CHAPTER **15**

Probability and Statistics

1. INTRODUCTION

The theory of probability has many applications in the physical sciences. It is of basic importance in quantum mechanics where results may be expressed in terms of probabilities (see Chapter 13, Schrödinger equation). It is needed whenever we are dealing with large numbers of particles or variables where it is impossible or impractical to have complete information about each one, such as in kinetic theory and statistical mechanics and a great variety of engineering problems. Statistics is the part of probability theory which deals with the interpretation of sets of data. You need statistical terms and methods every time you make a set of laboratory measurements. In this chapter, we shall discuss some of the basic ideas of probability and statistics which are most useful in applications.

The word "probably" is frequently used in everyday life. We say "The test will probably be hard," "It will probably snow today," "We will probably win this game," and so on. Such statements always imply a state of partial ignorance about the outcome of some event; we do not say "probably" about something whose outcome we know. The theory of probability tries to express more precisely just what our state of ignorance is. We say that the probability of getting a head in one toss of a coin is $\frac{1}{2}$, and similarly for a tail. We mean by this that there are two possible outcomes of the experiment (if we do not consider the possibility of the coin's standing on edge) and that we have no reason to expect one outcome more than the other; therefore we assign equal probabilities to the two possible outcomes. (See end of Section 2 for further discussion of this.)

Consider the following problem. You and I each toss a coin and look at our own coins but not each other's. The question is "What is the probability that both coins show heads?" Suppose you see that your coin shows tails; you say that the probability that both coins are heads is zero because you *know* that yours is tails. On the other hand, suppose I see that my coin is heads; then I say that the probability of both heads is $\frac{1}{2}$ because I don't know whether your coin shows heads or tails. Now suppose neither of us looks at either coin, but a third person looks at both coins and gives us the information that at least one is heads. Without this

information. there

(1.1)

to each of which we
and Section 3). The
information about t
we still consider the

Notice in the abo
on the state of know
also that in order to
equally likely outco
that these are mutua
collectively exhausti
have no information
assume the same pr
formalize this notion

If there a
tively exha
event E is

(1.2)

Example 1. Find the proba
will be either a diamo

There are 52 differe
we assume all cards e
and the 3 other kings)

Example 2. A three-digit n
dom." ("At random"
bility of being selected

There are 900 thre
all three digits the sa

PROBLEMS, SECTION 1

1. If you select a thre
digit is 7? What is

2. Three coins are to
That the first two
the probability tha

3. In a box there are
what is the probab

information, there are four possibilities, namely

$$hh \qquad tt \qquad th \qquad ht \tag{1.1}$$

to each of which we would ordinarily assign the probability $\frac{1}{4}$ (see end of Section 2, and Section 3). The information "at least one head" rules out tt, but gives no new information about the other three cases. Since hh, th, ht were equally likely before, we still consider them equally likely and say that the probability of hh is $\frac{1}{3}$.

Notice in the above discussion that the answer to a probability problem depends on the state of knowledge (or ignorance) of the person giving the answer. Notice also that in order to find the probability of an event, we consider all the different equally likely outcomes which are possible according to our information. We say that these are mutually exclusive (for example, if a coin is heads it cannot be tails), collectively exhaustive (we must consider *all* possibilities), and equally likely (we have no information which makes us expect one result more than another so we assume the same probability for each one of the set of outcomes). Let us now formalize this notion of probability as a definition (also see Section 2).

(1.2) **If there are several equally likely, mutually exclusive, and collectively exhaustive outcomes of an experiment, the probability of an event E is**

$$p = \frac{\text{number of outcomes favorable to } E}{\text{total number of outcomes}}.$$

Example 1. Find the probability that a single card drawn from a shuffled deck of cards will be either a diamond or a king (or both).

There are 52 different possible outcomes of the drawing; since the deck is shuffled, we assume all cards equally likely. Of the 52 cards, 16 are favorable (13 diamonds and the 3 other kings); therefore by (1.2) the desired probability is $\frac{16}{52} = \frac{4}{13}$.

Example 2. A three-digit number (that is, a number from 100–999) is selected "at random." ("At random" means that we assume all numbers to have the same probability of being selected.) What is the probability that all three digits are the same?

There are 900 three-digit numbers; 9 of them (namely 111, 222, $\cdots$, 999) have all three digits the same. Hence the desired probability is $\frac{9}{900} = \frac{1}{100}$.

PROBLEMS, SECTION 1

1. If you select a three-digit number at random, what is the probability that the units digit is 7? What is the probability that the hundreds digit is 7?

2. Three coins are tossed; what is the probability that two are heads and one tails? That the first two are heads and the third tails? If at least two are heads, what is the probability that all are heads?

3. In a box there are 2 white, 3 black, and 4 red balls. If a ball is drawn at random, what is the probability that it is black? That it is *not* red?

4. A single card is drawn at random from a shuffled deck. What is the probability that it is red? That it is the ace of hearts? That it is either a three or a five? That it is either an ace or red or both?

5. Given a family of two children (assume boys and girls equally likely, that is, probability 1/2 for each), what is the probability that both are boys? That at least one is a girl? Given that at least one is a girl, what is the probability that both are girls? Given that the first two are girls, what is the probability that an expected third child will be a boy?

6. A trick deck of cards is printed with the hearts and diamonds black, and the spades and clubs red. A card is chosen at random from this deck (after it is shuffled). Find the probability that it is either a red card or the queen of hearts. That it is either a red face card or a club. That it is either a red ace or a diamond.

7. A letter is selected at random from the alphabet. What is the probability that it is one of the letters in the word "probability?" What is the probability that it occurs in the first half of the alphabet? What is the probability that it is a letter after x?

8. An integer N is chosen at random with $1 \leq N \leq 100$. What is the probability that N is divisible by 11? That $N > 90$? That $N \leq 3$? That N is a perfect square?

9. You are trying to find instrument A in a laboratory. Unfortunately, someone has put both instruments A and another kind (which we shall call B) away in identical unmarked boxes mixed at random on a shelf. You know that the laboratory has 3 A's and 7 B's. If you take down one box, what is the probability that you get an A? If it is a B and you put it on the table and take down another box, what is the probability that you get an A this time?

10. A shopping mall has four entrances, one on the North, one on the South, and two on the East. If you enter at random, shop and then exit at random, what is the probability that you enter and exit on the same side of the mall?

2. SAMPLE SPACE

It is frequently convenient to make a list of the possible outcomes of an experiment [as we did in (1.1)]. Such a set of all possible mutually exclusive outcomes is called a *sample space*; each individual outcome is called a *point* of the sample space. There are many different sample spaces for any given problem. For example, instead of (1.1), we could say that a set of all mutually exclusive outcomes of two tosses of a coin is

(2.1) 2 heads, 1 head, no heads.

Still another sample space for the same problem is

(2.2) no heads, at least 1 head.

(Can you list some more examples?) On the other hand, the set of outcomes

2 heads, at least 1 head, exactly 1 tail.

cannot be used as a sample space, because these outcomes are not mutually exclusive. "At least 1 head" includes "2 heads" and also includes "exactly 1 tail" (which means also "exactly 1 head").

In order to use a sample space to solve problems, we need to have the probabilities corresponding to the different points in the sample space. We usually assign probability 1/4 to each of the outcomes listed in (1.1). (See end of Section 2 and Section 3.) We call such a list of equally likely outcomes a *uniform* sample space. Now suppose the outcomes are not equally likely. Satisfy yourself that the probabilities associated with the points of (2.1) and (2.2) are as follows.

	$2h$	$1h$	no h		no h	at least 1 h
For (2.1):	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	and for (2.2):	$\frac{1}{4}$	$\frac{3}{4}$

The sample spaces (2.1) and (2.2) with different probabilities associated with different points are called *nonuniform* sample spaces. For some problems, there may be both uniform and nonuniform sample spaces; for example, (1.1) is a uniform sample space, and (2.1) and (2.2) are nonuniform sample spaces for a toss of two coins. But sometimes there *is* no uniform sample space; for example, consider a weighted coin which has a probability $\frac{1}{3}$ for heads and $\frac{2}{3}$ for tails. In such cases, we cannot use the definition (1.2) of probability, and we need the following more general definition.

Definition of Probability. Given any sample space (uniform or not) and the probabilities associated with the points, we find the probability of an event by adding the probabilities associated with all the sample points favorable to the event.

For a given nonuniform sample space, we must use this definition since (1.2) does not apply. If the given sample space is uniform, or if there is an underlying uniform sample space [for example, (1.1) is the uniform space underlying (2.1) and (2.2)], then this definition is consistent with the definition (1.2) by equally likely cases (Problems 15 and 16), and we may use either definition. As an example, let us find from (2.1) the probability of at least one head; this is the probability of one head plus the probability of two heads or $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$. We get the same result from the uniform sample space (1.1) using either (1.2) or the definition above.

If we can easily construct several sample spaces for a given problem, we must choose an appropriate one for the question we want to answer. Suppose we ask the question: In two tosses of a coin, what is the probability that both are heads? From either (1.1) or (2.1) we find the answer $\frac{1}{4}$; (2.2) is not an appropriate sample space to use in answering this question. (Why not?) To find the probability of both tails, we could use any of the three listed sample spaces, and to find the probability that the first toss gave a head and the second a tail, we could use only (1.1) since the other sample spaces do not give enough information. Let us now consider some less trivial examples.

Example 1. A coin is tossed three times. A uniform sample space for this problem contains eight points.

$$(2.3) \qquad \begin{matrix} hhh & hth & ttt & tht \\ hht & thh & tth & htt \end{matrix}$$

and we attach probability $\frac{1}{8}$ to each. Now let us use this sample space to answer some questions.

What is the probability of at least two tails in succession? By actual count, we see that there are three such cases, so the probability is $\frac{3}{8}$.

What is the probability that two consecutive coins fall the same? Again by actual count, this is true in six cases, so the probability is $\frac{6}{8}$ or $\frac{3}{4}$.

If we know that there was at least one tail, what is the probability of all tails? The point hhh is now ruled out; we have a new sample space consisting of seven points. Since the new information (at least one tail) tells us nothing new about these seven outcomes, we consider them equally probable, each with probability $\frac{1}{7}$. Thus the probability of all tails when all heads is ruled out is $\frac{1}{7}$.

(See problems 11 and 12 for further discussion of this example.)

Example 2. Let two dice be thrown; the first die can show any number from 1 to 6 and similarly for the second die. Then there are 36 possible outcomes or points in a uniform sample space for this problem; with each point we associate the probability $\frac{1}{36}$. We can indicate a 3 on the first die and a 2 on the second die by the symbol 3,2. Then the sample space is as shown in (2.4). (Ignore the circling of some points and the letters a and b right now; they are for use in the problems below.)

(2.4)

	1,1	1,2	1,3	1,4	1,5	1,6	
	2,1	2,2	2,3	2,4	2,5	2,6	
	3,1	3,2	3,3	3,4	3,5	3,6	
a	4,1	4,2	4,3	4,4	4,5	4,6	b
	5,1	5,2	5,3	5,4	5,5	5,6	
	6,1	6,2	6,3	6,4	6,5	6,6	

Let us now ask some questions and use the sample space (2.4) to answer them.

(a) What is the probability that the sum of the numbers on the dice will be 5? The sample space points circled and marked a in (2.4) give all the cases for which the sum is 5. There are four of these sample points; therefore the probability that the sum is 5 is $\frac{4}{36}$ or $\frac{1}{9}$.

(b) What is the probability that the sum on the dice is divisible by 5? This means a sum of 5 or 10; the four points circled and marked a in (2.4) correspond to a sum of 5, and the three points circled and marked b correspond to a sum of 10. Thus there are seven points in the sample space corresponding to a sum divisible by 5, so the probability of a sum divisible by 5 is $\frac{7}{36}$ (7 favorable cases out of 36 possible cases, or 7 times the probability $\frac{1}{36}$ of each of the favorable sample points).

(c) Set up a sample space in which the points correspond to the possible sums of the two numbers on the dice, and find the probabilities associated with the points of this nonuniform sample space. The possible sums range from 2 (that is, $1+1$) to 12 (that is, $6+6$). From (2.4) we see that the points corresponding to any given sum lie on a diagonal (parallel to the diagonal elements marked a or b). There is one point corresponding to the sum 2; there are two points giving the sum 3, three

points for sum 4, e

(2.5) Sample
Associ
probabi

(d) What is the
answer this from th
that the sum 7 has
(e) What is the
to 9? Using (2.5),
and 12. Thus the de

So far we have b
that heads and tails
about this, you are
true, as a bent or w
the mathematical th
the physical world. M
a set of assumptions
results follow. The b
the probabilities asso
tossing problem, we
probability of tails are
in two tosses is $\frac{1}{4}$. (S
correct is not a math
are trying to solve. I
can somehow estimate
tails), then the mathem
absence of any informa
the "natural" or "intui
possible answer to the
If the results predicte
then the assumptions
Section 4, Example 5.)
In this chapter we
lating the probabilities
associated with the poi
these probabilities to b
applies, however, if we
problem, etc.) by any s

PROBLEMS, SECTION 2

1 to 10. Set up an appr
to solve the problem

11. Set up several nonu
(Example 1, above).

points for sum 4, etc. Thus we have:

(2.5)

Sample Space	2	3	4	5	6	7	8	9	10	11	12
Associated probabilities	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

(d) What is the most probable sum in a toss of two dice? Although we can answer this from the sample space (2.4) (Try it!), it is easier from (2.5). We see that the sum 7 has the largest probability, namely $\frac{6}{36} = \frac{1}{6}$.

