



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	66777888999999
29	9	001123
30	9	9

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

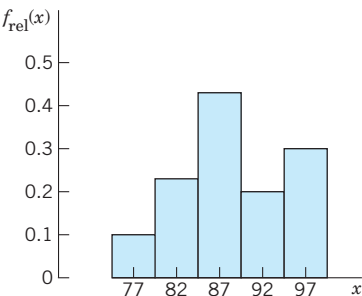


Fig. 508. Histogram of the data in Example 1 (grouped as in Fig. 507)

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 $R = 22$.

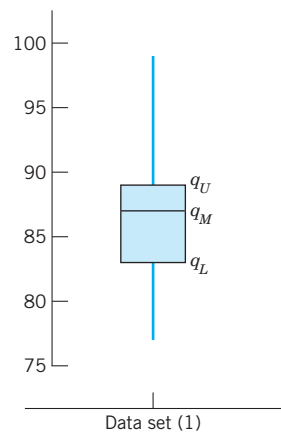


Fig. 509. Boxplot of the data set (1)

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 $89 + 1.5 IQR = 98$, 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 (kg/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 (Fe_2O_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 kg/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**¹ 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?).

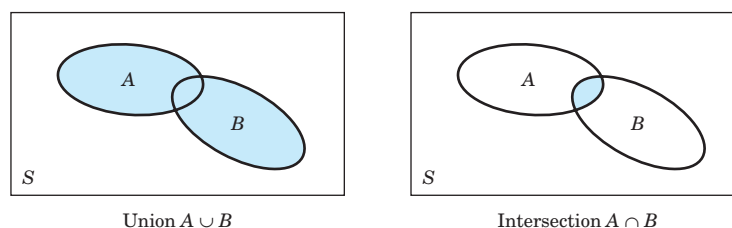


Fig. 510. Venn diagrams showing two events A and B in a sample space S and their union $A \cup B$ (colored) and intersection $A \cap B$ (colored)

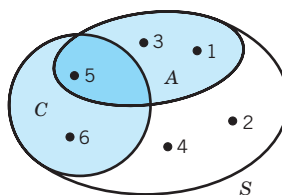


Fig. 511. Venn diagram for the experiment of rolling a die, showing S , $A = \{1, 3, 5\}$, $C = \{5, 6\}$, $A \cup C = \{1, 3, 5, 6\}$, $A \cap C = \{5\}$

PROBLEM SET 24.2

1–12 SAMPLE SPACES, EVENTS

Graph a sample space for the experiments:

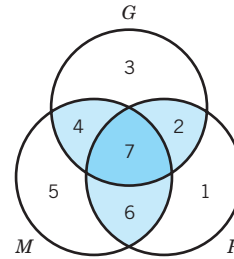
1. Drawing 3 screws from a lot of right-handed and left-handed screws
2. Tossing 2 coins
3. Rolling 2 dice
4. Rolling a die until the first *Six* appears
5. Tossing a coin until the first *Head* appears
6. Recording the lifetime of each of 3 lightbulbs

¹JOHN 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 σ -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. ■

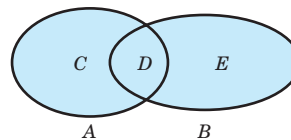


Fig. 512. Proof of Theorem 3

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**²

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

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

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

²JAMES 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!* ■

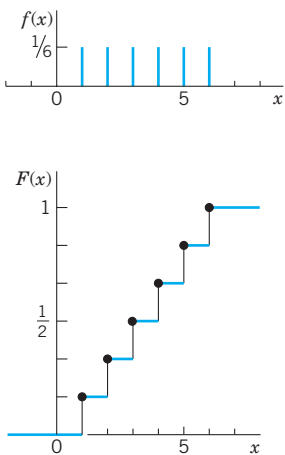


Fig. 513. Probability function $f(x)$ and distribution function $F(x)$ of the random variable $X = \text{Number obtained in tossing a fair die once}$

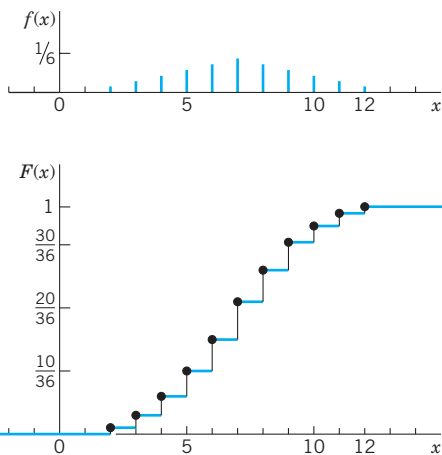


Fig. 514. Probability function $f(x)$ and distribution function $F(x)$ of the random variable $X = \text{Sum of the two numbers obtained in tossing two fair dice once}$

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

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

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 ($> a$) 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.

