sigma_X X^* \\ Y &= \mu_Y + \rho \sigma_Y X^* + \sqrt{1 - \rho^2} \sigma_Y Y^*. \end{aligned}$$

This gives the general **two-dimensional normal distribution** with the density

$$(21a) \quad f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-h(x,y)/2}$$

where

$$(21b) \quad h(x, y) = \frac{1}{1-\rho^2} \left[ \left( \frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \left( \frac{x-\mu_X}{\sigma_X} \right) \left( \frac{y-\mu_Y}{\sigma_Y} \right) + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 \right].$$

In Theorem 3(b), normality is important, as we can see from the following example.

### EXAMPLE 3 Uncorrelated But Dependent Random Variables

If  $X$  assumes  $-1, 0, 1$  with probability  $\frac{1}{3}$  and  $Y = X^2$ , then  $E(X) = 0$  and in (3)

$$\sigma_{XY} = E(XY) = E(X^3) = (-1)^3 \cdot \frac{1}{3} + 0^3 \cdot \frac{1}{3} + 1^3 \cdot \frac{1}{3} = 0,$$

so that  $\rho = 0$  and  $X$  and  $Y$  are uncorrelated. But they are certainly not independent since they are even functionally related. ■

## Test for the Correlation Coefficient $\rho$

Table 25.13 shows a test for  $\rho$  in the case of the two-dimensional normal distribution.  $t$  is an observed value of a random variable that has a  $t$ -distribution with  $n - 2$  degrees of freedom. This was shown by R. A. Fisher (*Biometrika* **10** (1915), 507–521).

**Table 25.13** Test of the Hypothesis  $\rho = 0$  Against the Alternative  $\rho > 0$  in the Case of the Two-Dimensional Normal Distribution

**Step 1.** Choose a significance level  $\alpha$  (5%, 1%, or the like).

**Step 2.** Determine the solution  $c$  of the equation

$$P(T \leq c) = 1 - \alpha$$

from the  $t$ -distribution (Table A9 in App. 5) with  $n - 2$  degrees of freedom.

**Step 3.** Compute  $r$  from (17), using a sample  $(x_1, y_1), \dots, (x_n, y_n)$ .

**Step 4.** Compute

$$t = r \left( \sqrt{\frac{n-2}{1-r^2}} \right).$$

If  $t \leq c$ , accept the hypothesis. If  $t > c$ , reject the hypothesis.

**EXAMPLE 4 Test for the Correlation Coefficient  $\rho$** 

Test the hypothesis  $\rho = 0$  (independence of  $X$  and  $Y$ , because of Theorem 3) against the alternative  $\rho > 0$ , using the data in the lower left corner of Fig. 544, where  $r = 0.6$  (manual soldering errors on 10 two-sided circuit boards done by 10 workers;  $x$  = front,  $y$  = back of the boards).

**Solution.** We choose  $\alpha = 5\%$ ; thus  $1 - \alpha = 95\%$ . Since  $n = 10$ ,  $n - 2 = 8$ , the table gives  $c = 1.86$ . Also,  $t = 0.6\sqrt{8/0.64} = 2.12 > c$ . We reject the hypothesis and assert that there is a **positive correlation**. A worker making few (many) errors on the front side also tends to make few (many) errors on the reverse side of the board. ■

**PROBLEM SET 25.9****1–10 SAMPLE REGRESSION LINE**

Find and graph the sample regression line of  $y$  on  $x$  and the given data as points on the same axes. Show the details of your work.

- (0, 1.0), (2, 2.1), (4, 2.9), (6, 3.6), (8, 5.2)
- (−2, 3.5), (1, 2.6), (3, 1.3), (5, 0.4)
- $x$  = Revolutions per minute,  $y$  = Power of a Diesel engine [hp]

$x$	400	500	600	700	750
$y$	5800	10,300	14,200	18,800	21,000

- $x$  = Deformation of a certain steel [mm],  $y$  = Brinell hardness [ $\text{kg/mm}^2$ ]

$x$	6	9	11	13	22	26	28	33	35
$y$	68	67	65	53	44	40	37	34	32

- $x$  = Brinell hardness,  $y$  = Tensile strength [in 1000 psi (pounds per square inch)] of steel with 0.45% C tempered for 1 hour

$x$	200	300	400	500
$y$	110	150	190	280

- Abrasion of quenched and tempered steel S620.**  $x$  = Sliding distance [km],  $y$  = Wear volume [ $\text{mm}^3$ ]

$x$	1.1	3.2	3.4	4.5	5.6
$y$	40	65	120	150	190

- Ohm's law (Sec. 2.9).**  $x$  = Voltage [V],  $y$  = Current [A]. Also find the resistance  $R$  [ $\Omega$ ].

$x$	40	40	80	80	110	110
$y$	5.1	4.8	0.0	10.3	13.0	12.7

- Hooke's law** (Sec. 2.4).  $x$  = Force [lb],  $y$  = Extension [in] of a spring. Also find the spring modulus.

$x$	2	4	6	8
$y$	4.1	7.8	12.3	15.8

- Thermal conductivity of water.**  $x$  = Temperature [ $^{\circ}\text{F}$ ],  $y$  = Conductivity [ $\text{Btu}/(\text{hr} \cdot \text{ft} \cdot ^{\circ}\text{F})$ ]. Also find  $y$  at room temperature  $66^{\circ}\text{F}$ .

$x$	32	50	100	150	212
$y$	0.337	0.345	0.365	0.380	0.395

- Stopping distance of a car.**  $x$  = Speed [mph],  $y$  = Stopping distance [ft]. Also find  $y$  at 35 mph.

$x$	30	40	50	60
$y$	160	240	330	435

- CAS EXPERIMENT. Moving Data.** Take a sample, for instance, that in Prob. 4, and investigate and graph the effect of changing  $y$ -values (a) for small  $x$ , (b) for large  $x$ , (c) in the middle of the sample.

**12–15 CONFIDENCE INTERVALS**

Find a 95% confidence interval for the regression coefficient  $\kappa_1$ , assuming (A2) and (A3) hold and using the sample.

- In Prob. 2
- In Prob. 3
- In Prob. 4
- $x$  = Humidity of air [%],  $y$  = Expansion of gelatin [%],

$x$	10	20	30	40
$y$	0.8	1.6	2.3	2.8

**CHAPTER 25 REVIEW QUESTIONS AND PROBLEMS**

- What is a sample? A population? Why do we sample in statistics?
- If we have several samples from the same population, do they have the same sample distribution function? The same mean and variance?
- Can we develop statistical methods without using probability theory? Apply the methods without using a sample?
- What is the idea of the maximum likelihood method? Why do we say “likelihood” rather than “probability”?