(e) What is the probability that the sum on the dice is greater than or equal to 9? Using (2.5), we add the probabilities associated with the sums 9, 10, 11, and 12. Thus the desired probability is

$$\frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{10}{36} = \frac{5}{18}.$$

So far we have been talking as if it were perfectly obvious and unquestionable that heads and tails are equally likely in the toss of a coin. If you have felt skeptical about this, you are perfectly right. It is *not* obvious; it is not even necessarily true, as a bent or weighted coin would show. We must distinguish here between the mathematical theory of probability and its application to a problem about the physical world. Mathematical probability (like all of mathematics) starts with a set of assumptions and shows that *if* the assumptions are true, *then* various results follow. The basic assumptions in a mathematical probability problem are the probabilities associated with the points of the sample space. Thus in a coin tossing problem, we *assume* that for each toss the probability of heads and the probability of tails are both $\frac{1}{2}$, and then we show that the probability of both heads in two tosses is $\frac{1}{4}$. (See Section 3.) The question of whether the assumptions are correct is not a mathematical one. Here we must ask what physical problem we are trying to solve. If we are dealing with a weighted coin, and if we know or can somehow estimate experimentally the probability p of heads (and so $1 - p$ of tails), then the mathematical theory starts with these values instead of $\frac{1}{2}, \frac{1}{2}$. In the absence of any information as to whether heads or tails is more likely, we often make the "natural" or "intuitive" assumption that the probabilities are both $\frac{1}{2}$. The only possible answer to the question of whether this is correct or not lies in experiment. If the results predicted on the basis of our assumptions agree with experiment, then the assumptions are good; otherwise we must revise the assumptions. (See Section 4, Example 5.)

In this chapter we shall consider mainly the mathematical methods of calculating the probabilities of complicated happenings if we are given the probabilities associated with the points of the sample space. For simplicity, we shall often assume these probabilities to be the "natural" ones; the mathematical theory we develop applies, however, if we replace these "natural" probabilities ($\frac{1}{2}, \frac{1}{2}$ in the coin toss problem, etc.) by any set of non-negative fractions whose sum is 1.

PROBLEMS, SECTION 2

1 to 10. Set up an appropriate sample space for each of Problems 1.1 to 1.10 and use it to solve the problem. Use either a uniform or nonuniform sample space or try both.

11. Set up several nonuniform sample spaces for the problem of three tosses of a coin (Example 1, above).

12. Use the sample space of Example 1 above, or one or more of your sample spaces in Problem 11, to answer the following questions.

(a) If there were more heads than tails, what is the probability of one tail?

(b) If two heads did not appear in succession, what is the probability of all tails?

(c) If the coins did not all fall alike, what is the probability that two in succession were alike?

(d) If N_t = number of tails and N_h = number of heads, what is the probability that $|N_h - N_t| = 1$?

(e) If there was at least one head, what is the probability of exactly two heads?

13. A student claims in Problem 1.5 that if one child is a girl, the probability that both are girls is $\frac{1}{2}$. Use appropriate sample spaces to show what is wrong with the following argument: It doesn't matter whether the girl is the older child or the younger; in either case the probability is $\frac{1}{2}$ that the other child is a girl.

14. Two dice are thrown. Use the sample space (2.4) to answer the following questions.

(a) What is the probability of being able to form a two-digit number greater than 33 with the two numbers on the dice? (Note that the sample point 1, 4 yields the two-digit number 41 which is greater than 33, etc.)

(b) Repeat part (a) for the probability of being able to form a two-digit number greater than or equal to 42.

(c) Can you find a two-digit number (or numbers) such that the probability of being able to form a larger number is the same as the probability of being able to form a smaller number? [See note, part (a).]

15. Use both the sample space (2.4) and the sample space (2.5) to answer the following questions about a toss of two dice.

(a) What is the probability that the sum is ≥ 4?

(b) What is the probability that the sum is even?

(c) What is the probability that the sum is divisible by 3?

(d) If the sum is odd, what is the probability that it is equal to 7?

(e) What is the probability that the product of the numbers on the two dice is 12?

16. Given an nonuniform sample space and the probabilities associated with the points, we defined the probability of an event A as the sum of the probabilities associated with the sample points favorable to A. [You used this definition in Problem 15 with the sample space (2.5).] Show that this definition is consistent with the definition by equally likely cases if there is also a uniform sample space for the problem (as there was in Problem 15). *Hint:* Let the uniform sample space have N points each with the probability N^{-1}. Let the nonuniform sample space have $n < N$ points, the first point corresponding to N_1 points of the uniform space, the second to N_2 points, etc. What is

$$N_1 + N_2 + \cdots + N_n?$$

What are $p_1, p_2, \ldots$, the probabilities associated with the first, second, etc., points of the nonuniform space? What is $p_1 + p_2 + \cdots + p_n$? Now consider an event for which several points, say i, j, k, of the nonuniform sample space are favorable. Then using the nonuniform sample space, we have, by definition of the probability p of the event, $p = p_i + p_j + p_k$. Write this in terms of the N's and show that the result is the same as that obtained by equally likely cases using the uniform space. Refer to Problem 15 as a specific example if you need to.

17. Two dice are thrown. Given the information that the number on the first die is even, and the number on the second is < 4, set up an appropriate sample space and answer the following questions.

(a) What are the possible sums and their probabilities?

(b) What is the most probable sum?

(c) What is the probability that the sum is even?

18. Are the following correct nonuniform sample spaces for a throw of two dice? If so, find the probabilities of the given sample points. If not show what is wrong. *Suggestion:* Copy sample space (2.4) and circle on it the regions corresponding to the points of the proposed nonuniform spaces.

(a) First die shows an even number.
First die shows an odd number.

(b) Sum of two numbers on dice is even.
First die is even and second odd.
First die is odd and second even.

(c) First die shows a number ≤ 3.
At least one die shows a number > 3.

19. Consider the set of all permutations of the numbers 1, 2, 3. If you select a permutation at random, what is the probability that the number 2 is in the middle position? In the first position? Do your answers suggest a simple way of answering the same questions for the set of all permutations of the numbers 1 to 7?

3. PROBABILITY THEOREMS

It is not always easy to make direct use of our definitions to calculate probabilities. Definition (1.2) asks us to find a uniform sample space for a problem, that is, a set of all possible *equally likely*, mutually exclusive outcomes of an experiment, and then determine how many of these are favorable to a given event. The definition in Section 2 similarly requires a sample space, that is, a list of the possible outcomes and their probabilities. Such lists may be prohibitively long; we want to consider some theorems which will shorten our work.

Suppose there are 5 black balls and 10 white balls in a box; we draw one ball "at random" (this means we are assuming that each ball has probability $\frac{1}{15}$ of being drawn), and then without replacing the first ball, we draw another. Let us ask for the probability that the first ball is white and the second one is black. The probability of drawing a white ball the first time is $\frac{10}{15}$ (10 of the 15 balls are white). The probability of *then* drawing a black ball is $\frac{5}{14}$ since there are 14 balls left and 5 of them are black. We are going to show that the probability of drawing first a white ball and then (without replacement) a black is the product $\frac{10}{15} \cdot \frac{5}{14}$. We reason in the following way, using a uniform sample space. Imagine that the balls are numbered 1 to 15. The symbol 5,3 will mean that ball 5 was drawn the first time and ball 3 the second time. In such pairs of two (different) numbers representing a drawing of two balls in succession, there are 15 choices for the first number and 14 for the second (the first ball was not replaced). Thus the uniform sample space representing all possible drawings consists of a rectangular array of symbols (like 5,3) with 15 columns (for the 15 different choices for the first number) and 14 rows (for the 14 choices for the second number). Thus there are 15 · 14 points in the sample space. [See also (4.1)]. How many of these sample points correspond to drawing first a white ball and then a black ball? Ten numbers correspond to white balls and the other five to black balls. Thus to obtain a sample point corresponding to drawing first a white and then a black ball, we can choose the first number in 10 ways and then the second number in 5 ways, and so choose the sample point in $10 \cdot 5$ ways; that is, there are $10 \cdot 5$ sample points favorable to the desired drawing. Then by the definition (1.2), the desired probability is $(10 \cdot 5)/(15 \cdot 14)$ as claimed.

Let us state in general the theorem we have just illustrated. We are interested in two successive events A and B. Let $P(A)$ be the probability that A will happen, $P(AB)$ be the probability that both A and B will happen, and $P_A(B)$ be the probability that B will happen if know that A has happened. Then

$$P(AB) = P(A) \cdot P_A(B) \tag{3.1}$$

or in words, the probability of the compound event "*A and B*" is the product of the probability that A will happen times the probability that B will happen if A does. Using the idea of a uniform sample space, we can prove (3.1) by following the method in the ball drawing problem. Let N be the total number of sample points in a uniform sample space, $N(A)$ and $N(B)$ be the numbers of sample points corresponding to the events A and B respectively, and $N(AB)$ be the number of sample points corresponding to the compound event *A and B*. It is useful to picture the sample space geometrically (Figure 3.1) as an array of N points [compare with sample space (2.4)]. We can then circle all points which correspond to A's happening and mark this region A; it contains $N(A)$ points. Similarly, we can circle the $N(B)$ points which correspond to B's happening and call this region B. The overlapping region we call AB; it is part of both A and B and contains $N(AB)$ points which correspond to the compound event *A and B*. Then by the definition (1.2):

$$P(AB) = \frac{N(AB)}{N},$$

$$P(A) = \frac{N(A)}{N}, \tag{3.2}$$

$$P_A(B) = \frac{N(AB)}{N(A)}.$$

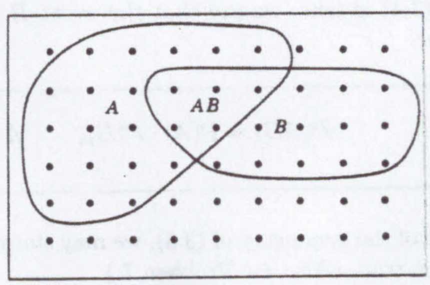


Figure 3.1

Perhaps this last formula for $P_A(B)$ needs some discussion. Recall from Section 2, Example 1, the uniform sample space (2.3) for three tosses of a coin. To find the probability of all tails given that there was at least one tail, we reduced our sample space to seven points (eliminating *hhh*). We then assumed that the seven points of the new sample space had the same relative probability as before the deletion of the point *hhh*; thus each of the seven points had probability $\frac{1}{7}$. (This is no more and no less "obvious" than the original assumption that the eight points had equal probability; it is an additional assumption which we make in the absence of any information to the contrary; see end of Section 2.) Now let us look at the third equation of (3.2). $N(A)$ is the number of sample points corresponding to event A; the N points in the original sample apace all had the same probability so we now assume that when we cross off all the points corresponding to A's *not* happening, the remaining $N(A)$ points also have equal probability. Thus we have a new uniform sample space consisting of $N(A)$ points. $N(AB)$ of these $N(A)$ points correspond to the event B (assuming A). Thus by (1.2), the probability of "B if A" is $N(AB)/N(A)$. From the three equations (3.2), we then have (3.1). In a similar way we can show that

$$P(BA) = P(B) \cdot P_B(A) = P(AB) \tag{3.3}$$

(see Problem 1). [We have proved (3.1) assuming a uniform sample space. This assumption is not necessary; (3.1) is true whether or not we can construct a uniform sample space: see Problem 2.]

Suppose, now, in our example of 5 black and 10 white balls in a box, we draw a ball and replace it and then draw a second ball. The probability of a black ball on the second drawing is then $\frac{5}{15} = \frac{1}{3}$; this is exactly the same result we would get if we had not drawn and replaced the first ball. In the notation of the last paragraph

$$P(B) = P_A(B), \qquad A \text{ and } B \text{ independent.} \tag{3.4}$$

When (3.4) is true, we say that the event B is *independent* of event A and (3.1) becomes

$$P(AB) = P(A) \cdot P(B), \qquad A \text{ and } B \text{ independent.} \tag{3.5}$$

Because of the symmetry of (3.5), we may simply say that A and B are independent if (3.5) is true. (Also see Problem 7.)

Example 1. (a) In three tosses of a coin, what is the probability that all three are heads? We found $p = \frac{1}{8}$ for this problem in Section 2 by seeing that one sample point out of eight corresponds to all heads. Now we can do the problem more simply by saying that the probability of heads on each toss is $\frac{1}{2}$, the tosses are independent, and therefore

$$p = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

(b) If we should want the probability of all heads when a coin is tossed ten times, the sample space would be unwieldy; instead of using the sample space, we can say that since the tosses are independent, the desired probability is $p = (\frac{1}{2})^{10}$.

(c) To find the probability of at least one tail in ten tosses, we see that this event corresponds to all the rest of the sample space except the "all heads" point. Since the sum of the probabilities of all the sample points is 1, the desired probability is $1 - (\frac{1}{2})^{10}$.

In Figure 3.1 or Figure 3.2 the region AB corresponds to the happening of *both* A and B. The whole region consisting of points in A or B or both corresponds to the happening of *either* A *or* B *or* both. We write $P(AB)$ for the probability that both A and B occur. We shall write $P(A + B)$ for the probability that either or both occur. Then we can prove that

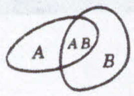


Figure 3.2

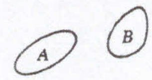


Figure 3.3

$$(3.6) \qquad P(A+B) = P(A) + P(B) - P(AB).$$

To see why this is true, consider Figure 3.2. To find $P(A+B)$ we add the probabilities of all the sample points in the region consisting of A or B or both. But if we add $P(A)$ and $P(B)$, we have included the probabilities of all the sample points in AB twice [once in $P(A)$ and once in $P(B)$]. Thus we must subtract $P(AB)$, which is the sum of the probabilities of all the sample points in AB. This is just what (3.6) says.

If the sample space diagram is like the one in Figure 3.3, so that $P(AB) = 0$, we say that A and B are mutually exclusive. Then (3.6) becomes

$$(3.7) \qquad P(A+B) = P(A) + P(B), \qquad A \text{ and } B \text{ mutually exclusive.}$$

Example 2. Two students are working separately on the same problem. If the first student has probability $\frac{1}{2}$ of solving it and the second student has probability $\frac{3}{4}$ of solving it, what is the probability that at least one of them solves it?

Let A be the event "first student succeeds," and B be the event "second student succeeds." Then $P(AB) = \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$ (assume A and B independent since the students work separately). Then by (3.6) the probability that one or the other or both students solve the problem is

$$P(A+B) = \frac{1}{2} + \frac{3}{4} - \frac{3}{8} = \frac{7}{8}.$$

Conditional Probability; Bayes' Formula If we are asked for the probability of event B assuming that event A occurs [that is, $P_A(B)$], it is often useful to find it from (3.1):

$$(3.8) \qquad P_A(B) = \frac{P(AB)}{P(A)}.$$

Equation (3.8) is called Bayes' formula. In any conditional probability problem to which the answer is not immediately obvious, you should consider whether you can easily find $P(A)$ and $P(AB)$; if so, the conditional probability $P_A(B)$ is given by (3.8).