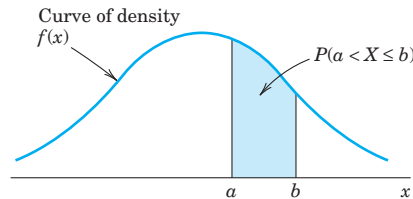


Fig. 515. Example illustrating formula (9)

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 thinner 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 μ and variance σ^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** μ (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** σ^2 (sigma square) by

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

σ (the positive square root of σ^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 μ 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 μ and σ^2 that measure certain properties of a distribution are called **parameters**. μ and σ^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 μ and σ^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 σ^2 is large. ■

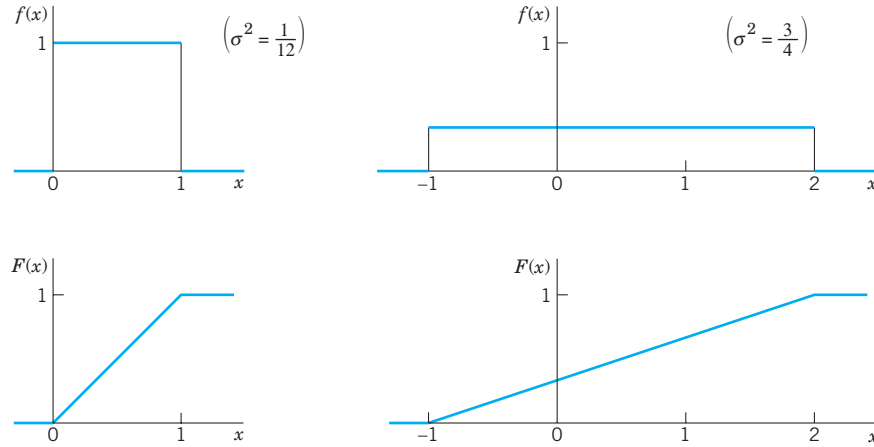


Fig. 516. Uniform distributions having the same mean (0.5) but different variances σ^2

Symmetry. We can obtain the mean μ 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 μ and variance σ^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 μ and variance σ^2 , then the random variable

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

has the mean μ^* and variance σ^{*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 Δx on the x -axis there corresponds the probability $f(x)\Delta x$ [approximately; the area of a rectangle of base Δx 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 Δx^* 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 μ . This proves (5) for μ^* . 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 μ 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 , μ , and σ^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** γ 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.

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

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

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

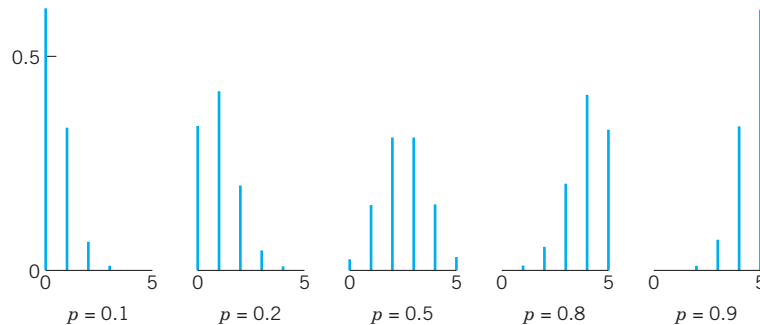


Fig. 517. Probability function (2) of the binomial distribution for $n = 5$ and various values of p

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 μ . 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 μ 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.

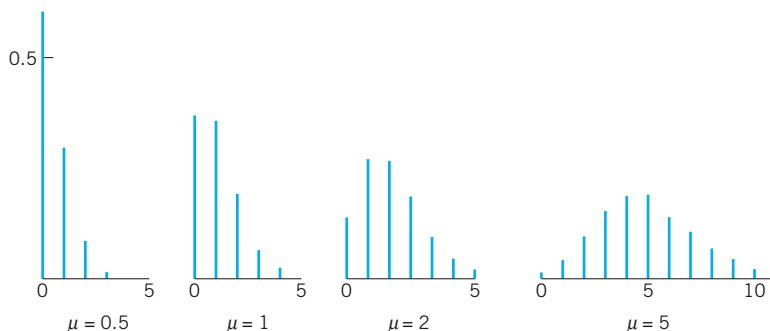


Fig. 518. Probability function (5) of the Poisson distribution for various values of μ

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

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

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 μ 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

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

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. μ is the mean and σ 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 ∞ 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 σ is, as it should be (Fig. 519).