5. Couldn't we make the error of interval estimation zero simply by choosing the confidence level 1?
6. What is testing? Why do we test? What are the errors involved?
7. When did we use the  $t$ -distribution? The  $F$ -distribution?
8. What is the chi-square ( $\chi^2$ ) test? Give a sample example from memory.
9. What are one-sided and two-sided tests? Give typical examples.
10. How do we test in quality control? In acceptance sampling?
11. What is the power of a test? What could you perhaps do when it is low?
12. What is Gauss's least squares principle (which he found at age 18)?
13. What is the difference between regression and correlation?
14. Find the mean, variance, and standard deviation of the sample 21.0 21.6 19.9 19.6 15.6 20.6 22.1 22.2.
15. Assuming normality, find the maximum likelihood estimates of mean and variance from the sample in Prob. 14.
16. Determine a 95% confidence interval for the mean  $\mu$  of a normal population with variance  $\sigma^2 = 25$ , using a sample of size 500 with mean 22.
17. Determine a 99% confidence interval for the mean of a normal population, using the sample 32, 33, 32, 34, 35, 29, 29, 27.
18. Assuming normality, find a 95% confidence interval for the variance from the sample 145.3, 145.1, 145.4, 146.2.
19. Using a sample of 10 values with mean 14.5 from a normal population with variance  $\sigma^2 = 0.25$ , test the hypothesis  $\mu_0 = 15.0$  against the alternative  $\mu_1 = 14.5$  on the 5% level. Find the power.
20. Three specimens of high-quality concrete had compressive strength 357, 359, 413 [kg/cm<sup>2</sup>], and for three specimens of ordinary concrete the values were 346, 358, 302. Test for equality of the population means,  $\mu_1 = \mu_2$ , against the alternative  $\mu_1 > \mu_2$ . Assume normality and equality of variance. Choose  $\alpha = 5\%$ .
21. Assume the thickness  $X$  of washers to be normal with mean 2.75 mm and variance 0.00024 mm<sup>2</sup>. Set up a control chart for  $\mu$  and graph the means of the five samples (2.74, 2.76), (2.74, 2.74), (2.79, 2.81), (2.78, 2.76), (2.71, 2.75) on the chart.
22. The OC curve in acceptance sampling cannot have a strictly vertical portion. Why?
23. Find the risks in the sampling plan with  $n = 6$  and  $c = 0$ , assuming that the AQL is  $\theta_0 = 1\%$  and the RQL is  $\theta_1 = 15\%$ . How do the risks change if we increase  $n$ ?
24. Does a process of producing plastic rods of length  $\tilde{\mu} = 2$  meters need adjustment if in a sample, 2 rods have the exact length and 15 are shorter and 3 longer than 2 meters? (Use the sign test.)
25. Find the regression line of  $y$  on  $x$  for the data  $(x, y) = (0, 4), (2, 0), (4, -5), (6, -9), (8, -10)$ .

## SUMMARY OF CHAPTER 25

### Mathematical Statistics

We recall from Chap. 24 that, with an experiment in which we observe some quantity (number of defectives, height of persons, etc.), there is associated a random variable  $X$  whose probability distribution is given by a distribution function

$$(1) \quad F(x) = P(X \leq x) \quad (\text{Sec. 24.5})$$

which for each  $x$  gives the probability that  $X$  assumes any value not exceeding  $x$ .

In statistics we take random samples  $x_1, \dots, x_n$  of size  $n$  by performing that experiment  $n$  times (Sec. 25.1) and draw conclusions from properties of samples about properties of the distribution of the corresponding  $X$ . We do this by calculating *point estimates* or *confidence intervals* or by performing a *test* for **parameters** ( $\mu$  and  $\sigma^2$  in the normal distribution,  $p$  in the binomial distribution, etc.) or by a test for distribution functions.

A **point estimate** (Sec. 25.2) is an approximate value for a parameter in the distribution of  $X$  obtained from a sample. Notably, the **sample mean** (Sec. 25.1)

$$(2) \quad \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j = \frac{1}{n} (x_1 + \cdots + x_n)$$

is an estimate of the mean  $\mu$  of  $X$ , and the **sample variance** (Sec. 25.1)

$$(3) \quad s^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2 = \frac{1}{n-1} [(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2]$$

is an estimate of the variance  $\sigma^2$  of  $X$ . Point estimation can be done by the basic **maximum likelihood method** (Sec. 25.2).

**Confidence intervals** (Sec. 25.3) are intervals  $\theta_1 \leq \theta \leq \theta_2$  with endpoints calculated from a sample such that, with a high probability  $\gamma$ , we obtain an interval that contains the unknown true value of the parameter  $\theta$  in the distribution of  $X$ . Here,  $\gamma$  is chosen at the beginning, usually 95% or 99%. We denote such an interval by  $\text{CONF}_\gamma \{ \theta_1 \leq \theta \leq \theta_2 \}$ .

In a **test** for a parameter we test a *hypothesis*  $\theta = \theta_0$  against an *alternative*  $\theta = \theta_1$  and then, on the basis of a sample, accept the hypothesis, or we reject it in favor of the alternative (Sec. 25.4). Like any conclusion about  $X$  from samples, this may involve errors leading to a false decision. There is a small probability  $\alpha$  (which we can choose, 5% or 1%, for instance) that we reject a true hypothesis, and there is a probability  $\beta$  (which we can compute and decrease by taking larger samples) that we accept a false hypothesis.  $\alpha$  is called the **significance level** and  $1 - \beta$  the **power** of the test. Among many other engineering applications, testing is used in **quality control** (Sec. 25.5) and **acceptance sampling** (Sec. 25.6).

If not merely a parameter but the kind of distribution of  $X$  is unknown, we can use the **chi-square test** (Sec. 25.7) for testing the hypothesis that some function  $F(x)$  is the unknown distribution function of  $X$ . This is done by determining the discrepancy between  $F(x)$  and the distribution function  $\tilde{F}(x)$  of a given sample.

“Distribution-free” or **nonparametric tests** are tests that apply to any distribution, since they are based on combinatorial ideas. These tests are usually very simple. Two of them are discussed in Sec. 25.8.

The last section deals with samples of **pairs of values**, which arise in an experiment when we simultaneously observe two quantities. In **regression analysis**, one of the quantities,  $x$ , is an ordinary variable and the other,  $Y$ , is a random variable whose mean  $\mu$  depends on  $x$ , say,  $\mu(x) = \kappa_0 + \kappa_1 x$ . In **correlation analysis** the relation between  $X$  and  $Y$  in a two-dimensional random variable  $(X, Y)$  is investigated, notably in terms of the **correlation coefficient**  $\rho$ .







# APPENDIX 5

## Tables

**For Tables of Laplace Transforms see Secs. 6.8 and 6.9.**

**For Tables of Fourier Transforms see Sec. 11.10.**

*If you have a Computer Algebra System (CAS), you may not need the present tables, but you may still find them convenient from time to time.*

**Table A1 Bessel Functions**

For more extensive tables see Ref. [GenRef1] in App. 1.