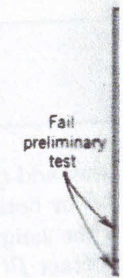
Example 3. A preliminary tes
certain course. The foll

(a) 95% of the students

(b) 96% of the students

(c) 25% of the students

What is the probability
the course?



Let A be the event
The probability we wan
$P(AB)$ is the probability
the course; this is $P(A$
students passed the cou
also want $P(A)$, the pr
event corresponds to th
probabilities of the two
failing test." Then

$P(A)$

(See Figure 3.4; of the 95
nary test; of the 5% of th
test since we are given t

that is, half of the stud
course.

Note that in Figure 3.
test). We are interested
of the original sample sp
sample space (shaded ar
sample space correspond
which we computed.

Example 3. A preliminary test is customarily given to the students at the beginning of a certain course. The following data are accumulated after several years:

(a) 95% of the students pass the course, 5% fail.

(b) 96% of the students who pass the course also passed the preliminary test.

(c) 25% of the students who fail the course passed the preliminary test.

What is the probability that a student who has failed the preliminary test will pass the course?

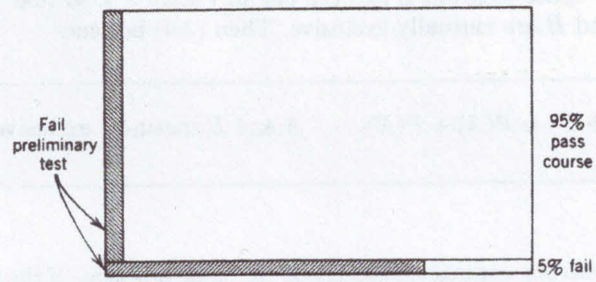


Figure 3.4

Let A be the event "fails preliminary test" and B be the event "Passes course." The probability we want is then $P_A(B)$ in (3.8), so we need $P(AB)$ and $P(A)$. $P(AB)$ is the probability that the student both fails the preliminary test and passes the course; this is $P(AB) = (0.95)(0.04) = 0.038$. (See Figure 3.4; 95% of the students passed the course and of these 4% had failed the preliminary test.) We also want $P(A)$, the probability that a students fails the preliminary test; this event corresponds to the shaded area in Figure 3.4. Thus $P(A)$ is the sum of the probabilities of the two events "passes course after failing test," "fails course after failing test." Then

$$P(A) = (0.095)(0.04) + (0.05)(0.75) = 0.0755$$

(See Figure 3.4; of the 95% of students who passed the course, 4% failed the preliminary test; of the 5% of the students who failed the course, 75% failed the preliminary test since we are given that 25% passed.) By (3.8) we have

$$P_A(B) = \frac{P(AB)}{P(A)} = \frac{0.038}{0.0755} = 50\%,$$

that is, half of the students who fail the preliminary test succeed in passing the course.

Note that in Figure 3.4, the shaded area corresponds to event A (fails preliminary test). We are interested in event B (passes course) given event A. Thus instead of the original sample space (whole rectangle in Figure 3.4) we consider a smaller sample space (shaded area in Figure 3.4). We then want to know what part of this sample space corresponds to event B (passes course). This fraction is $P(AB)/P(A)$ which we computed.

PROBLEMS, SECTION 3

1. (a) Set up a sample space for the 5 black and 10 white balls in a box discussed above assuming the first ball is not replaced. *Suggestions:* Number the balls, say 1 to 5 for black and 6 to 15 for white. Then the sample points form an array something like (2.4), but the point 3,3 for example is not allowed. (Why? What other points are not allowed?) You might find it helpful to write the numbers for black balls and the numbers for white balls in different colors.

(b) Let A be the event "first ball is white" and B be the event "second ball is black." Circle the region of your sample space containing points favorable to A and mark this region A. Similarly, circle and mark region B. Count the number of sample points in A and in B; these are $N(A)$ and $N(B)$. The region AB is the region inside both A and B; the number of points in this region is $N(AB)$. Use the numbers you have found to verify (3.2) and (3.1). Also find $P(B)$ and $P_B(A)$ and verify (3.3) numerically.

(c) Use Figure 3.1 and the ideas of part (b) to prove (3.3) in general.

2. Prove (3.1) for a nonuniform sample space. *Hints:* Remember that the probability of an event is the sum of the probabilities of the sample points favorable to it. Using Figure 3.1, let the points in A but not in AB have probabilities $p_1, p_2, \ldots, p_n$, the points in AB have probabilities $p_{n+1}, p_{n+2}, \ldots, p_{n+k}$, and the points in B but not in AB have probabilities $p_{n+k+1}, p_{n+k+2}, \ldots, p_{n+k+l}$. Find each of the probabilities in (3.1) in terms of the p's and show that you then have an identity.

3. What is the probability of getting the sequence *hhhttt* in six tosses of a coin? If you know the first three are heads, what is the probability that the last three are tails?

4. (a) A weighted coin has probability of $\frac{2}{3}$ of showing heads and $\frac{1}{3}$ of showing tails. Find the probabilities of *hh*, *ht*, *th* and *tt* in two tosses of the coin. Set up the sample space and the associated probabilities. Do the probabilities add to 1 as they should? What is the probability of at least one head? What is the probability of two heads if you know there was at least one head?

(b) For the coin in (a), set up the sample space for three tosses, find the associated probabilities, and use it to answer the questions in Problem 2.12.

5. What is the probability that a number n, $1 \le n \le 99$, is divisible by *both* 6 and 10? By *either* 6 *or* 10 or both?

6. A card is selected from a shuffled deck. What is the probability that it is either a king or a club? That it is both a king and a club?

7. (a) Note that (3.4) assumes $P(A) \neq 0$ since $P_A(B)$ is meaningless if $P(A) = 0$. Assuming both $P(A) \neq 0$ and $P(B) \neq 0$, show that if (3.4) is true, then $P(A) = P_B(A)$; that is if B is independent of A, then A is independent of B. If either $P(A)$ or $P(B)$ is zero, then we use (3.5) to define independence.

(b) When is an event E independent of itself? When is E independent of "not E"?

8. Show that

$$P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

Hint: Start with Figure 3.2 and sketch in a region C overlapping some of the points of each of the regions A, B, and AB.

9. Two cards are drawn at random from a shuffled deck and laid aside without being examined. Then a third card is drawn. Show that the probability that the third card is a spade is $\frac{1}{4}$ just as it was for the first card. *Hint:* Consider all the (mutually exclusive) possibilities (two discarded cards spades, third card spade or not spade, etc.).

10. (a) Three typed letters and their envelopes are piled on a desk. If someone puts the letters into the envelopes at random (one letter in each), what is the probability that each letter gets into its own envelope? Call the envelopes A, B, C, and the corresponding letters a, b, c, and set up the sample space. Note that "a in C, b in B, c in A" is *one* point in the sample space.

(b) What is the probability that at least one letter gets into its own envelope? *Hint:* What is the probability that no letter gets into its own envelope?

(c) Let A mean that a got into envelope A, and so on. Find the probability $P(A)$ that a got into A. Find $P(B)$ and $P(C)$. Find the probability $P(A+B)$ that either a or b or both got into their correct envelopes, and the probability $P(AB)$ that both got into their correct envelopes. Verify equation (3.6).

11. In paying a bill by mail, you want to put your check and the bill (with a return address printed on it) into a window envelope so that the address shows right side up and is not blocked by the check. If you put check and bill at random into the envelope, what is the probability that the address shows correctly?

12. (a) A loaded die has probabilities $\frac{1}{21}, \frac{2}{21}, \frac{3}{21}, \frac{4}{21}, \frac{5}{21}, \frac{6}{21}$, of showing 1, 2, 3, 4, 5, 6. What is the probability of throwing two 3's in succession?

(b) What is the probability of throwing a 4 the first time and not a 4 the second time with a die loaded as in (a)?

(c) If two dice loaded as in (a) are thrown, and we know that the sum of the numbers on the faces is greater than or equal to 10, what is the probability that both are 5's?

(d) How many times must we throw a die loaded as in (a) to have probability greater than $\frac{1}{2}$ of getting an ace?

(e) A die, loaded as in (a), is thrown twice. What is the probability that the number on the die is even the first time > 4 the second time?

13. (a) A candy vending machine is out of order. The probability that you get a candy bar (with or without return of your money) is $\frac{1}{2}$, the probability that you get your money back (with or without candy) is $\frac{1}{3}$, and the probability that you get both the candy and your money back is $\frac{1}{12}$. What is the probability that you get nothing at all? *Suggestion:* Sketch a geometric diagram similar to Figure 3.1, indicate regions representing the various possibilities and their probabilities; then set up a four-point sample space and the associated probabilities of the points.

(b) Suppose you try again to get a candy bar as in part (a). Set up the 16-point sample space corresponding to the possible results of your two attempts to buy a candy bar, and find the probability that you get two candy bars (and no money back); that you get no candy and lose your money both times; that you just get your money back both times.

14. A basketball player succeeds in making a basket 3 tries out of 4. How many tries are necessary in order to have probability > 0.99 of at least one basket?

15. Use Bayes' formula (3.8) to repeat these simple problems previously done by using a reduced sample space.

(a) In a family of two children, what is the probability that both are girls if at least one is a girl?

(b) What is the probability of all heads in three tosses of a coin if you know that at least one is a head?

16. Suppose you have 3 nickels and 4 dimes in your right pocket and 2 nickels and a quarter in your left pocket. You pick a pocket at random and from it select a coin at random. If it is a nickel, what is the probability that it came from your right pocket?

17. (a) There are 3 red and 5 black balls in one box and 6 red and 4 white balls in another. If you pick a box at random, and then pick a ball from it at random, what is the probability that it is red? Black? White? That it is either red or white?

(b) Suppose the first ball selected is red and is not replaced before a second ball is drawn. What is the probability that the second ball is red also?

(c) If both balls are red, what is the probability that they both came from the same box?

18. Two cards are drawn at random from a shuffled deck.

(a) What is the probability that at least one is a heart?

(b) If you know that at least one is a heart, what is the probability that both are hearts?

19. Suppose it is known that 1% of the population have a certain kind of cancer. It is also known that a test for this kind of cancer is positive in 99% of the people who have it but is also positive in 2% of the people who do not have it. What is the probability that a person who tests positive has cancer of this type?

20. Some transistors of two different kinds (call them N and P) are stored in two boxes. You know that there are 6 N's in one box and that 2 N's and 3 P's got mixed in the other box, but you don't know which box is which. You select a box and a transistor from it at random and find that it is an N; what is the probability that it came from the box with the 6 N's? From the other box? If another transistor is picked from the same box as the first, what is the probability that it is also an N?

21. Two people are taking turns tossing a pair of coins; the first person to toss two alike wins. What are the probabilities of winning for the first player and for the second player? *Hint:* Although there are an infinite number of possibilities here (win on first turn, second turn, third turn, etc.), the sum of the probabilities is a geometric series which can be summed; see Chapter 1 if necessary.

22. Repeat Problem 21 if the players toss a pair of dice trying to get a double (that is, both dice showing the same number).

23. A thick coin has probability $\frac{3}{7}$ of falling heads, $\frac{3}{7}$ of falling tails, and $\frac{1}{7}$ of standing on edge. Show that if it is tossed repeatedly it has probability 1 of eventually standing on edge.

4. METHODS OF COUNTING

Let us digress for a bit to review some ideas and formulas we need in computing probabilities in more complicated problems.

Let us ask how many two-digit numbers have either 5 or 7 for the tens digit and either 3, 4, or 6 for the units digit. The answer becomes obvious if we arrange the possible numbers in a rectangle

$$\begin{matrix} 53 & 54 & 56 \\ 73 & 74 & 76 \end{matrix}$$

with two rows corresponding to the two choices of the tens digit and three columns corresponding to the three choices of the units digit. This is an example of the *fundamental principle of counting*:

> (4.1) **If one thing can be done N_1 ways, and after that a second thing can be done in N_2 ways, the two things can be done in succession in that order in $N_1 \cdot N_2$ ways. This can be extended to doing any number of things one after the other, the first N_1 ways, the second N_2 ways, the third N_3 ways, etc. Then the total number of ways to perform the succession of acts is the product $N_1 N_2 N_3 \cdots$.**

Now consider a set of n things lined up in a row; we ask how many ways we can arrange (permute) them. This result is called the number of *permutations* of n things n at a time, and is denoted by ${}_nP_n$ or $P(n,n)$ or P_n^n. To find this number, we think of seating n people in a row of n chairs. We can place anyone in the first chair, that is, we have n possible ways of filling the first chair. Once we have selected someone for the first chair, there are $(n-1)$ choices left for the second chair, then $(n-2)$ choices for the third chair, and so on. Thus by the fundamental principle, there are $n(n-1)(n-2)\cdots 2 \cdot 1 = n!$ ways of arranging the n people in the row of n chairs. The number of permutations of n things n at a time is

$$P(n,n) = n!. \tag{4.2}$$

Next suppose there are n people but only $r < n$ chairs and we ask how many ways we can select groups of r people and seat them in the r chairs. The result is called the number of permutations of n things r at a time and is denoted by ${}_nP_r$ or $P(n,r)$ or P_r^n. Arguing as before, we find that there are n ways to fill the first chair, $(n-1)$ ways to fill the second chair, $(n-2)$ ways for the third [note that we could write $(n-2)$ as $(n-3+1)$], etc., and finally $(n-r+1)$ ways of filling chair r. Thus we have for the number of permutations of n things r at a time

$$P(n,r) = n(n-1)(n-2)\cdots(n-r+1).$$

By multiplying and dividing by $(n-r)!$ we can write this as

$$P(n,r) = n(n-1)(n-2)\cdots(n-r+1)\frac{(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}. \tag{4.3}$$

So far we have been talking about arranging things in a definite order. Suppose, instead that we ask how many committees of r people can be chosen from a group of n people $(n \geq r)$. Here the order of the people in the committee is not considered; the committee made up of people A, B, C, is the same as the committee made up of people B, A, C. We call the number of such committees of r people which we can select from n people, the number of *combinations* or *selections* of n things r at a time, and denote this number by ${}_nC_r$ or $C(n,r)$ or $\binom{n}{r}$. To find $C(n,r)$, we go back to the problem of selecting r people from a group of n and seating them in r chairs; we found that the number of ways of doing this is $P(n,r)$ as given in

(4.3). We can perform this job by first selecting r people from the total n and then arranging the r people in r chairs. The selection of r people can be done in $C(n, r)$ ways (this is the number we are trying to find), and after r people are selected, they can be arranged in r chairs in $P(r, r)$ ways by (4.2). By the fundamental principle (4.1), the total number of ways $P(n, r)$ of selecting and seating r people out of n is the product $C(n, r) \cdot P(r, r)$. Thus we have

$$P(n, r) = C(n, r) \cdot P(r, r). \tag{4.4}$$

We can solve this equation to find the value $C(n, r)$ which we wanted. Substituting the values of $P(n, r)$ and $P(r, r)$ from (4.3) and (4.2) into (4.4) and solving for $C(n, r)$, we find for the number of combinations of n things r at a time

$$C(n, r) = \frac{P(n, r)}{P(r, r)} = \frac{n!}{(n-r)!r!} = \binom{n}{r}. \tag{4.5}$$

Each time we select r people to be seated, we leave $n - r$ people without chairs. Then there are exactly the same number of combinations of n things $n - r$ at a time as there are combinations of n things r at a time. Hence we write

$$C(n, n-r) = C(n, r) = \frac{n!}{(n-r)!r!}. \tag{4.6}$$

We can also obtain (4.6) from (4.5) by replacing r by $(n - r)$.