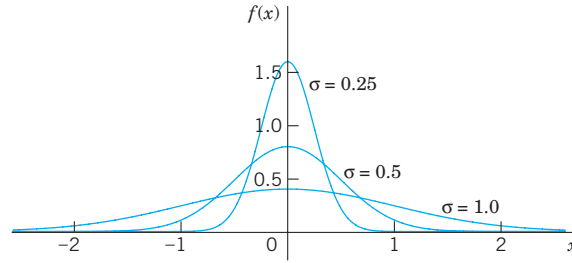


Fig. 519. Density (1) of the normal distribution with $\mu = 0$ for various values of σ

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 μ and σ 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.

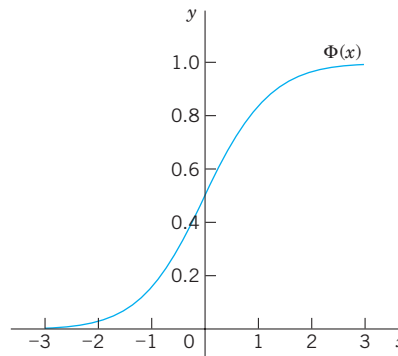


Fig. 520. Distribution function $\Phi(z)$ of the normal distribution with mean 0 and variance 1

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

The distribution function $F(x)$ of the normal distribution with any μ and σ [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 σ 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 μ and standard deviation σ 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. \blacksquare

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 μ by more than σ , 2σ , or 3σ will occur in one of about 3, 20, and 300 trials, respectively.

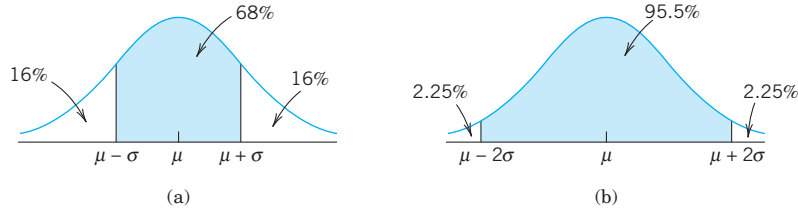


Fig. 521. Illustration of formula (6)

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,

$$\begin{aligned}
 (7) \quad & \text{(a)} \quad P(\mu - 1.96\sigma < X \leq \mu + 1.96\sigma) = 95\% \\
 & \text{(b)} \quad P(\mu - 2.58\sigma < X \leq \mu + 2.58\sigma) = 99\% \\
 & \text{(c)} \quad P(\mu - 3.29\sigma < X \leq \mu + 3.29\sigma) = 99.9\%.
 \end{aligned}$$

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 ± 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 ± 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 ∞ . 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 α and β 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 148Ω and 152Ω ? Between 140Ω and 160Ω ?
- 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 σ 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 μ and variance σ^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

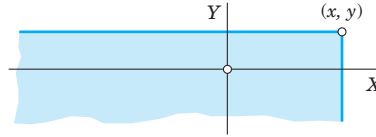


Fig. 522. Formula (1)

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

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

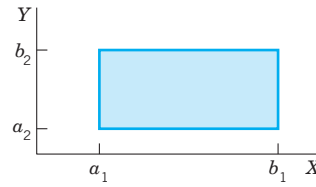


Fig. 523. Notion of a two-dimensional distribution

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

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

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. ■

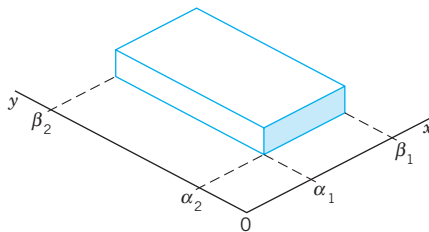


Fig. 524. Density function (8) of the uniform distribution

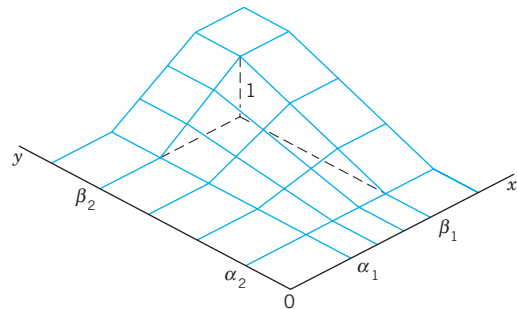


Fig. 525. Distribution function of the uniform distribution defined by (8)

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 μ and σ^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 σ^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 σ_1^2 and σ_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** μ and **variance** σ^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 $f(x, y)$.