$x$	$J_0(x)$	$J_1(x)$	$x$	$J_0(x)$	$J_1(x)$	$x$	$J_0(x)$	$J_1(x)$
0.0	1.0000	0.0000	3.0	-0.2601	0.3391	6.0	0.1506	-0.2767
0.1	0.9975	0.0499	3.1	-0.2921	0.3009	6.1	0.1773	-0.2559
0.2	0.9900	0.0995	3.2	-0.3202	0.2613	6.2	0.2017	-0.2329
0.3	0.9776	0.1483	3.3	-0.3443	0.2207	6.3	0.2238	-0.2081
0.4	0.9604	0.1960	3.4	-0.3643	0.1792	6.4	0.2433	-0.1816
0.5	0.9385	0.2423	3.5	-0.3801	0.1374	6.5	0.2601	-0.1538
0.6	0.9120	0.2867	3.6	-0.3918	0.0955	6.6	0.2740	-0.1250
0.7	0.8812	0.3290	3.7	-0.3992	0.0538	6.7	0.2851	-0.0953
0.8	0.8463	0.3688	3.8	-0.4026	0.0128	6.8	0.2931	-0.0652
0.9	0.8075	0.4059	3.9	-0.4018	-0.0272	6.9	0.2981	-0.0349
1.0	0.7652	0.4401	4.0	-0.3971	-0.0660	7.0	0.3001	-0.0047
1.1	0.7196	0.4709	4.1	-0.3887	-0.1033	7.1	0.2991	0.0252
1.2	0.6711	0.4983	4.2	-0.3766	-0.1386	7.2	0.2951	0.0543
1.3	0.6201	0.5220	4.3	-0.3610	-0.1719	7.3	0.2882	0.0826
1.4	0.5669	0.5419	4.4	-0.3423	-0.2028	7.4	0.2786	0.1096
1.5	0.5118	0.5579	4.5	-0.3205	-0.2311	7.5	0.2663	0.1352
1.6	0.4554	0.5699	4.6	-0.2961	-0.2566	7.6	0.2516	0.1592
1.7	0.3980	0.5778	4.7	-0.2693	-0.2791	7.7	0.2346	0.1813
1.8	0.3400	0.5815	4.8	-0.2404	-0.2985	7.8	0.2154	0.2014
1.9	0.2818	0.5812	4.9	-0.2097	-0.3147	7.9	0.1944	0.2192
2.0	0.2239	0.5767	5.0	-0.1776	-0.3276	8.0	0.1717	0.2346
2.1	0.1666	0.5683	5.1	-0.1443	-0.3371	8.1	0.1475	0.2476
2.2	0.1104	0.5560	5.2	-0.1103	-0.3432	8.2	0.1222	0.2580
2.3	0.0555	0.5399	5.3	-0.0758	-0.3460	8.3	0.0960	0.2657
2.4	0.0025	0.5202	5.4	-0.0412	-0.3453	8.4	0.0692	0.2708
2.5	-0.0484	0.4971	5.5	-0.0068	-0.3414	8.5	0.0419	0.2731
2.6	-0.0968	0.4708	5.6	0.0270	-0.3343	8.6	0.0146	0.2728
2.7	-0.1424	0.4416	5.7	0.0599	-0.3241	8.7	-0.0125	0.2697
2.8	-0.1850	0.4097	5.8	0.0917	-0.3110	8.8	-0.0392	0.2641
2.9	-0.2243	0.3754	5.9	0.1220	-0.2951	8.9	-0.0653	0.2559

$J_0(x) = 0$  for  $x = 2.40483, 5.52008, 8.65373, 11.7915, 14.9309, 18.0711, 21.2116, 24.3525, 27.4935, 30.6346$

$J_1(x) = 0$  for  $x = 3.83171, 7.01559, 10.1735, 13.3237, 16.4706, 19.6159, 22.7601, 25.9037, 29.0468, 32.1897$

**Table A1** (continued)

$x$	$Y_0(x)$	$Y_1(x)$	$x$	$Y_0(x)$	$Y_1(x)$	$x$	$Y_0(x)$	$Y_1(x)$
0.0	$(-\infty)$	$(-\infty)$	2.5	0.498	0.146	5.0	-0.309	0.148
0.5	-0.445	-1.471	3.0	0.377	0.325	5.5	-0.339	-0.024
1.0	0.088	-0.781	3.5	0.189	0.410	6.0	-0.288	-0.175
1.5	0.382	-0.412	4.0	-0.017	0.398	6.5	-0.173	-0.274
2.0	0.510	-0.107	4.5	-0.195	0.301	7.0	-0.026	-0.303

**Table A2** Gamma Function [see (24) in App. A3.1]

$\alpha$	$\Gamma(\alpha)$	$\alpha$	$\Gamma(\alpha)$	$\alpha$	$\Gamma(\alpha)$	$\alpha$	$\Gamma(\alpha)$	$\alpha$	$\Gamma(\alpha)$
1.00	1.000 000	1.20	0.918 169	1.40	0.887 264	1.60	0.893 515	1.80	0.931 384
1.02	0.988 844	1.22	0.913 106	1.42	0.886 356	1.62	0.895 924	1.82	0.936 845
1.04	0.978 438	1.24	0.908 521	1.44	0.885 805	1.64	0.898 642	1.84	0.942 612
1.06	0.968 744	1.26	0.904 397	1.46	0.885 604	1.66	0.901 668	1.86	0.948 687
1.08	0.959 725	1.28	0.900 718	1.48	0.885 747	1.68	0.905 001	1.88	0.955 071
1.10	0.951 351	1.30	0.897 471	1.50	0.886 227	1.70	0.908 639	1.90	0.961 766
1.12	0.943 590	1.32	0.894 640	1.52	0.887 039	1.72	0.912 581	1.92	0.968 774
1.14	0.936 416	1.34	0.892 216	1.54	0.888 178	1.74	0.916 826	1.94	0.976 099
1.16	0.929 803	1.36	0.890 185	1.56	0.889 639	1.76	0.921 375	1.96	0.983 743
1.18	0.923 728	1.38	0.888 537	1.58	0.891 420	1.78	0.926 227	1.98	0.991 708
1.20	0.918 169	1.40	0.887 264	1.60	0.893 515	1.80	0.931 384	2.00	1.000 000

**Table A3** Factorial Function and Its Logarithm with Base 10

$n$	$n!$	$\log(n!)$	$n$	$n!$	$\log(n!)$	$n$	$n!$	$\log(n!)$
1	1	0.000 000	6	720	2.857 332	11	39 916 800	7.601 156
2	2	0.301 030	7	5 040	3.702 431	12	479 001 600	8.680 337
3	6	0.778 151	8	40 320	4.605 521	13	6 227 020 800	9.794 280
4	24	1.380 211	9	362 880	5.559 763	14	87 178 291 200	10.940 408
5	120	2.079 181	10	3 628 800	6.559 763	15	1 307 674 368 000	12.116 500