Example 1. A club consists of 50 members. In how many ways can a president, vice-president, secretary, and treasurer be chosen? In how many ways can a committee of 4 members be chosen?

In the selection of officers, we must not only select 4 people, but decide which one is president, etc.; we could think of seating the 4 people in chairs labeled president, vice-president, etc. Thus the number of ways of selecting the officers is

$$P(50, 4) = \frac{50!}{(50-4)!} = \frac{50!}{46!} = 50 \cdot 49 \cdot 48 \cdot 47.$$

The committee members, however, are all equivalent (we are neglecting the possibility that one is named chairman), so the number of ways of selecting committees of 4 people is

$$C(50, 4) = \frac{50!}{46!4!} = \frac{50 \cdot 49 \cdot 48 \cdot 47}{24}.$$

Example 2. Find the coefficient of x^8 in the binomial expansion of $(1 + x)^{15}$.

Think of multiplying out

$$(1+x)(1+x)(1+x)\cdots(1+x), \quad \text{(with 15 factors)}.$$

We obtain a term in x^8 each time we multiply 1's from seven of the parentheses by x's from eight of the parentheses. The number of ways of selecting 8 parentheses out of 15 is

$$C(15, 8) = \frac{15!}{8!7!}.$$

This is the desired coefficient of x^8.

Generalizing this example, we see that in the expansion of $(a+b)^n$, the coefficient of $a^{n-r}b^r$ is $C(n,r)$, usually written $\binom{n}{r}$ when used in connection with a binomial expansion (see Chapter 1, Section 13C). Thus the expressions $C(n,r)$ are just the binomial coefficients, and we can write

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^{n-r} b^r. \tag{4.7}$$

Example 3. A basic problem in statistical mechanics is this: Given N balls, and n boxes, in how many ways can the balls be put into the boxes so that there will be given numbers of balls in the boxes, say N_1 balls in the first box, N_2 balls in the second box, N_3 in the third, $\cdots$, N_n in the nth, and what is the probability that this given distribution will occur when the balls are put into the boxes? In statistical mechanics the "balls" may be molecules, electrons, photons, etc., and each "box" corresponds to a small range of values of position and momentum of a particle. We can state many other problems in this same language of putting balls into boxes. For example, in tossing a coin, we can equate heads with box 1, and tails with box 2; in tossing a die, there are six "boxes." In putting letters into envelopes, the letters are the balls, and the envelopes are the boxes. In dealing cards, the cards are the balls and the players who receive them are the boxes. In an alpha scattering experiment, the alpha particles are the balls, and the boxes are elements of area on the detecting screen which the particles hit after they are scattered. (Also see Problems 14 and 21 and Feller, pp. 10–11.)

Let us do a special case of this problem in which we have 15 balls and 6 boxes, and the numbers of balls we are to put into the various boxes are:

Number of balls:	3	1	4	2	3	2
In box number:	1	2	3	4	5	6

We first ask how many ways we can select 3 balls to go in the first box from the 15 balls; this is $C(15,3)$. (Note that the order of the balls in the boxes is not considered; this is like the committee problem in Example 1.) Now we have 12 balls left, of which we are to select 1 for box 2; we can do this in $C(12,1)$ ways. We can then select the 4 balls for box 3 from the remaining 11 balls in $C(11,4)$ ways, the 2 balls for box 4 in $C(7,2)$ ways, the 3 balls for box 5 in $C(5,3)$ ways, and finally the balls for box 6 in $C(2,2)$ ways (verify that this is 1). By the fundamental principle, the total number of ways of putting the required numbers of balls into the boxes is

$$\begin{aligned} &C(15,3)\cdot C(12,1)\cdot C(11,4)\cdot C(7,2)\cdot C(5,3)\cdot C(2,2) \\ &= \frac{15!}{3!\cdot 12!}\cdot\frac{12!}{1!\cdot 11!}\cdot\frac{11!}{4!\cdot 7!}\cdot\frac{7!}{2!\cdot 5!}\cdot\frac{5!}{3!\cdot 2!}\cdot\frac{2!}{2!\cdot 0!} \\ &= \frac{15!}{3!\cdot 1!\cdot 4!\cdot 2!\cdot 3!\cdot 2!}. \end{aligned}$$

(Remember from Chapters 1 and 11 that $0! = 1$.)

Next we want the probability of this particular distribution. Let us assume that the balls are distributed "at random" into the boxes; by this we mean that a ball has the same probability (namely $\frac{1}{6}$) of being put into any one box as into any other box. We can put the first ball into any one of the 6 boxes, the second ball into any one of the 6 boxes, and so on. Thus by the fundamental principle, the total number of ways of distributing the 15 balls into the 6 boxes is $6 \cdot 6 \cdot 6 \cdot 6 \cdots 6 = 6^{15}$ and we are assuming that these distributions are equally probable. Then the probability that, when 15 balls are distributed "at random" into 6 boxes, there will be 3 balls in box 1, 1 in box 2, etc., as given, is, by (1.2) (favorable cases ÷ total)

$$\frac{15!}{3!\cdot 1!\cdot 4!\cdot 2!\cdot 3!\cdot 2!} \div 6^{15}.$$

Example 4. In Example 3, we assumed that the 6^{15} possible distributions of 15 balls into 6 boxes were equally likely. This seems very reasonable if we think of putting the balls into the boxes by tossing a die for each ball; if the die shows 1 we put the ball into box 1, etc. However, we can think of situations to which this method and result do not apply. For example, suppose we are putting letters into envelopes or seating people in chairs; then we may reasonably require only one letter per envelope, not more than one person per chair, that is, one ball (or none) per box. Consider the problem of seating 4 people in 6 chairs, that is of putting 4 balls into 6 boxes. If we number the chairs from 1 to 6 and let each person choose a chair by tossing a die, we may have two or more people choosing the same chair. The result 6^4 (which the method of Example 3 gives for the problem of 4 balls in 6 boxes) then does not apply to this problem. However, let us consider the uniform sample space of 6^4 points and select from it the points corresponding to our restriction (one ball or none per box). The new sample space contains $C(6,4)\cdot 4!$ points (number of ways of selecting the 4 chairs to be occupied times the number of ways of then arranging 4 people in 4 chairs). Since these points were equally probable in the original (uniform) sample space, we still consider them equally probable. Now let us ask for the probability that the first two chairs are vacant when the 4 people are seated. The number of sample points corresponding to this event is 4! (the number of ways of arranging the 4 people in the last 4 chairs). Thus the desired probability is

$$\frac{4!}{C(6,4)\cdot 4!} = \frac{1}{C(6,4)}.$$

We can now see an easier way of doing problems of this kind. The factor 4!, which canceled in the probability calculation, was the number of rearrangements of the 4 people among the 4 occupied chairs. Since this is the same for any given set of 4 chairs, we can lump together all the sample points corresponding to each given set of 4 chairs, and have a smaller (still uniform) sample space of $C(6,4)$ points. Each point now corresponds to a given set of 4 occupied chairs; the quantity $C(6,4)$ is just the number of ways of picking 4 occupied chairs out of 6. The probability that the first two chairs are vacant when 4 people are seated is $1/C(6,4)$ since there is only one way to select 4 occupied chairs leaving the first two chairs vacant.

Another useful way of looking at this problem is to consider a set of 4 *identical* balls to be put into 6 boxes. Since the balls are identical, the 4! arrangements of the 4 balls in 4 given boxes all look alike. We can say that there are $C(6,4)$ *distinguishable* arrangements of the 4 identical balls in 6 boxes (one ball or none per box). Since all these arrangements are equally probable, the probability of any one arrangement (say the first two boxes empty) is $1/C(6,4)$ as we found previously.

Example 5. In Example 4 w
ular boxes were empty
was true because the
Without the restrictio
are not equally probal
ple, the probability o
of no balls in the firs
$4! \div 6^4 = \frac{1}{54}$. We see t
box) are less probable
Now we are going
rangements *are* equall
and the 4 balls are pe
if the people are frien
and the probabilities
the concentrated arra
model. (This is a mo
to 6, and 4 balls. Fro
ball in the box numbe
also add another card
the number first draw
in the corresponding
8 cards. We repeat t
Then the probability
of one ball in each of
probability that the fi
this the probability th
are 4! such possibilitie
the distributions "all
equally probable. Furt
arrangements are equa
To find the number
ture of the 4 balls in t

Box number:
Number of ba

The lines mean the s
requires 7 lines to pic
arrangements of the 4
the beginning and at t
can be arranged in any
of the balls in the boxes
arrangements is just th
out of 9 positions for t
arrangements in this p

We see then that p
must say *how* we prop
what practical problem
space and the probabili

Example 5. In Example 4 we found the same answer for the probability that two particular boxes were empty whether or not we considered the balls distinguishable. This was true because the allowed distinguishable arrangements were equally probable. Without the restriction of one ball or none per box, all distinguishable arrangements are not equally probable according to the methods of Examples 3 and 4. For example, the probability of all balls in box 1 is $1/6^4$; compare this with the probability of no balls in the first 2 boxes and one ball in each of the other 4 boxes, which is $4! \div 6^4 = \frac{1}{54}$. We see that the concentrated arrangements (all or several balls in one box) are less probable than the more uniform arrangements.

Now we are going to try to imagine a situation in which *all* distinguishable arrangements *are* equally probable. Suppose the 6 boxes are benches in a waiting room and the 4 balls are people who are going to come in and sit on the benches. Then if the people are friends, there will be a certain tendency for them to sit together and the probabilities we have been calculating will not apply—the probabilities of the concentrated arrangements will increase. Consider the following mathematical model. (This is a modification of Pólya's urn model.) We have 6 boxes labeled 1 to 6, and 4 balls. From 6 cards labeled 1 to 6 we draw one at random and place a ball in the box numbered the same as the card drawn. We then replace the card and also add another card of the same number so that there are now 7 cards, two with the number first drawn. We now select a card at random from these 7, put a ball in the corresponding box and again replace the card adding a duplicate to make 8 cards. We repeat this process two more times (until all balls are distributed). Then the probability that all balls are in box 1 is $\frac{1}{6} \cdot \frac{2}{7} \cdot \frac{3}{8} \cdot \frac{4}{9}$. The probability of one ball in each of the first 4 boxes is $\frac{1}{6} \cdot \frac{1}{7} \cdot \frac{1}{8} \cdot \frac{1}{9} \cdot 4!$ (here $\frac{1}{6} \cdot \frac{1}{7} \cdot \frac{1}{8} \cdot \frac{1}{9}$ is the probability that the first ball is in box 1, the second in box 2, etc.; we must add to this the probability that the first ball is in box 3, the second in box 1, etc.; there are 4! such possibilities all giving one ball in each of the first 4 boxes). We see that the distributions "all balls in box 1" and "one ball in each of the first 4 boxes" are equally probable. Further calculation (Problem 20) shows that all distinguishable arrangements are equally probable.

To find the number of distinguishable arrangements, consider the following picture of the 4 balls in the 6 boxes.

```
                     | o |   | o o |   | o |   |
Box number:            1   2    3    4   5   6
Number of balls:       1   0    2    0   1   0
```

The lines mean the sides of the boxes and the circles are the balls; note that it requires 7 lines to picture the 6 boxes. This picture shows one of many possible arrangements of the 4 balls in 6 boxes. In any such picture there must be a line at the beginning and at the end, but the rest of the lines (5 of them) and the 4 circles can be arranged in any order. You should convince yourself that every arrangement of the balls in the boxes can be so pictured. Then the number of such distinguishable arrangements is just the number of ways we can select 4 positions for the 4 circles out of 9 positions for the 5 lines and 4 circles. Thus there are $C(9, 4)$ equally likely arrangements in this problem.

We see then that putting balls in boxes is not quite as simple as we thought; we must say *how* we propose to distribute them and even before that we must think what practical problem we are trying to solve; this is what determines the sample space and the probabilities to be associated with the sample points. Unfortunately, it may not always be clear what the sample space probabilities should be: then the best we can do is to try various assumptions. In statistical mechanics it is found that certain particles (for example, the molecules of a gas) are correctly described if we assume that they behave like the balls of Example 3 (all 6^{15} arrangements equally likely); we then say that they obey Maxwell-Boltzmann statistics. Other particles (for example, electrons) behave like the people to be seated in Example 4 (one particle or none per box); we say that such particles obey Fermi-Dirac statistics. Finally some particles (for example, photons) act something like the friends who want to sit near each other (all distinguishable arrangements of identical particles are equally likely); we say that these particles obey Bose-Einstein statistics. For the problem of 4 particles in 6 boxes, there are then 6^4 equally likely arrangements for Maxwell-Boltzmann particles, $C(6,4)$ for Fermi-Dirac particles, and $C(9,4)$ for Bose-Einstein particles. (See Problems 15 to 20.)

PROBLEMS, SECTION 4

1. (a) There are 10 chairs in a row and 8 people to be seated. In how many ways can this be done?

(b) There are 10 questions on a test and you are to do 8 of them. In how many ways can you choose them?

(c) In part (a) what is the probability that the first two chairs in the row are vacant?

(d) In part (b), what is the probability that you omit the first two problems in the test?

(e) Explain why the answer to parts (a) and (b) are different, but the answers to (c) and (d) are the same.

2. In the expansion of $(a+b)^n$ (see Example 2), let $a = b = 1$, and interpret the terms of the expansion to show that the total number of combinations of n things taken $1, 2, 3, \cdots, n$ at a time, is $2^n - 1$.

3. A bank allows one person to have only one savings account insured to \$100,000. However, a larger family may have accounts for each individual, and also accounts in the names of any 2 people, any 3 and so on. How many accounts are possible for a family of 2? Of 3? Of 5? Of n? *Hint:* See Problem 2.

4. Five cards are dealt from a shuffled deck. What is the probability that they are all of the same suit? That they are all diamond? That they are all face cards? That the five cards are a sequence in the same suit (for example, 3, 4, 5, 6, 7 of hearts)?

5. A bit (meaning binary digit) is 0 or 1. An ordered array of eight bits (such as 01101001) is a byte. How many different bytes are there? If you select a byte at random, what is the probability that you select 11000010? What is the probability that you select a byte containing three 1's and five 0's?

6. A so-called 7-way lamp has three 60-watt bulbs which may be turned on one or two or all three at a time, and a large bulb which may be turned to 100 watts, 200 watts or 300 watts. How many different light intensities can the lamp be set to give if the completely off position is not included? (The answer is *not* 7.)

7. What is the probability that the 2 and 3 of clubs are next to each other in a shuffled deck? *Hint:* Imagine the two cards accidentally stuck together and shuffled as one card.

8. Two cards are d… aces? If you kno… aces? If you knc… are aces?

9. Two cards are d… red? If at least o… is a red ace, wha… what is the prob…

10. What is the prob… plicity, let a year… three different bi… birthdays is

$p =$

Estimate this for… for $x \ll 1$]. Find… group of 23 peopl… the same birthda… presidents of the…

11. The following gan… on each license pl… the same last two… must you observe… with the same last…

12. Consider Problem… people for which t… same month?

13. Generalize Examp… with N_1 in box 1,…

14. (a) Find the pro… That in six t… a 12-sided di… faces show up…

(b) The last prob… n balls are di… ball. Show th…

15. Set up the uniform… for Maxwell-Boltzn… particles. See Exan… 6 for BE.)

16. Do Problem 15 for… find the probability… (You should find th…

17. Find the number of… kinds of statistics.

ılities should be; then the mechanics it is found that correctly described if we 6^{15} arrangements equally statistics. Other particles ated in Example 4 (one ey Fermi-Dirac statistics. ning like the friends who ents of identical particles -Einstein statistics. For ually likely arrangements particles, and $C(9,4)$ for

ed. In how many ways can

8 of them. In how many

wo chairs in the row are

first two problems in the

erent, but the answers to

, and interpret the terms nations of n things taken

unt insured to $100,000. vidual, and also accounts accounts are possible for

bability that they are all are all face cards? That 3, 4, 5, 6, 7 of hearts)?

y of eight bits (such as If you select a byte at What is the probability

be turned on one or two to 100 watts, 200 watts mp be set to give if the 7.)

each other in a shuffled gether and shuffled as

8. Two cards are drawn from a shuffled deck. What is the probability that both are aces? If you know that at least one is an ace, what is the probability that both are aces? If you know that one is the ace of spades, what is the probability that both are aces?

9. Two cards are drawn from a shuffled deck. What is the probability that both are red? If at least one is red, what is the probability that both are red? If at least one is a red ace, what is the probability that both are red? If exactly one is a red ace, what is the probability that both are red?

10. What is the probability that you and a friend have different birthdays? (For simplicity, let a year have 365 days.) What is the probability that three people have three different birthdays? Show that the probability that n people have n different birthdays is

$$p = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\left(1 - \frac{3}{365}\right)\cdots\left(1 - \frac{n-1}{365}\right).$$

Estimate this for $n \ll 365$ by calculating $\ln p$ [recall that $\ln(1+x)$ is approximately x for $x \ll 1$]. Find the smallest (integral) n for which $p < \frac{1}{2}$. Hence, show that for a group of 23 people or more, the probability is greater than $\frac{1}{2}$ that two of them have the same birthday. (Try it with a group of friends or a list of people such as the presidents of the United States.)

11. The following game was being played on a busy street: Observe the last two digits on each license plate. What is the probability of observing at least two cars with the same last two digits among the first 5 cars? 10 cars? 15 cars? How many cars must you observe in order for the probability to be greater than $\frac{1}{2}$ of observing two with the same last two digits?

12. Consider Problem 10 for different months of birth. What is the smallest number of people for which the probability is greater than $\frac{1}{2}$ that two of them were born in the same month?

13. Generalize Example 3 to show that the number of ways of putting N balls in n boxes with N_1 in box 1, N_2 in box 2, etc., is

$$\left(\frac{N!}{N_1! \cdot N_2! \cdot N_3! \cdots N_n!}\right).$$

14. (a) Find the probability that in two tosses of a coin, one is heads and one tails. That in six tosses of a die, all six of the faces show up. That in 12 tosses of a 12-sided die, all 12 faces show up. That in n tosses of an n-sided die, all n faces show up.

(b) The last problem in part (a) is equivalent to finding the probability that, when n balls are distributed at random into n boxes, each box contains exactly one ball. Show that for large n, this is approximately $e^{-n}\sqrt{2\pi n}$.

15. Set up the uniform sample spaces for the problem of putting 2 particles in 3 boxes: for Maxwell-Boltzmann particles, for Fermi-Dirac particles, and for Bose-Einstein particles. See Example 5. (You should find 9 sample points for MB, 3 for FD, and 6 for BE.)

16. Do Problem 15 for 2 particles in 2 boxes. Using the model discussed in Example 5, find the probability of each of the three sample points in the Bose-Einstein case. (You should find that each has probability $\frac{1}{3}$, that is, they are equally probable.)

17. Find the number of ways of putting 2 particles in 4 boxes according to the three kinds of statistics.

18. Find the number of ways of putting 3 particles in 5 boxes according to the three kinds of statistics.

19. (a) Following the methods of Examples 3, 4, and 5, show that the number of equally likely ways of putting N particles in n boxes, $n > N$, is n^N for Maxwell-Boltzmann particles, $C(n, N)$ for Fermi-Dirac particles, and $C(n-1+N, N)$ for Bose-Einstein particles.

(b) Show that if n is much larger than N (think , for example, of $n = 10^6$, $N = 10$), then both the Bose-Einstein and the Fermi-Dirac results in part (a) contain products of N numbers, each number approximately equal to n. Thus show that for $n \gg N$, both the BE and the FD results are approximately equal to $n^N/N!$, which is $1/N!$ times the MB result.

20. (a) In Example 5, a mathematical model is discussed which claims to give a distribution of identical balls into boxes in such a way that all distinguishable arrangements are equally probable (Bose-Einstein statistics). Prove this by showing that the probability of a distribution of N balls into n boxes (according to this model) with N_1 balls in the first box, N_2 in the second, $\cdots$, N_n in the nth, is $1/C(n-1+N, N)$ for any set of numbers N_i such that $\sum_{i=1}^{n} N_i = N$.

(b) Show that the model in (a) leads to Maxwell-Boltzmann statistics if the drawn card is replaced (but no extra card added) and to Fermi-Dirac statistics if the drawn card is not replaced. *Hint:* Calculate in each case the number of possible arrangements of the balls in the boxes. First do the problem of 4 particles in 6 boxes as in the example, and then do N particles in n boxes ($n > N$) to get the results in Problem 19.

21. The following problem arises in quantum mechanics (see Chapter 13, Problem 7.21). Find the number of ordered triples of nonnegative integers a, b, c whose sum $a+b+c$ is a given positive integer n. (For example, if $n = 2$, we could have $(a, b, c) = (2, 0, 0)$ or $(0, 2, 0)$ or $(0, 0, 2)$ or $(0, 1, 1)$ or $(1, 0, 1)$ or $(1, 1, 0)$.) *Hint:* Show that this is the same as the number of distinguishable distributions of n identical balls in 3 boxes, and follow the method of the diagram in Example 5.

22. Suppose 13 people want to schedule a regular meeting one evening a week. What is the probability that there is an evening when everyone is free if each person is already busy one evening a week?

23. Do Problem 22 if one person is busy 3 evenings, one is busy 2 evenings, two are each busy one evening, and the rest are free every evening.

5. RANDOM VARIABLES

In the problem of tossing two dice (Example 2, Section 2), we may be more interested in the value of the sum of the numbers on the two dice than we are in the individual numbers. Let us call this sum x; then for each point of the sample space in (2.4), x has a value. For example, for the point 2,1, we have $x = 2+1 = 3$; for the point 6,2, we have $x = 8$, etc. Such a variable, x, which has a definite value for each sample point, is called a *random variable*. We can easily construct many more examples of random variables for the sample space (2.4); here are a few (Can you construct some more?):

$$x = \text{number on first die minus number on second;}$$

$$x = \text{number on second die;}$$

$$x = \text{probability } p \text{ associated with the sample point;}$$

$$x = \begin{cases} 1 \text{ if the sum is 7 or 11,} \\ 0 \text{ otherwise.} \end{cases}$$

For each of these random variables x, we could set up a table listing all the sample points in (2.4) and, next to each sample point, the corresponding value of x. This table may remind you of the tables of values we could use in plotting the graph of a function. In analytical geometry or in a physics problem, knowing x as a function of t means that for any given t we can find the corresponding value of x. In probability the sample point corresponds to the independent variable t; given the sample point, we can find the corresponding value of the random variable x if we are given a description of x (for example, $x =$ the sum of numbers on dice). The "description" corresponds to the formula $x(t)$ that we use in plotting a graph in analytic geometry. Thus we may say that a *random variable x is a function defined on a sample space.*

Probability Functions Let us consider further the random variable $x =$ "sum of numbers on dice" for a toss of two dice [sample space (2.4)]. We note that there are several sample points for which $x = 5$, namely the points marked a in (2.4). Similarly, there are several sample points for most of the other values of x. It is then convenient to lump together all the sample points corresponding to a given value of x, and consider a new sample space in which each point corresponds to one value of x; this is the sample space (2.5). The probability associated with each point of the new sample space is obtained as in Section 2, by adding the probabilities associated with all the points in the original sample space corresponding to the particular value of x. Each value of x, say x_i, has a probability p_i of occurrence; we may write $p_i = f(x_i) =$ probability that $x = x_i$, and call the function $f(x)$ the *probability function* for the random variable x. In (2.5) we have listed on the first line the values of x and on the second line the values of $f(x)$. [In this problem, x and $f(x)$ take on only a finite number of discrete values; in some later problems they will take on a continuous set of values.] We could also exhibit these values graphically (Figure 5.1).

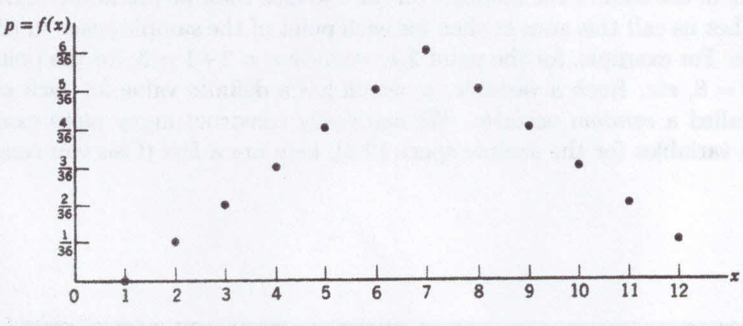


Figure 5.1

Now that we have the table of values (2.5) or the graph (Figure 5.1) to describe the random variable x and its probability function $f(x)$, we can dispense with the original sample space (2.4). But since we used (2.4) in defining what is meant by a random variable, let us now give another definition using (2.5) or Figure 5.1. We can say that *x is a random variable if it takes various values x_i with probabilities $p_i = f(x_i)$*. This definition may explain the name random variable; x is called a variable since it takes various values. A random (or stochastic) process is one whose outcome is not known in advance. The way the two dice fall is such an unknown outcome, so the value of x is unknown in advance, and we call x a *random* variable.

You may note that at first we thought of x as a dependent variable or function with the sample point as the independent variable. Although we didn't say much about it, there was also a value of the probability p attached to each sample point, that is p and x were both functions of the sample point. In the last paragraph, we have thought of x as an independent variable with p as a function of x. This is quite analogous to having both x and p given as functions of t and eliminating t to obtain p as a function of x. We have here eliminated the sample point from the forefront of our discussion in order to consider directly the probability function $p = f(x)$.

Example 1. Let x = number of heads when three coins are tossed. The uniform sample space is (2.3) and we could write the value of x for each sample point in (2.3). Instead, let us go immediately to a table of x and $p = f(x)$. [Can you verify this table by using (2.3), or otherwise?]

(5.1)

x	0	1	2	3
$p = f(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Other terms used for the probability function $p = f(x)$ are: *probability density function*, *frequency function*, or *probability distribution* (*caution*: **not** distribution function, which means the *cumulative distribution* as we will discuss later; see Figure 5.2). The origins of these terms will become clearer as we go on (Sections 6 and 7) but we can get some idea of the terms frequency and distribution from (5.1). Suppose we toss three coins repeatedly; we might reasonably expect to get three heads in about $\frac{1}{8}$ of the tosses, two heads in about $\frac{3}{8}$ of the tosses, etc. That is, each value of $p = f(x)$ is proportional to the *frequency* of occurrence of that value of x—hence the term *frequency function* (see also Section 7). Again in (5.1), imagine four boxes labeled $x = 0, 1, 2, 3$, and put a marble into the appropriate box for each toss of three coins. Then $p = f(x)$ indicates approximately how the marbles are distributed into the boxes after many tosses—hence the term *distribution*.