**Table A4** Error Function, Sine and Cosine Integrals [see (35), (40), (42) in App. A3.1]

$x$	$\operatorname{erf} x$	$\operatorname{Si}(x)$	$\operatorname{ci}(x)$	$x$	$\operatorname{erf} x$	$\operatorname{Si}(x)$	$\operatorname{ci}(x)$
0.0	0.0000	0.0000	$\infty$	2.0	0.9953	1.6054	-0.4230
0.2	0.2227	0.1996	1.0422	2.2	0.9981	1.6876	-0.3751
0.4	0.4284	0.3965	0.3788	2.4	0.9993	1.7525	-0.3173
0.6	0.6039	0.5881	0.0223	2.6	0.9998	1.8004	-0.2533
0.8	0.7421	0.7721	-0.1983	2.8	0.9999	1.8321	-0.1865
1.0	0.8427	0.9461	-0.3374	3.0	1.0000	1.8487	-0.1196
1.2	0.9103	1.1080	-0.4205	3.2	1.0000	1.8514	-0.0553
1.4	0.9523	1.2562	-0.4620	3.4	1.0000	1.8419	0.0045
1.6	0.9763	1.3892	-0.4717	3.6	1.0000	1.8219	0.0580
1.8	0.9891	1.5058	-0.4568	3.8	1.0000	1.7934	0.1038
2.0	0.9953	1.6054	-0.4230	4.0	1.0000	1.7582	0.1410

**Table A5 Binomial Distribution**Probability function  $f(x)$  [see (2), Sec. 24.7] and distribution function  $F(x)$ 

$n$	$x$	$p = 0.1$		$p = 0.2$		$p = 0.3$		$p = 0.4$		$p = 0.5$	
		$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
1	0	0.	0.9000	0.	0.8000	0.	0.7000	0.	0.6000	0.	0.5000
	1	9000	1.0000	8000	1.0000	7000	1.0000	6000	1.0000	5000	1.0000
2	0	8100	0.8100	6400	0.6400	4900	0.4900	3600	0.3600	2500	0.2500
	1	1800	0.9900	3200	0.9600	4200	0.9100	4800	0.8400	5000	0.7500
	2	0100	1.0000	0400	1.0000	0900	1.0000	1600	1.0000	2500	1.0000
3	0	7290	0.7290	5120	0.5120	3430	0.3430	2160	0.2160	1250	0.1250
	1	2430	0.9720	3840	0.8960	4410	0.7840	4320	0.6480	3750	0.5000
	2	0270	0.9990	0960	0.9920	1890	0.9730	2880	0.9360	3750	0.8750
	3	0010	1.0000	0080	1.0000	0270	1.0000	0640	1.0000	1250	1.0000
4	0	6561	0.6561	4096	0.4096	2401	0.2401	1296	0.1296	0625	0.0625
	1	2916	0.9477	4096	0.8192	4116	0.6517	3456	0.4752	2500	0.3125
	2	0486	0.9963	1536	0.9728	2646	0.9163	3456	0.8208	3750	0.6875
	3	0036	0.9999	0256	0.9984	0756	0.9919	1536	0.9744	2500	0.9375
	4	0001	1.0000	0016	1.0000	0081	1.0000	0256	1.0000	0625	1.0000
5	0	5905	0.5905	3277	0.3277	1681	0.1681	0778	0.0778	0313	0.0313
	1	3281	0.9185	4096	0.7373	3602	0.5282	2592	0.3370	1563	0.1875
	2	0729	0.9914	2048	0.9421	3087	0.8369	3456	0.6826	3125	0.5000
	3	0081	0.9995	0512	0.9933	1323	0.9692	2304	0.9130	3125	0.8125
	4	0005	1.0000	0064	0.9997	0284	0.9976	0768	0.9898	1563	0.9688
	5	0000	1.0000	0003	1.0000	0024	1.0000	0102	1.0000	0313	1.0000
6	0	5314	0.5314	2621	0.2621	1176	0.1176	0467	0.0467	0156	0.0156
	1	3543	0.8857	3932	0.6554	3025	0.4202	1866	0.2333	0938	0.1094
	2	0984	0.9841	2458	0.9011	3241	0.7443	3110	0.5443	2344	0.3438
	3	0146	0.9987	0819	0.9830	1852	0.9295	2765	0.8208	3125	0.6563
	4	0012	0.9999	0154	0.9984	0595	0.9891	1382	0.9590	2344	0.8906
	5	0001	1.0000	0015	0.9999	0102	0.9993	0369	0.9959	0938	0.9844
	6	0000	1.0000	0001	1.0000	0007	1.0000	0041	1.0000	0156	1.0000
7	0	4783	0.4783	2097	0.2097	0824	0.0824	0280	0.0280	0078	0.0078
	1	3720	0.8503	3670	0.5767	2471	0.3294	1306	0.1586	0547	0.0625
	2	1240	0.9743	2753	0.8520	3177	0.6471	2613	0.4199	1641	0.2266
	3	0230	0.9973	1147	0.9667	2269	0.8740	2903	0.7102	2734	0.5000
	4	0026	0.9998	0287	0.9953	0972	0.9712	1935	0.9037	2734	0.7734
	5	0002	1.0000	0043	0.9996	0250	0.9962	0774	0.9812	1641	0.9375
	6	0000	1.0000	0004	1.0000	0036	0.9998	0172	0.9984	0547	0.9922
	7	0000	1.0000	0000	1.0000	0002	1.0000	0016	1.0000	0078	1.0000
8	0	4305	0.4305	1678	0.1678	0576	0.0576	0168	0.0168	0039	0.0039
	1	3826	0.8131	3355	0.5033	1977	0.2553	0896	0.1064	0313	0.0352
	2	1488	0.9619	2936	0.7969	2965	0.5518	2090	0.3154	1094	0.1445
	3	0331	0.9950	1468	0.9437	2541	0.8059	2787	0.5941	2188	0.3633
	4	0046	0.9996	0459	0.9896	1361	0.9420	2322	0.8263	2734	0.6367
	5	0004	1.0000	0092	0.9988	0467	0.9887	1239	0.9502	2188	0.8555
	6	0000	1.0000	0011	0.9999	0100	0.9987	0413	0.9915	1094	0.9648
	7	0000	1.0000	0001	1.0000	0012	0.9999	0079	0.9993	0313	0.9961
	8	0000	1.0000	0000	1.0000	0001	1.0000	0007	1.0000	0039	1.0000

Probability function  $f(x)$  [see (5), Sec. 24.7] and distribution function  $F(x)$

[illegible]

**Table A7 Normal Distribution**Values of the distribution function  $\Phi(z)$  [see (3), Sec. 24.8].  $\Phi(-z) = 1 - \Phi(z)$ 