Mean Value; Standard Deviation The probability function $f(x)$ of a random variable x gives us detailed information about it, but for many purposes we want a simpler description. Suppose, for example, that x represents experimental measurements of the length of a rod, and that we have a large number N of measurements x_i. We might reasonably take $p_i = f(x_i)$ proportional to the number of times N_i we obtained the value x_i, that is $p_i = N_i/N$. We are especially interested in two numbers, namely a mean or average value of all our measurements, and some number which indicates how widely the original set of values spreads out about that average. Let us define two such quantities which are customarily used to describe a random variable. To calculate the average of a set of N numbers, we add them and

divide by N. Instea
each measurement
for the average of t

By analogy with th
random variable x

(5.2)

To obtain a meas
first list how much
deviations are posit
(Problem 10). Inste
define the *variance*

(5.3)

(The variance is some
are very close to μ,
Var(x) is large. Thu
surements; this is wh
deviation of x, is ofte

(5.4)

Example 2. For the data in
By (5.2), μ = average
By (5.3), Var(x) = (0
$= \frac{9}{4}$
By (5.4), σ_x = standar

The mean or avera
or its *expected value*
Instead of μ, the symb
of x.

ph (Figure 5.1) to describe
we can dispense with the
lefining what is meant by
ig (2.5) or Figure 5.1. We
alues x_i *with probabilities*
om variable: x is called a
stic) process is one whose
fall is such an unknown
call x a *random* variable.
dent variable or function
ough we didn't say much
ied to each sample point,
n the last paragraph, we
a function of x. This is
s of t and eliminating t
the sample point from
the probability function

ed. The uniform sample
sample point in (2.3).
r). [Can you verify this

are: *probability density*
ation: not distribution
ll discuss later; see Fig-
s we go on (Sections 6
distribution from (5.1).
oly expect to get three
e tosses, etc. That is,
ccurrence of that value
. Again in (5.1), imag-
he appropriate box for
ately how the marbles
term *distribution*.

nction $f(x)$ of a ran-
for many purposes we
presents experimental
ge number N of mea-
onal to the number of
e especially interested
asurements, and some
preads out about that
ily used to describe a
ers, we add them and

divide by N. Instead of adding the large number of measurements, we can multiply each measurement by the number of times it occurs and add the results. This gives for the average of the measurements, the value

$$\frac{1}{N}\cdot\sum_i N_i x_i = \sum_i p_i x_i.$$

By analogy with this calculation, we now define the *average* or *mean value* μ *of a random variable* x whose probability function is $f(x)$ by the equation

$$\mu = \text{average of } x = \sum_i x_i p_i = \sum_i x_i f(x_i). \tag{5.2}$$

To obtain a measure of the spread or dispersion of our measurements, we might first list how much each measurement differs from the average. Some of these deviations are positive and some are negative; if we average them, we get zero (Problem 10). Instead, let us square each deviation and average the squares. We define the *variance* of a random variable x by the equation

$$\operatorname{Var}(x) = \sum_i (x_i - \mu)^2 f(x_i). \tag{5.3}$$

(The variance is sometimes called the dispersion.) If nearly all the measurements x_i are very close to μ, then $\operatorname{Var}(x)$ is small; if the measurements are widely spread, $\operatorname{Var}(x)$ is large. Thus we have a number which indicates the spread of the measurements; this is what we wanted. The square root of $\operatorname{Var}(x)$, called the *standard deviation* of x, is often used instead of $\operatorname{Var}(x)$:

$$\sigma_x = \text{standard deviation of } x = \sqrt{\operatorname{Var}(x)}. \tag{5.4}$$

Example 2. For the data in (5.1) we can compute:
By (5.2), $\mu = \text{average of } x = 0\cdot\frac{1}{8} + 1\cdot\frac{3}{8} + 2\cdot\frac{3}{8} + 3\cdot\frac{1}{8} = \frac{12}{8} = \frac{3}{2}$.
By (5.3), $\operatorname{Var}(x) = \left(0-\frac{3}{2}\right)^2\cdot\frac{1}{8} + \left(1-\frac{3}{2}\right)^2\cdot\frac{3}{8} + \left(2-\frac{3}{2}\right)^2\cdot\frac{3}{8} + \left(3-\frac{3}{2}\right)^2\cdot\frac{1}{8}$
$= \frac{9}{4}\cdot\frac{1}{8} + \frac{1}{4}\cdot\frac{3}{8} + \frac{1}{4}\cdot\frac{3}{8} + \frac{9}{4}\cdot\frac{1}{8} = \frac{3}{4}$.
By (5.4), $\sigma_x = \text{standard deviation of } x = \sqrt{\operatorname{Var}(x)} = \frac{1}{2}\sqrt{3}$.

The mean or average value of a random variable x is also called its *expectation* or its *expected value* or (especially in quantum mechanics) its *expectation value*. Instead of μ, the symbols $\bar{x}$ or $E(x)$ or $\langle x\rangle$ may be used to denote the mean value of x.

$$(5.5) \qquad \bar{x} = E(x) = \langle x \rangle = \mu = \sum_i x_i f(x_i).$$

The term expectation comes from games of chance.

Example 3. Suppose you will be paid \$5 if a die shows a 5, \$2 if it shows a 2 or a 3, and nothing otherwise. Let x represent your gain in playing the game. Then the possible values of x and the corresponding probabilities are $x = 5$ with $p = \frac{1}{6}$, $x = 2$ with $p = \frac{1}{3}$, and $x = 0$ with $p = \frac{1}{2}$. We find for the average or expectation of x:

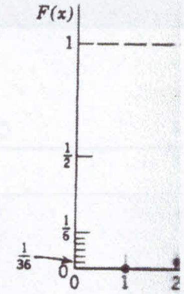
$$E(x) = \sum x_i p_i = \$5 \cdot \frac{1}{6} + \$2 \cdot \frac{1}{3} + \$0 \cdot \frac{1}{2} = \$1.50.$$

If you play the game many times, this is a reasonable estimate of your average gain per game; this is what your expectation means. It is also a reasonable amount to pay as a fee for each game you play. The term *expected value* (which means the same as *expectation* or *average*) may be somewhat confusing and misleading if you try to interpret "expected" in an everyday sense. Note that the expected value (\$1.50) of x is not one of the possible values of x, so you cannot ever "expect" to have $x = \$1.50$. If you think of expected value as a technical term meaning the same as average, then there is no difficulty. Of course, in some cases, it makes reasonable sense with its everyday meaning; for example, if a coin is tossed n times, the expected number of heads is $n/2$ (Problem 11) and it is true that we may reasonably "expect" a fair approximation to this result (see Section 7).

Cumulative Distribution Functions So far we have been using the probability function $f(x)$ which gives the probability $p_i = f(x_i)$ that x is exactly x_i. In some problems we may be more interested in the probability that x is less than some particular value. For example, in an election we would like to know the probability that less than half the votes would be cast for the opposing candidate, that is, that our candidate would win. In an experiment on radioactivity, we would like to know the probability that the background radiation always remains below a certain level. Given the probability function $f(x)$, we can obtain the probability that x is less than or equal to a certain value x_i by adding all the probabilities of values of x less than or equal to x_i. For example, consider the sum of the numbers on two dice; the probability function $p = f(x)$ is plotted in Figure 5.1. The probability that x is, say, less than or equal to 4 is the sum of the probabilities that x is 2 or 3 or 4, that is, $\frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{1}{6}$. Similarly, we could find the probability that x is less than or equal to any given number. The resulting function of x is plotted in Figure 5.2. Such a function $F(x)$ is called a *cumulative distribution function*; we can write

$$(5.6) \qquad F(x_i) = (\text{probability that } x \le x_i) = \sum_{x_j \le x_i} f(x_j).$$

Note carefully that, although the probability function $f(x)$ may be referred to as a *probability distribution*, the term *distribution function* means the *cumulative distribution* $F(x)$.



PROBLEMS, SECTION 5

Set up sample spaces fo
the indicated random w
Make a table of the diffe
Compute the mean, the
cumulative distribution

1. Three coins are to
2. Two dice are thro
3. A coin is tossed re
4. Suppose that Mart
a pair of these dice
the dice if the pro
5. A random variable
6. A card is drawn f
$x = -1$ if it is a 2
7. A weighted coin w
number of heads n
8. Would you pay \$1
equal to the produ
If it is more than \$
9. Show that the ex
same sample space
probabilities assoc
$\cdots, y_n$, be the valu
out $E(x)$, $E(y)$, an
10. Let μ be the avera
deviations of x fro
Hint: Remember t
11. Show that the exp
ways that the expe
(a) Let x = numb
(b) Let x = numb
average of x
number of he

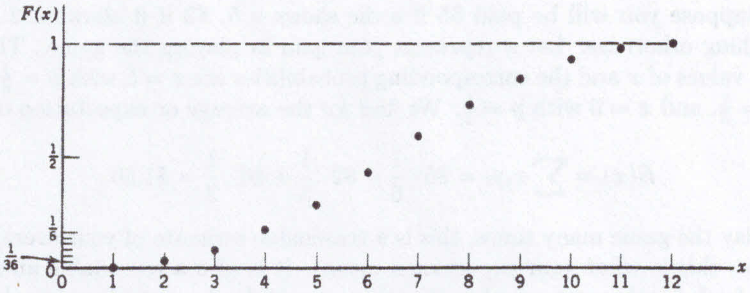


Figure 5.2

PROBLEMS, SECTION 5

Set up sample spaces for Problems 1 to 7 and list next to each sample point the value of the indicated random variable x, and the probability associated with the sample point. Make a table of the different values x_i of x and the corresponding probabilities $p_i = f(x_i)$. Compute the mean, the variance, and the standard deviation for x. Find and plot the cumulative distribution function $F(x)$.

1. Three coins are tossed; $x =$ number of heads minus number of tails.
2. Two dice are thrown; $x =$ sum of the numbers on the dice.
3. A coin is tossed repeatedly; $x =$ number of the toss at which a head first appears.
4. Suppose that Martian dice are 4-sided (tetrahedra) with points labeled 1 to 4. When a pair of these dice is tossed, let x be the product of the two numbers at the tops of the dice if the product is odd; otherwise $x = 0$.
5. A random variable x takes the values 0, 1, 2, 3, with probabilities $\frac{5}{12}, \frac{1}{3}, \frac{1}{12}, \frac{1}{6}$.
6. A card is drawn from a shuffled deck. Let $x = 10$ if it is an ace or a face card; $x = -1$ if it is a 2; and $x = 0$ otherwise.
7. A weighted coin with probability p of coming down heads is tossed three times; $x =$ number of heads minus number of tails.
8. Would you pay \$10 per throw of two dice if you were to receive a number of dollars equal to the product of the numbers on the dice? *Hint:* What is your expectation? If it is more than \$10, then the game would be favorable for you.
9. Show that the expectation of the sum of two random variables defined over the same sample space is the sum of the expectations. *Hint:* Let $p_1, p_2, \cdots, p_n$ be the probabilities associated with the n sample points; let $x_1, x_2, \cdots, x_n$, and $y_1, y_2, \cdots, y_n$, be the values of the random variables x and y for the n sample points. Write out $E(x)$, $E(y)$, and $E(x+y)$.
10. Let μ be the average of the random variable x. Then the quantities $(x_i - \mu)$ are the deviations of x from its average. Show that the average of these deviations is zero. *Hint:* Remember that the sum of all the p_i must equal 1.
11. Show that the expected number of heads in a single toss of a coin is $\frac{1}{2}$. Show in two ways that the expected number of heads in two tosses of a coin is 1:

    (a) Let $x =$ number of heads in two tosses and find $\bar{x}$.

    (b) Let $x =$ number of heads in toss 1 and $y =$ number of heads in toss 2; find the average of $x + y$ by Problem 9. Use this method to show that the expected number of heads in n tosses of a coin is $\frac{1}{2}n$.

12. Use Problem 9 to find the expected value of the sum of the numbers on the dice in Problem 2.

13. Show that adding a constant K to a random variable increases the average by K but does not change the variance. Show that multiplying a random variable by K multiplies both the average and the standard deviation by K.

14. As in Problem 11, show that the expected number of 5's in n tosses of a die is $n/6$.

15. Use Problem 9 to find $\bar{x}$ in Problem 7.

16. Show that $\sigma^2 = E(x^2) - \mu^2$. *Hint:* Write the definition of σ^2 from (5.3) and (5.4) and use Problems 9 and 13.

17. Use Problem 16 to find σ in Problems 2, 6, and 7.

6. CONTINUOUS DISTRIBUTIONS

In Section 5, we discussed random variables x which took a discrete set of values x_i. It is not hard to think of cases in which a random variable takes a continuous set of values.

Example 1. Consider a particle moving back and forth along the x axis from $x = 0$ to $x = l$, rebounding elastically at the turning points so that its speed is constant. (This could be a simple-minded model of an alpha particle in a radioactive nucleus, or of a gas molecule bouncing back and forth between the walls of a container.) Let the position x of the particle be the random variable; then x takes a continuous set of values from $x = 0$ to $x = l$. Now suppose that, following Section 5, we ask for the probability that the particle is *at* a particular point x; this probability must be the same, say k, for all points (because the speed is constant). In Section 5, with a finite number of points, we would say $k = 1/N$. In the continuous case, there are an infinite number of points so we would find $k = 0$, that is, the probability that the particle is *at* a given point) must be zero. But this is not a very useful result. Let us instead divide $(0, l)$ into small intervals dx; since the particle has constant speed, the time it spends in each dx is proportional to the length of dx. In fact, since the particle spends the fraction $(dx)/l$ of its time in a given interval dx, the probability of finding it in dx is just $(dx)/l$.

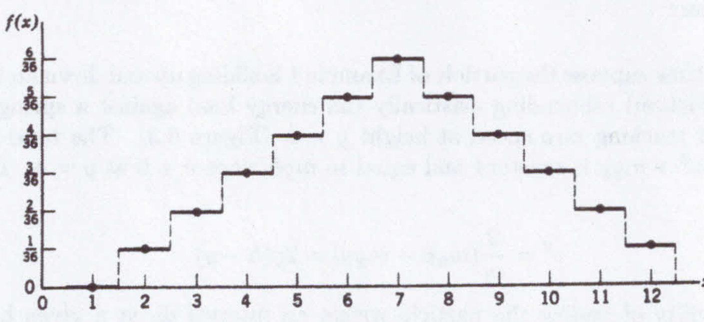


Figure 6.1

Section 6

Comparison of Di
how to define a prob
discussion with the d
we plotted a vertical
of x. Instead of a dot
horizontal line segme
area under the horizo
(since the length of e
instead of the ordina
histogram.

Example 2. Now let us app
plotted the function

If we consider any
on $(0, l)$, the area und
$1/l$ for this interval is
and this is just the p
particle is in this int
bility that the particl
subinterval of $(0, l)$, sa
or $\int_a^b f(x)\,dx$, that is,
curve from a to b. If t
outside $(0, l)$, then $\int_a^b$
value of the probabilit

When $f(x)$ is cons
uniformly distributed
is not constant.

Example 3. This time suppo
plane (no friction) reb
bottom and reaching
namely $\frac{1}{2}mv^2 + mgy$ is
have

(6.1)

The probability of find
is proportional to the
$dt = (ds)/v$; from Figu
we have

Since the probability
is proportional to dt, w

Comparison of Discrete and Continuous Probability Functions To see how to define a probability function for the continuous case and to correlate this discussion with the discrete case, let us return for a moment to Figure 5.1. There we plotted a vertical *distance* to represent the probability $p = f(x)$ of each value of x. Instead of a dot (as in Figure 5.1) to indicate p for each x, let us now draw a horizontal line segment of length 1 centered on each dot, as in Figure 6.1. Then the *area* under the horizontal line segment at a particular x_i is $f(x_i) \cdot 1 = f(x_i) = p_i$ (since the length of each horizontal line segment is 1), and we could use this *area* instead of the ordinate as a measure of the probability. Such a graph is called a *histogram.*

Example 2. Now let us apply this area idea to Example 1. Consider Figure 6.2. We have plotted the function

$$f(x) = \begin{cases} 1/l, & 0 < x < l, \\ 0, & x < 0 \text{ and } x > l. \end{cases}$$

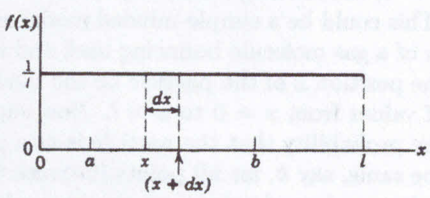


Figure 6.2

If we consider any interval x to $x + dx$ on $(0, l)$, the area under the curve $f(x) = 1/l$ for this interval is $(1/l)\,dx$ or $f(x)\,dx$, and this is just the probability that the particle is in this interval. The probability that the particle is in some longer subinterval of $(0, l)$, say (a, b), is $(b - a)/l$ or $\int_a^b f(x)\,dx$, that is, the area under the curve from a to b. If the interval (a, b) is outside $(0, l)$, then $\int_a^b f(x)\,dx = 0$ since $f(x)$ is zero, and again this is the correct value of the probability of finding the particle on the given interval.

When $f(x)$ is constant over an interval (as in Figure 6.2), we say that x is *uniformly* distributed on that interval. Let us consider an example in which $f(x)$ is not constant.

Example 3. This time suppose the particle of Example 1 is sliding up and down an inclined plane (no friction) rebounding elastically (no energy loss) against a spring at the bottom and reaching zero speed at height $y = h$ (Figure 6.3). The total energy, namely $\frac{1}{2}mv^2 + mgy$ is constant and equal to mgh since $v = 0$ at $y = h$. Thus we have

$$v^2 = \frac{2}{m}(mgh - mgy) = 2g(h - y). \tag{6.1}$$

The probability of finding the particle within an interval dy at a given height y is proportional to the time dt spent in that interval. From $v = ds/dt$, we have $dt = (ds)/v$; from Figure 6.3, we find $ds = (dy)\csc\alpha$. Combining these with (6.1) we have

$$dt = \frac{ds}{v} = \frac{(dy)\csc\alpha}{\sqrt{2g}\sqrt{h-y}}$$

Since the probability $f(y)\,dy$ of finding the particle in the interval dy at height y is proportional to dt, we can drop the constant factor $(\csc\alpha)/\sqrt{2g}$, and say that

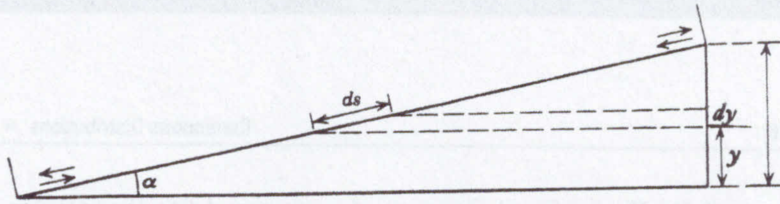


Figure 6.3

$f(y)\,dy$ is proportional to $dy/\sqrt{h-y}$. In order to find $f(y)$, we must multiply by a constant factor which makes the total probability $\int_0^h f(y)\,dy$ equal to 1 since this is the probability that the particle is *somewhere*. You can easily verify that

$$f(y)\,dy = \frac{1}{2\sqrt{h}}\frac{dy}{\sqrt{h-y}} \quad \text{or} \quad f(y) = \frac{1}{2\sqrt{h(h-y)}}$$

A graph of $f(y)$ is plotted in Figure 6.4. Note that although $f(y)$ becomes infinite at $y = h$, the area under the $f(y)$ curve for any interval is finite; this area represents the probability that the particle is in that height interval.

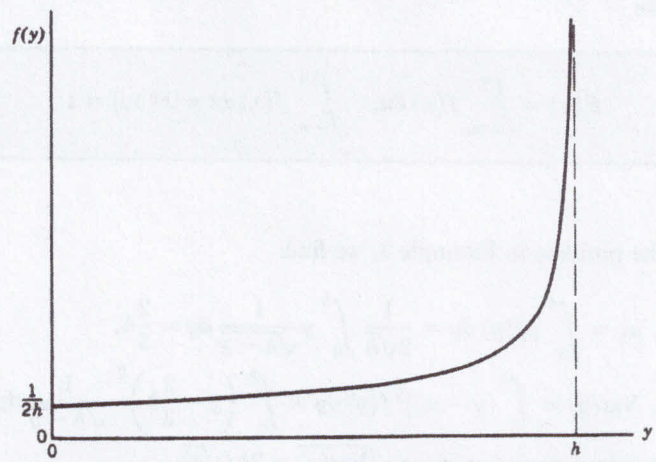


Figure 6.4

We can now extend the definitions of mean (expectation), variance, standard deviation, and cumulative distribution function to the continuous case. Let $f(x)$ be a probability density function; remember that $\int_{-\infty}^{\infty} f(x)\,dx = 1$ just as $\sum_{i=1}^{n} p_i = 1$. The average of a random variable x with probability density function $f(x)$ is

$$\text{(6.2)} \qquad \mu = \overline{x} = E(x) = \langle x \rangle = \int_{-\infty}^{\infty} x f(x)\,dx.$$

(In writing the limits $-\infty, \infty$ here, we assume that $f(x)$ is defined to be zero on intervals where the probability is zero.) Note that (6.2) is a natural extension of the sum in (5.5). Having found the mean of x, we now define the variance as in Section 5 as the average of $(x-\mu)^2$, that is,

$$\text{(6.3)} \qquad \operatorname{Var}(x) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \sigma_x^2.$$

As before, the standard deviation σ_x is the square root of the variance. Finally, the cumulative distribution function $F(x)$ gives for each x the probability that the random variable is less than or equal to that x. But this probability is just the area under the $f(x)$ curve from $-\infty$ up to the point x. Also, of course, the integral of $f(x)$ from $-\infty$ to ∞ must $= 1$ since that is the total probability for all values of x. Thus we have

$$\text{(6.4)} \qquad F(x) = \int_{-\infty}^{x} f(u)\,du, \qquad \int_{-\infty}^{\infty} f(x)\,dx = F(\infty) = 1.$$

Example 4. For the problem in Example 3, we find:

By (6.2), $\mu_y = \int_0^h y f(y)\,dy = \dfrac{1}{2\sqrt{h}}\int_0^h y \dfrac{1}{\sqrt{h-y}}\,dy = \dfrac{2}{3}h.$

By (6.3), $\operatorname{Var}(y) = \int_0^h (y-\mu_y)^2 f(y)\,dy = \int_0^h \left(y - \dfrac{2}{3}h\right)^2 \dfrac{1}{\sqrt{h-y}}\,dy = \dfrac{4h^2}{45},$

so standard deviation $\sigma_y = \sqrt{\operatorname{Var}(y)} = 2h/\sqrt{45}.$

By (6.4), cumulative distribution function $F(y) = \int_0^y f(u)\,du$

$$= \frac{1}{2\sqrt{h}}\int_0^y \frac{du}{\sqrt{h-u}}.$$

Why "density function"? In Section 5, we mentioned that the probability function $f(x)$ is often called the *probability density.* We can now explain why. Consider (6.2). If $f(x)$ represents the density (mass per unit length) of a thin rod, then the center of mass of the rod is given by [see Chapter 5, (3.3)]

$$\text{(6.5)} \qquad \bar{x} = \int x f(x)\,dx \Big/ \int f(x)\,dx,$$

where the integrals are over the length of the rod, or from $-\infty$ to ∞ as in (6.2) with $f(x) = 0$ outside the rod. But in (6.2), $\int f(x)\,dx$ is the total probability that x has *some* value, and so this integral is equal to 1. Then (6.5) and (6.2) are really the same; we see that it is reasonable to call $f(x)$ a density, and also that the mean of x corresponds to the center of mass of a linear mass distribution of density $f(x)$. In a similar way, we can interpret (6.3) as giving the moment of inertia of the mass distribution about the center of mass (see Chapter 5, Section 3).

Joint Distributions We can easily generalize the ideas and formulas above to two (or more) dimensions. Suppose we have two random variables x and y; we define their joint probability density function $f(x,y)$ so that $f(x_i, y_j)\,dx\,dy$ is the probability that the point (x,y) is in an element of area $dx\,dy$ at $x = x_i$, $y = y_j$. Then the probability that the point (x,y) is in a given region of the (x,y) plane, is the integral of $f(x,y)$ over that area. The average or expected values of x and y, the variances and standard deviations of x and y, and the covariance of x, y (see Problems 13 to 16) are given by

$$\bar{x} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,dx\,dy,$$

$$\bar{y} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y)\,dx\,dy,$$

$$\text{(6.6)}\qquad \operatorname{Var}(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\bar{x})^2 f(x,y)\,dx\,dy = \sigma_x^2,$$

$$\operatorname{Var}(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y-\bar{y})^2 f(x,y)\,dx\,dy = \sigma_y^2,$$

$$\operatorname{Cov}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\bar{x})(y-\bar{y}) f(x,y)\,dx\,dy.$$

You should see that these are generalizations of (6.2) and (6.3); that (6.6) can be interpreted as giving the coordinates of the center of mass and the moments of inertia of a two-dimensional mass distribution; and that similar formulas can be written for three (or more) random variables (that is, in three or more dimensions). Also note that the formulas in (6.6) could be written in terms of polar coordinates (see Problems 6 to 9).

We have discussed a number of probability distributions both discrete and continuous, and you will find others in the problems. We will discuss three very important named distributions (binomial, normal, and Poisson) in the following sections. Learning about these and related graphs, formulas, and terminology should make it possible for you to cope with any of the many other named distributions you find in texts, reference books, and computer programs.

PROBLEMS, SECTION 6

1. (a) Find the probability density function $f(x)$ for the position x of a particle which is executing simple harmonic motion on $(-a, a)$ along the x axis. (See Chapter 7, Section 2, for a discussion of simple harmonic motion.) *Hint:* The value of x at time t is $x = a\cos\omega t$. Find the velocity dx/dt; then the probability of finding the particle in a given dx is proportional to the time it spends there which is inversely proportional to its speed there. Don't forget that the total probability of finding the particle *somewhere* must be 1.

(b) Sketch the probability density function $f(x)$ found in part (a) and also the cumulative distribution function $F(x)$ [see equation (6.4)].

(c) Find the average and the standard deviation of x in part (a).

2. It is shown in the kinetic theory of gases that the probability for the distance a molecule travels between collisions to be between x and $x + dx$, is proportional to $e^{-x/\lambda}\,dx$, where λ is a constant. Show that the average distance between collisions (called the "mean free path") is λ. Find the probability of a free path of length $\geq 2\lambda$.

3. A ball is thrown
density function
h and $h + dh$. H

4. In Problem 1 we
oscillator. In qua
oscillator (in the
x takes values fro
of x. (In quantum
in position and is

5. The probability f
proportional to e
function $F(t)$. Fi
particle. Compare
value of t when e

6. A circular garden
distributed over t
some particular a
seed to be in the
particular seed) i
probability for a

7. (a) Repeat Prob
earth, say al
on the earth
seeds could
be uniformly

(b) Also find F
Do your ans

8. Given that a parti
of being found in
distribution funct
function $f(r)$. H
radius r. Find $\bar{r}$ a

9. A hydrogen atom
the electron revol
the ground state)
distance r (from
the electron is in
is proportional to
coordinates (see
$f(r)\,dr$ is the pro
from the proton.
must be 1.) Com
then say that the
of r^{-1} is a^{-1}.

10. Do Problem 5.10

11. Do Problem 5.13

12. Do Problem 5.16

13. Given a joint dist
$E(y)$ and $\operatorname{Var}(x +$

s and formulas above to
n variables x and y; we
hat $f(x_i, y_j)\,dx\,dy$ is the
$dx\,dy$ at $x = x_i$, $y = y_j$.
ion of the (x, y) plane, is
ected values of x and y,
e covariance of x, y (see

$y = \sigma_x^2$,

$= \sigma_y^2$,

$) \,dx\,dy$.

(6.3): that (6.6) can be
ss and the moments of
similar formulas can be
ee or more dimensions).
ms of polar coordinates

both discrete and con-
iscuss three very impor-
n the following sections.
rminology should make
d distributions you find

position x of a particle
a) along the x axis. (See
onic motion.) *Hint:* The
x/dt; then the probability
the time it spends there
on't forget that the total
e 1.

in part (a) and also the
6.4)].

part (a).

bility for the distance a
+ dx, is proportional to
stance between collisions
free path of length $\geq 2\lambda$.

3. A ball is thrown straight up and falls straight back down. Find the probability density function $f(h)$ so that $f(h)\,dh$ is the probability of finding it between height h and $h + dh$. *Hint:* Look at Example 3.

4. In Problem 1 we found the probability density function for a classical harmonic oscillator. In quantum mechanics, the probability density function for a harmonic oscillator (in the ground state) is proportional to $e^{-\alpha^2 x^2}$, where α is a constant and x takes values from $-\infty$ to ∞. Find $f(x)$ and the average and standard deviation of x. (In quantum mechanics, the standard deviation of x is called the uncertainty in position and is written Δx.)

5. The probability for a radioactive particle to decay between time t and time $t + dt$ is proportional to $e^{-\lambda t}$. Find the density function $f(t)$ and the cumulative distribution function $F(t)$. Find the expected lifetime (called the mean life) of the radioactive particle. Compare the mean life and the so-called "half life" which is defined as the value of t when $e^{-\lambda t} = 1/2$.