$z$	$\Phi(z)$	$z$	$\Phi(z)$	$z$	$\Phi(z)$	$z$	$\Phi(z)$	$z$	$\Phi(z)$	$z$	$\Phi(z)$
	<b>0.</b>		<b>0.</b>		<b>0.</b>		<b>0.</b>		<b>0.</b>		<b>0.</b>
0.01	5040	0.51	6950	1.01	8438	1.51	9345	2.01	9778	2.51	9940
0.02	5080	0.52	6985	1.02	8461	1.52	9357	2.02	9783	2.52	9941
0.03	5120	0.53	7019	1.03	8485	1.53	9370	2.03	9788	2.53	9943
0.04	5160	0.54	7054	1.04	8508	1.54	9382	2.04	9793	2.54	9945
0.05	5199	0.55	7088	1.05	8531	1.55	9394	2.05	9798	2.55	9946
0.06	5239	0.56	7123	1.06	8554	1.56	9406	2.06	9803	2.56	9948
0.07	5279	0.57	7157	1.07	8577	1.57	9418	2.07	9808	2.57	9949
0.08	5319	0.58	7190	1.08	8599	1.58	9429	2.08	9812	2.58	9951
0.09	5359	0.59	7224	1.09	8621	1.59	9441	2.09	9817	2.59	9952
0.10	5398	0.60	7257	1.10	8643	1.60	9452	2.10	9821	2.60	9953
0.11	5438	0.61	7291	1.11	8665	1.61	9463	2.11	9826	2.61	9955
0.12	5478	0.62	7324	1.12	8686	1.62	9474	2.12	9830	2.62	9956
0.13	5517	0.63	7357	1.13	8708	1.63	9484	2.13	9834	2.63	9957
0.14	5557	0.64	7389	1.14	8729	1.64	9495	2.14	9838	2.64	9959
0.15	5596	0.65	7422	1.15	8749	1.65	9505	2.15	9842	2.65	9960
0.16	5636	0.66	7454	1.16	8770	1.66	9515	2.16	9846	2.66	9961
0.17	5675	0.67	7486	1.17	8790	1.67	9525	2.17	9850	2.67	9962
0.18	5714	0.68	7517	1.18	8810	1.68	9535	2.18	9854	2.68	9963
0.19	5753	0.69	7549	1.19	8830	1.69	9545	2.19	9857	2.69	9964
0.20	5793	0.70	7580	1.20	8849	1.70	9554	2.20	9861	2.70	9965
0.21	5832	0.71	7611	1.21	8869	1.71	9564	2.21	9864	2.71	9966
0.22	5871	0.72	7642	1.22	8888	1.72	9573	2.22	9868	2.72	9967
0.23	5910	0.73	7673	1.23	8907	1.73	9582	2.23	9871	2.73	9968
0.24	5948	0.74	7704	1.24	8925	1.74	9591	2.24	9875	2.74	9969
0.25	5987	0.75	7734	1.25	8944	1.75	9599	2.25	9878	2.75	9970
0.26	6026	0.76	7764	1.26	8962	1.76	9608	2.26	9881	2.76	9971
0.27	6064	0.77	7794	1.27	8980	1.77	9616	2.27	9884	2.77	9972
0.28	6103	0.78	7823	1.28	8997	1.78	9625	2.28	9887	2.78	9973
0.29	6141	0.79	7852	1.29	9015	1.79	9633	2.29	9890	2.79	9974
0.30	6179	0.80	7881	1.30	9032	1.80	9641	2.30	9893	2.80	9974
0.31	6217	0.81	7910	1.31	9049	1.81	9649	2.31	9896	2.81	9975
0.32	6255	0.82	7939	1.32	9066	1.82	9656	2.32	9898	2.82	9976
0.33	6293	0.83	7967	1.33	9082	1.83	9664	2.33	9901	2.83	9977
0.34	6331	0.84	7995	1.34	9099	1.84	9671	2.34	9904	2.84	9977
0.35	6368	0.85	8023	1.35	9115	1.85	9678	2.35	9906	2.85	9978
0.36	6406	0.86	8051	1.36	9131	1.86	9686	2.36	9909	2.86	9979
0.37	6443	0.87	8078	1.37	9147	1.87	9693	2.37	9911	2.87	9979
0.38	6480	0.88	8106	1.38	9162	1.88	9699	2.38	9913	2.88	9980
0.39	6517	0.89	8133	1.39	9177	1.89	9706	2.39	9916	2.89	9981
0.40	6554	0.90	8159	1.40	9192	1.90	9713	2.40	9918	2.90	9981
0.41	6591	0.91	8186	1.41	9207	1.91	9719	2.41	9920	2.91	9982
0.42	6628	0.92	8212	1.42	9222	1.92	9726	2.42	9922	2.92	9982
0.43	6664	0.93	8238	1.43	9236	1.93	9732	2.43	9925	2.93	9983
0.44	6700	0.94	8264	1.44	9251	1.94	9738	2.44	9927	2.94	9984
0.45	6736	0.95	8289	1.45	9265	1.95	9744	2.45	9929	2.95	9984
0.46	6772	0.96	8315	1.46	9279	1.96	9750	2.46	9931	2.96	9985
0.47	6808	0.97	8340	1.47	9292	1.97	9756	2.47	9932	2.97	9985
0.48	6844	0.98	8365	1.48	9306	1.98	9761	2.48	9934	2.98	9986
0.49	6879	0.99	8389	1.49	9319	1.99	9767	2.49	9936	2.99	9986
0.50	6915	1.00	8413	1.50	9332	2.00	9772	2.50	9938	3.00	9987

**Table A8 Normal Distribution**

Values of  $z$  for given values of  $\Phi(z)$  [see (3), Sec. 24.8] and  $D(z) = \Phi(z) - \Phi(-z)$   
 Example:  $z = 0.279$  if  $\Phi(z) = 61\%$ ;  $z = 0.860$  if  $D(z) = 61\%$ .

%	$z(\Phi)$	$z(D)$	%	$z(\Phi)$	$z(D)$	%	$z(\Phi)$	$z(D)$
1	-2.326	0.013	41	-0.228	0.539	81	0.878	1.311
2	-2.054	0.025	42	-0.202	0.553	82	0.915	1.341
3	-1.881	0.038	43	-0.176	0.568	83	0.954	1.372
4	-1.751	0.050	44	-0.151	0.583	84	0.994	1.405
5	-1.645	0.063	45	-0.126	0.598	85	1.036	1.440
6	-1.555	0.075	46	-0.100	0.613	86	1.080	1.476
7	-1.476	0.088	47	-0.075	0.628	87	1.126	1.514
8	-1.405	0.100	48	-0.050	0.643	88	1.175	1.555
9	-1.341	0.113	49	-0.025	0.659	89	1.227	1.598
10	-1.282	0.126	50	0.000	0.674	90	1.282	1.645
11	-1.227	0.138	51	0.025	0.690	91	1.341	1.695
12	-1.175	0.151	52	0.050	0.706	92	1.405	1.751
13	-1.126	0.164	53	0.075	0.722	93	1.476	1.812
14	-1.080	0.176	54	0.100	0.739	94	1.555	1.881
15	-1.036	0.189	55	0.126	0.755	95	1.645	1.960
16	-0.994	0.202	56	0.151	0.772	96	1.751	2.054
17	-0.954	0.215	57	0.176	0.789	97	1.881	2.170
18	-0.915	0.228	58	0.202	0.806	97.5	1.960	2.241
19	-0.878	0.240	59	0.228	0.824	98	2.054	2.326
20	-0.842	0.253	60	0.253	0.842	99	2.326	2.576
21	-0.806	0.266	61	0.279	0.860	99.1	2.366	2.612
22	-0.772	0.279	62	0.305	0.878	99.2	2.409	2.652
23	-0.739	0.292	63	0.332	0.896	99.3	2.457	2.697
24	-0.706	0.305	64	0.358	0.915	99.4	2.512	2.748
25	-0.674	0.319	65	0.385	0.935	99.5	2.576	2.807
26	-0.643	0.332	66	0.412	0.954	99.6	2.652	2.878
27	-0.613	0.345	67	0.440	0.974	99.7	2.748	2.968
28	-0.583	0.358	68	0.468	0.994	99.8	2.878	3.090
29	-0.553	0.372	69	0.496	1.015	99.9	3.090	3.291
30	-0.524	0.385	70	0.524	1.036			
31	-0.496	0.399	71	0.553	1.058	99.91	3.121	3.320
32	-0.468	0.412	72	0.583	1.080	99.92	3.156	3.353
33	-0.440	0.426	73	0.613	1.103	99.93	3.195	3.390
34	-0.412	0.440	74	0.643	1.126	99.94	3.239	3.432
35	-0.385	0.454	75	0.674	1.150	99.95	3.291	3.481
36	-0.358	0.468	76	0.706	1.175	99.96	3.353	3.540
37	-0.332	0.482	77	0.739	1.200	99.97	3.432	3.615
38	-0.305	0.496	78	0.772	1.227	99.98	3.540	3.719
39	-0.279	0.510	79	0.806	1.254	99.99	3.719	3.891
40	-0.253	0.524	80	0.842	1.282			

**Table A9 t-Distribution**

Values of  $z$  for given values of the distribution function  $F(z)$  (see (8) in Sec. 25.3).  
 Example: For 9 degrees of freedom,  $z = 1.83$  when  $F(z) = 0.95$ .

$F(z)$	Number of Degrees of Freedom									
	1	2	3	4	5	6	7	8	9	10
0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.6	0.32	0.29	0.28	0.27	0.27	0.26	0.26	0.26	0.26	0.26
0.7	0.73	0.62	0.58	0.57	0.56	0.55	0.55	0.55	0.54	0.54
0.8	1.38	1.06	0.98	0.94	0.92	0.91	0.90	0.89	0.88	0.88
0.9	3.08	1.89	1.64	1.53	1.48	1.44	1.41	1.40	1.38	1.37
0.95	6.31	2.92	2.35	2.13	2.02	1.94	1.89	1.86	1.83	1.81
0.975	12.7	4.30	3.18	2.78	2.57	2.45	2.36	2.31	2.26	2.23
0.99	31.8	6.96	4.54	3.75	3.36	3.14	3.00	2.90	2.82	2.76
0.995	63.7	9.92	5.84	4.60	4.03	3.71	3.50	3.36	3.25	3.17
0.999	318.3	22.3	10.2	7.17	5.89	5.21	4.79	4.50	4.30	4.14

$F(z)$	Number of Degrees of Freedom									
	11	12	13	14	15	16	17	18	19	20
0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.6	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
0.7	0.54	0.54	0.54	0.54	0.54	0.54	0.53	0.53	0.53	0.53
0.8	0.88	0.87	0.87	0.87	0.87	0.86	0.86	0.86	0.86	0.86
0.9	1.36	1.36	1.35	1.35	1.34	1.34	1.33	1.33	1.33	1.33
0.95	1.80	1.78	1.77	1.76	1.75	1.75	1.74	1.73	1.73	1.72
0.975	2.20	2.18	2.16	2.14	2.13	2.12	2.11	2.10	2.09	2.09
0.99	2.72	2.68	2.65	2.62	2.60	2.58	2.57	2.55	2.54	2.53
0.995	3.11	3.05	3.01	2.98	2.95	2.92	2.90	2.88	2.86	2.85
0.999	4.02	3.93	3.85	3.79	3.73	3.69	3.65	3.61	3.58	3.55

$F(z)$	Number of Degrees of Freedom									
	22	24	26	28	30	40	50	100	200	$\infty$
0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.6	0.26	0.26	0.26	0.26	0.26	0.26	0.25	0.25	0.25	0.25
0.7	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.52
0.8	0.86	0.86	0.86	0.85	0.85	0.85	0.85	0.85	0.84	0.84
0.9	1.32	1.32	1.31	1.31	1.31	1.30	1.30	1.29	1.29	1.28
0.95	1.72	1.71	1.71	1.70	1.70	1.68	1.68	1.66	1.65	1.65
0.975	2.07	2.06	2.06	2.05	2.04	2.02	2.01	1.98	1.97	1.96
0.99	2.51	2.49	2.48	2.47	2.46	2.42	2.40	2.36	2.35	2.33
0.995	2.82	2.80	2.78	2.76	2.75	2.70	2.68	2.63	2.60	2.58
0.999	3.50	3.47	3.43	3.41	3.39	3.31	3.26	3.17	3.13	3.09

**Table A10 Chi-square Distribution**

Values of  $x$  for given values of the distribution function  $F(z)$  (see Sec. 25.3 before (17)).

Example: For 3 degrees of freedom,  $z = 11.34$  when  $F(z) = 0.99$ .

$F(z)$	Number of Degrees of Freedom									
	1	2	3	4	5	6	7	8	9	10
0.005	0.00	0.01	0.07	0.21	0.41	0.68	0.99	1.34	1.73	2.16
0.01	0.00	0.02	0.11	0.30	0.55	0.87	1.24	1.65	2.09	2.56
0.025	0.00	0.05	0.22	0.48	0.83	1.24	1.69	2.18	2.70	3.25
0.05	0.00	0.10	0.35	0.71	1.15	1.64	2.17	2.73	3.33	3.94
0.95	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31
0.975	5.02	7.38	9.35	11.14	12.83	14.45	16.01	17.53	19.02	20.48
0.99	6.63	9.21	11.34	13.28	15.09	16.81	18.48	20.09	21.67	23.21
0.995	7.88	10.60	12.84	14.86	16.75	18.55	20.28	21.95	23.59	25.19

$F(z)$	Number of Degrees of Freedom									
	11	12	13	14	15	16	17	18	19	20
0.005	2.60	3.07	3.57	4.07	4.60	5.14	5.70	6.26	6.84	7.43
0.01	3.05	3.57	4.11	4.66	5.23	5.81	6.41	7.01	7.63	8.26
0.025	3.82	4.40	5.01	5.63	6.26	6.91	7.56	8.23	8.91	9.59
0.05	4.57	5.23	5.89	6.57	7.26	7.96	8.67	9.39	10.12	10.85
0.95	19.68	21.03	22.36	23.68	25.00	26.30	27.59	28.87	30.14	31.41
0.975	21.92	23.34	24.74	26.12	27.49	28.85	30.19	31.53	32.85	34.17
0.99	24.72	26.22	27.69	29.14	30.58	32.00	33.41	34.81	36.19	37.57
0.995	26.76	28.30	29.82	31.32	32.80	34.27	35.72	37.16	38.58	40.00

$F(z)$	Number of Degrees of Freedom									
	21	22	23	24	25	26	27	28	29	30
0.005	8.0	8.6	9.3	9.9	10.5	11.2	11.8	12.5	13.1	13.8
0.01	8.9	9.5	10.2	10.9	11.5	12.2	12.9	13.6	14.3	15.0
0.025	10.3	11.0	11.7	12.4	13.1	13.8	14.6	15.3	16.0	16.8
0.05	11.6	12.3	13.1	13.8	14.6	15.4	16.2	16.9	17.7	18.5
0.95	32.7	33.9	35.2	36.4	37.7	38.9	40.1	41.3	42.6	43.8
0.975	35.5	36.8	38.1	39.4	40.6	41.9	43.2	44.5	45.7	47.0
0.99	38.9	40.3	41.6	43.0	44.3	45.6	47.0	48.3	49.6	50.9
0.995	41.4	42.8	44.2	45.6	46.9	48.3	49.6	51.0	52.3	53.7

$F(z)$	Number of Degrees of Freedom							
	40	50	60	70	80	90	100	> 100 (Approximation)
0.005	20.7	28.0	35.5	43.3	51.2	59.2	67.3	$\frac{1}{2}(h - 2.58)^2$
0.01	22.2	29.7	37.5	45.4	53.5	61.8	70.1	$\frac{1}{2}(h - 2.33)^2$
0.025	24.4	32.4	40.5	48.8	57.2	65.6	74.2	$\frac{1}{2}(h - 1.96)^2$
0.05	26.5	34.8	43.2	51.7	60.4	69.1	77.9	$\frac{1}{2}(h - 1.64)^2$
0.95	55.8	67.5	79.1	90.5	101.9	113.1	124.3	$\frac{1}{2}(h + 1.64)^2$
0.975	59.3	71.4	83.3	95.0	106.6	118.1	129.6	$\frac{1}{2}(h + 1.96)^2$
0.99	63.7	76.2	88.4	100.4	112.3	124.1	135.8	$\frac{1}{2}(h + 2.33)^2$
0.995	66.8	79.5	92.0	104.2	116.3	128.3	140.2	$\frac{1}{2}(h + 2.58)^2$

In the last column,  $h = \sqrt{2m - 1}$ , where  $m$  is the number of degrees of freedom.



**Table A11 F-Distribution with  $(m, n)$  Degrees of Freedom**

Values of  $z$  for which the distribution function  $F(z)$  [see (13), Sec. 25.4] has the value **0.95**  
 Example: For (7, 4) d.f.,  $z = 6.09$  if  $F(z) = 0.95$ .

$n$	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$	$m = 8$	$m = 9$
1	161	200	216	225	230	234	237	239	241
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89
$\infty$	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

**Table A11 F-Distribution with  $(m, n)$  Degrees of Freedom (continued)**Values of  $z$  for which the distribution function  $F(z)$  [see (13), Sec. 25.4] has the value **0.95**

$n$	$m = 10$	$m = 15$	$m = 20$	$m = 30$	$m = 40$	$m = 50$	$m = 100$	$\infty$
1	242	246	248	250	251	252	253	254
2	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5
3	8.79	8.70	8.66	8.62	8.59	8.58	8.55	8.53
4	5.96	5.86	5.80	5.75	5.72	5.70	5.66	5.63
5	4.74	4.62	4.56	4.50	4.46	4.44	4.41	4.37
6	4.06	3.94	3.87	3.81	3.77	3.75	3.71	3.67
7	3.64	3.51	3.44	3.38	3.34	3.32	3.27	3.23
8	3.35	3.22	3.15	3.08	3.04	3.02	2.97	2.93
9	3.14	3.01	2.94	2.86	2.83	2.80	2.76	2.71
10	2.98	2.85	2.77	2.70	2.66	2.64	2.59	2.54
11	2.85	2.72	2.65	2.57	2.53	2.51	2.46	2.40
12	2.75	2.62	2.54	2.47	2.43	2.40	2.35	2.30
13	2.67	2.53	2.46	2.38	2.34	2.31	2.26	2.21
14	2.60	2.46	2.39	2.31	2.27	2.24	2.19	2.13
15	2.54	2.40	2.33	2.25	2.20	2.18	2.12	2.07
16	2.49	2.35	2.28	2.19	2.15	2.12	2.07	2.01
17	2.45	2.31	2.23	2.15	2.10	2.08	2.02	1.96
18	2.41	2.27	2.19	2.11	2.06	2.04	1.98	1.92
19	2.38	2.23	2.16	2.07	2.03	2.00	1.94	1.88
20	2.35	2.20	2.12	2.04	1.99	1.97	1.91	1.84
22	2.30	2.15	2.07	1.98	1.94	1.91	1.85	1.78
24	2.25	2.11	2.03	1.94	1.89	1.86	1.80	1.73
26	2.22	2.07	1.99	1.90	1.85	1.82	1.76	1.69
28	2.19	2.04	1.96	1.87	1.82	1.79	1.73	1.65
30	2.16	2.01	1.93	1.84	1.79	1.76	1.70	1.62
32	2.14	1.99	1.91	1.82	1.77	1.74	1.67	1.59
34	2.12	1.97	1.89	1.80	1.75	1.71	1.65	1.57
36	2.11	1.95	1.87	1.78	1.73	1.69	1.62	1.55
38	2.09	1.94	1.85	1.76	1.71	1.68	1.61	1.53
40	2.08	1.92	1.84	1.74	1.69	1.66	1.59	1.51
50	2.03	1.87	1.78	1.69	1.63	1.60	1.52	1.44
60	1.99	1.84	1.75	1.65	1.59	1.56	1.48	1.39
70	1.97	1.81	1.72	1.62	1.57	1.53	1.45	1.35
80	1.95	1.79	1.70	1.60	1.54	1.51	1.43	1.32
90	1.94	1.78	1.69	1.59	1.53	1.49	1.41	1.30
100	1.93	1.77	1.68	1.57	1.52	1.48	1.39	1.28
150	1.89	1.73	1.64	1.54	1.48	1.44	1.34	1.22
200	1.88	1.72	1.62	1.52	1.46	1.41	1.32	1.19
1000	1.84	1.68	1.58	1.47	1.41	1.36	1.26	1.08
$\infty$	1.83	1.67	1.57	1.46	1.39	1.35	1.24	1.00

**Table A11 F-Distribution with  $(m, n)$  Degrees of Freedom (continued)**Values of  $z$  for which the distribution function  $F(z)$  [see (13), Sec. 25.4] has the value **0.99**

$n$	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$	$m = 8$	$m = 9$
1	4052	4999	5403	5625	5764	5859	5928	5981	6022
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
32	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02
34	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98
36	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95
38	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59
150	6.81	4.75	3.91	3.45	3.14	2.92	2.76	2.63	2.53
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50
1000	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43
$\infty$	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

**Table A11 F-Distribution with  $(m, n)$  Degrees of Freedom (continued)**Values of  $z$  for which the distribution function  $F(z)$  [see (13), Sec. 25.4] has the value **0.99**

$n$	$m = 10$	$m = 15$	$m = 20$	$m = 30$	$m = 40$	$m = 50$	$m = 100$	$\infty$
1	6056	6157	6209	6261	6287	6303	6334	6366
2	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5
3	27.2	26.9	26.7	26.5	26.4	26.4	26.2	26.1
4	14.5	14.2	14.0	13.8	13.7	13.7	13.6	13.5
5	10.1	9.72	9.55	9.38	9.29	9.24	9.13	9.02
6	7.87	7.56	7.40	7.23	7.14	7.09	6.99	6.88
7	6.62	6.31	6.16	5.99	5.91	5.86	5.75	5.65
8	5.81	5.52	5.36	5.20	5.12	5.07	4.96	4.86
9	5.26	4.96	4.81	4.65	4.57	4.52	4.42	4.31
10	4.85	4.56	4.41	4.25	4.17	4.12	4.01	3.91
11	4.54	4.25	4.10	3.94	3.86	3.81	3.71	3.60
12	4.30	4.01	3.86	3.70	3.62	3.57	3.47	3.36
13	4.10	3.82	3.66	3.51	3.43	3.38	3.27	3.17
14	3.94	3.66	3.51	3.35	3.27	3.22	3.11	3.00
15	3.80	3.52	3.37	3.21	3.13	3.08	2.98	2.87
16	3.69	3.41	3.26	3.10	3.02	2.97	2.86	2.75
17	3.59	3.31	3.16	3.00	2.92	2.87	2.76	2.65
18	3.51	3.23	3.08	2.92	2.84	2.78	2.68	2.57
19	3.43	3.15	3.00	2.84	2.76	2.71	2.60	2.49
20	3.37	3.09	2.94	2.78	2.69	2.64	2.54	2.42
22	3.26	2.98	2.83	2.67	2.58	2.53	2.42	2.31
24	3.17	2.89	2.74	2.58	2.49	2.44	2.33	2.21
26	3.09	2.81	2.66	2.50	2.42	2.36	2.25	2.13
28	3.03	2.75	2.60	2.44	2.35	2.30	2.19	2.06
30	2.98	2.70	2.55	2.39	2.30	2.25	2.13	2.01
32	2.93	2.65	2.50	2.34	2.25	2.20	2.08	1.96
34	2.89	2.61	2.46	2.30	2.21	2.16	2.04	1.91
36	2.86	2.58	2.43	2.26	2.18	2.12	2.00	1.87
38	2.83	2.55	2.40	2.23	2.14	2.09	1.97	1.84
40	2.80	2.52	2.37	2.20	2.11	2.06	1.94	1.80
50	2.70	2.42	2.27	2.10	2.01	1.95	1.82	1.68
60	2.63	2.35	2.20	2.03	1.94	1.88	1.75	1.60
70	2.59	2.31	2.15	1.98	1.89	1.83	1.70	1.54
80	2.55	2.27	2.12	1.94	1.85	1.79	1.65	1.49
90	2.52	2.24	2.09	1.92	1.82	1.76	1.62	1.46
100	2.50	2.22	2.07	1.89	1.80	1.74	1.60	1.43
150	2.44	2.16	2.00	1.83	1.73	1.66	1.52	1.33
200	2.41	2.13	1.97	1.79	1.69	1.63	1.48	1.28
1000	2.34	2.06	1.90	1.72	1.61	1.54	1.38	1.11
$\infty$	2.32	2.04	1.88	1.70	1.59	1.52	1.36	1.00

**Table A12** Distribution Function  $F(x) = P(T \leq x)$  of the Random Variable  $T$  in Section 25.8

x	n=3
0	0.
1	167 500

x	n=4
0	0.
1	042 167
2	375

x	n=5
0	0.
1	008 042
2	117 242
3	408
4	

x	n=6
0	0.
1	001 008
2	028 068
3	136 235
4	360 500
5	
6	
7	

x	n=7
1	0.
2	001 005
3	015 035
4	068 119
5	191 281
6	386 500
7	
8	
9	
10	

x	n=8
2	0.
3	001 003
4	007 016
5	031 054
6	089 138
7	199 274
8	360 452
9	
10	
11	
12	
13	

x	n=9
4	0.
5	001 003
6	006 012
7	022 038
8	060 130
9	179 238
10	306 381
11	460
12	
13	
14	
15	
16	
17	

x	n=10
6	0.
7	001 002
8	005 008
9	014 023
10	036 054
11	078 108
12	146 190
13	242 300
14	364 431
15	500
16	
17	
18	
19	
20	
21	
22	

x	n=11
8	0.
9	001 002
10	003 005
11	008 013
12	020 030
13	043 060
14	082 109
15	141 179
16	223 271
17	324 381
18	440 500
19	
20	
21	
22	
23	
24	
25	
26	
27	

x	n=20
50	0.
51	001 002
52	002 003
53	004 005
54	006 007
55	008 010
56	012 014
57	017 020
58	023 027
59	032 037
60	043 049
61	056 062
62	067 072
63	078 082
64	087 093
65	104 117
66	130 144
67	159 176
68	193 211
69	230 257
70	271 293
71	315 339
72	362 387
73	411 436
74	462 487
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	

x	n=19
43	0.
44	001 002
45	002 003
46	003 004
47	005 006
48	010 018
49	025 034
50	047 054
51	062 072
52	082 093
53	105 119
54	133 149
55	166 184
56	203 223
57	245 267
58	290 314
59	339 365
60	408 431
61	445 473
62	500
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	

x	n=18
38	0.
39	001 002
40	002 003
41	004 007
42	011 016
43	021 026
44	032 038
45	046 054
46	064 076
47	088 100
48	115 130
49	147 