6. A circular garden bed of radius 1 m is to be planted so that N seeds are uniformly distributed over the circular area. Then we can talk about the number n of seeds in some particular area A, or we can call n/N the probability for any one particular seed to be in the area A. Find the probability $F(r)$ that a seed (that is, some particular seed) is within r of the center. (*Hint:* What is F(1)?) Find $f(r)\,dr$, the probability for a seed to be between r and $r + dr$ from the center. Find $\bar{r}$ and σ.

7. (a) Repeat Problem 6 where the "circular" area is now on the curved surface of the earth, say all points at distance s from Chicago (measured along a great circle on the earth's surface) with $s \leq \pi R/3$ where R = radius of the earth. The seeds could be replaced by, say, radioactive fallout particles (assuming these to be uniformly distributed over the surface of the earth). Find $F(s)$ and $f(s)$.

(b) Also find $F(s)$ and $f(s)$ if $s \leq 1 \ll R$ (say $s \leq 1$ mile where $R = 4000$ miles). Do your answers then reduce to those in Problem 6?

8. Given that a particle is inside a sphere of radius 1, and that it has equal probabilities of being found in any two volume elements of the same size, find the cumulative distribution function $F(r)$ for the spherical coordinate r, and from it find the density function $f(r)$. *Hint:* $F(r)$ is the probability that the particle is inside a sphere of radius r. Find $\bar{r}$ and σ.

9. A hydrogen atom consists of a proton and an electron. According to the Bohr theory, the electron revolves about the proton in a circle of radius a ($a = 5 \cdot 10^{-9}$cm for the ground state). According to quantum mechanics, the electron may be at any distance r (from 0 to ∞) from the proton; for the ground state, the probability that the electron is in a volume element dV, at a distance r to $r + dr$ from the proton, is proportional to $e^{-2r/a}dV$, where a is the Bohr radius. Write dV in spherical coordinates (see Chapter 5, Section 4) and find the density function $f(r)$ so that $f(r)\,dr$ is the probability that the electron is at a distance between r and $r + dr$ from the proton. (Remember that the probability for the electron to be somewhere must be 1.) Computer plot $f(r)$ and show that its maximum value is at $r = a$; we then say that the most probable value of r is a. Also show that the average value of r^{-1} is a^{-1}.

10. Do Problem 5.10 for a continuous distribution.

11. Do Problem 5.13 for a continuous distribution.

12. Do Problem 5.16 for a continuous distribution.

13. Given a joint distribution function $f(x, y)$ as in (6.6), show that $E(x + y) = E(x) + E(y)$ and $\mathrm{Var}(x + y) = \mathrm{Var}(x) + \mathrm{Var}(y) + 2\,\mathrm{Cov}(x, y)$.

14. Recall that two events A and B are called independent if $p(AB) = p(A)p(B)$. Similarly two random variables x and y are called independent if the joint probability function $f(x,y) = g(x)h(y)$. Show that if x and y are independent, then the expectation or average of xy is $E(xy) = E(x)E(y) = \mu_x \mu_y$.

15. Show that the covariance of two independent (see Problem 14) random variables is zero, and so by Problem 13, the variance of the sum of two independent random variables is equal to the sum of their variances.

16. By Problem 15, if x and y are independent, then $\text{Cov}(x,y) = 0$. The converse is not always true, that is, if $\text{Cov}(x,y) = 0$, it is not necessarily true that the joint distribution function is of the form $f(x,y) = g(x)h(y)$. For example, suppose $f(x,y) = (3y^2 + \cos x)/4$ on the rectangle $-\pi/2 < x < \pi/2, -1 < y < 1$, and $f(x,y) = 0$ elsewhere. Show that $\text{Cov}(x,y) = 0$, but x and y are not independent, that is, $f(x,y)$ is not of the form $g(x)h(y)$. Can you construct some more examples?

7. BINOMIAL DISTRIBUTION

Example 1. Let a coin be tossed 5 times; what is the probability of exactly 3 heads out of the 5 tosses? We can represent any sequence of 5 tosses by a symbol such as *thhth*. The probability of this particular sequence (or any other particular sequence) is $(\frac{1}{2})^5$ since the tosses are independent (see Example 1 of Section 3). The number of such sequences containing 3 heads and 2 tails is the number of ways we can select 3 positions out of 5 for heads (or 2 for tails), namely $C(5,3)$. Hence, the probability of exactly 3 heads in 5 tosses of a coin is $C(5,3)(\frac{1}{2})^5$. Suppose a coin is tossed repeatedly, say n times; let x be the number of heads in the n tosses. We want to find the probability density function $p = f(x)$ which gives the probability of exactly x heads in n tosses. By generalizing the case of 3 heads in 5 tosses, we see that

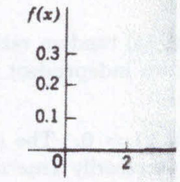
$$f(x) = C(n,x)(\tfrac{1}{2})^n \tag{7.1}$$

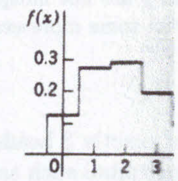
Example 2. Let us do a similar problem with a die, asking this time for the probability of exactly 3 aces in 5 tosses of the die. If A means ace and N not ace, the probability of a particular sequence such as $ANNAA$ is $\frac{1}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}$ since the probability of A is $\frac{1}{6}$, the probability of N is $\frac{5}{6}$, and the tosses are independent. The number of such sequences containing 3 A's and 2 N's is $C(5,3)$; thus the probability of exactly 3 aces in 5 tosses of a die is $C(5,3)(\frac{1}{6})^3(\frac{5}{6})^2$. Generalizing this, we find that the probability of exactly x aces in n tosses of a die is

$$f(x) = C(n,x)(\tfrac{1}{6})^x(\tfrac{5}{6})^{n-x}. \tag{7.2}$$

Bernoulli Trials In the two examples we have just done, we have been concerned with repeated independent trials, each trial having two possible outcomes (h or t, A or N) of given probability. There are many examples of such problems; let's consider a few. A manufactured item is good or defective; given the probability of a defect we want the probability of x defectives out of n items. An archer has probability p of hitting a target; we ask for the probability of x hits out of n tries. Each atom of a radioactive substance has probability p of emitting an alpha particle during the next minute; we are to find the probability that x alpha particles will be emitted in the next minute from the n atoms in the sample. A particle moves back and forth along the x axis in unit jumps; it has, at each step, equal probabilities of

Graphs





jumping forward or ba
as a model of a diffusi
jumps, the particle is
d = number x o
from its starting point
a total of n jumps.
In all these problem
possible outcomes of p

Binom

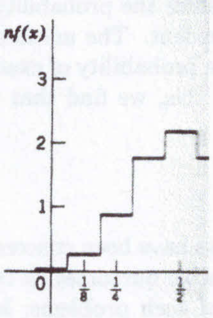


Figure 7

$p(AB) = p(A)p(B)$. Sim-
ent if the joint probability
ependent, then the expec-

m 14) random variables is
two independent random

$x, y) = 0$. The converse
necessarily true that the
$y)$. For example, suppose
$< \pi/2, -1 < y < 1$, and
d y are not independent,
uct some more examples?

of exactly 3 heads out of
a symbol such as *thhth*.
particular sequence) is
tion 3). The number of
of ways we can select 3
Hence, the probability
ppose a coin is tossed
e n tosses. We want to
e probability of exactly
5 tosses, we see that

e for the probability of
not ace, the probability
since the probability of
ndent. The number of
e probability of exactly
this, we find that the

e have been concerned
sible outcomes (h or t,
of such problems; let's
given the probability
items. An archer has
of x hits out of n tries.
itting an alpha particle
alpha particles will be
A particle moves back
equal probabilities of

Graphs of the binomial distribution, $f(x) = C(n, x)p^x q^{n-x}$

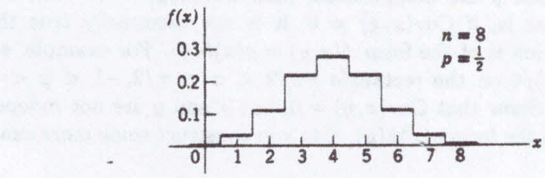


Figure 7.1

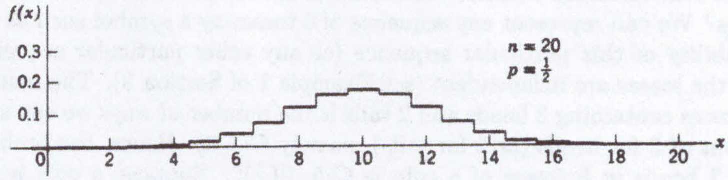


Figure 7.2

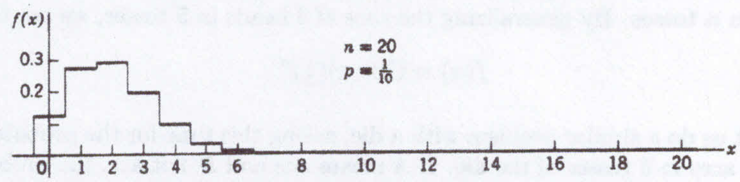


Figure 7.3

jumping forward or backward. (This motion is called a *random walk*; it can be used as a model of a diffusion process.) We want to know the probability that, after n jumps, the particle is at a distance

$$d = \text{number } x \text{ of positive jumps} - \text{number } (n - x) \text{ of negative jumps},$$

from its starting point; this probability is the probability of x positive jumps out of a total of n jumps.

In all these problems, something is tried repeatedly. At each trial there are two possible outcomes of probabilities p (usually called the probability of "success") and

Binomial distribution graphs of $nf(x)$ plotted against x/n

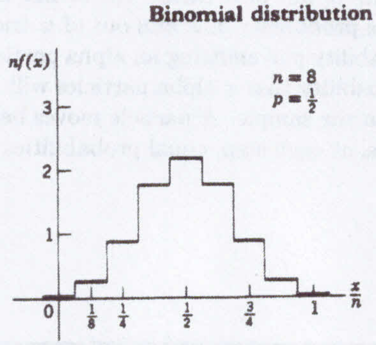


Figure 7.4

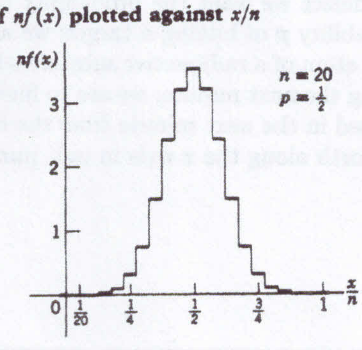


Figure 7.5

$q = 1 - p$ (where q = probability of "failure"). Such repeated independent trials with constant probabilities p and q are called *Bernoulli trials.*

Binomial Probability Functions Let us generalize (7.1) and (7.2) to obtain a formula which applies to any similar problem, namely the probability $f(x)$ of exactly x successes in n Bernoulli trials. Reasoning as we did to obtain (7.1) and (7.2), we find that

$$f(x) = C(n, x)p^x q^{n-x}. \tag{7.3}$$

We might also ask for the probability of *not more than* x successes in n trials. This is the sum of the probabilities of $0, 1, 2, \cdots, x$ successes, that is, it is the cumulative distribution function $F(x)$ for the random variable x whose probability density function is (7.3) [see (5.6)]. We can write

$$\begin{aligned} F(x) &= f(0) + f(1) + \cdots + f(x) \\ &= C(n, 0)p^0 q^n + C(n, 1)p^1 q^{n-1} + \cdots + C(n, x)p^x q^{n-x} \\ &= \sum_{u=0}^{x} C(n, u)p^u q^{n-u} = \sum_{u=0}^{x} \binom{n}{u} p^u q^{n-u}. \end{aligned} \tag{7.4}$$

Observe that (7.3) is one term of the binomial expansion of $(p + q)^n$ and (7.4) is a sum of several terms of this expansion (see Section 4, Example 2). For this reason, the functions $f(x)$ in (7.1), (7.2), or (7.3) are called *binomial probability* (or *density) functions* or *binomial distributions*, and the function $F(x)$ in (7.4) is called a *binomial cumulative distribution function.*

We shall find it very useful to computer plot graphs of the binomial density function $f(x)$ for various values of p and n. (See Figures 7.1 to 7.5 and Problems 1 to 8.) Instead of a point at $y = f(x)$ for each x, we plot a horizontal line segment of length 1 centered on each x as in Figure 6.1; the probabilities are then represented by *areas* under the broken line, rather than by ordinates. From Figures 7.1 to 7.3 and similar graphs, we can draw a number of conclusions. The most probable value of x [corresponding to the largest value of $f(x)$] is approximately $x = np$ (Problems 10 and 11); for example for $p = \frac{1}{2}$, the most probable value of x is $\frac{1}{2}n$ for even n; for odd n, there are two consecutive values of x, namely $\frac{1}{2}(n \pm 1)$, for which the probability is largest. The graphs for $p = \frac{1}{2}$ are symmetric about $x = \frac{1}{2}n$. For $p \neq \frac{1}{2}$, the curve is asymmetric, favoring small x values for small p and large x values for large p. As n increases, the graph of $f(x)$ becomes wider and flatter (the total area under the graph must remain 1). The probability of the most probable value of x decreases with n. For example, the most probable number of heads in 8 tosses of a coin is 4 with probability 0.27; the most probable number of heads in 20 tosses is 10 with probability 0.17; for 10^6 tosses, the probability of exactly 500,000 heads is less than 10^{-3}.

Let us redraw Figures 7.1 and 7.2 plotting $nf(x)$ against the relative number of successes x/n (Figures 7.4 and 7.5). Since this change of scale (ordinate times n, abscissa divided by n) leaves the area unchanged, we can still use the area to represent probability. Note that now the curves become narrower and taller as n

increases. This me
probable value, na
difference "number
with n (Figures 7.
is apt to be closer
reason that we ca
estimate of p.

Chebyshev's Ine
find useful. We co
let μ be the mean
that if we select an
by more than t, is
by more than a fe
deviation σ, we fin
less than $\sigma^2/t^2 = \sigma$

where the sum is o
$|x - \mu| \geq t$, we get

(7.5)

If we replace each x

(7.6) $\sigma^2 > \sum$

But $\sum_{|x-\mu|\geq t} f(x)$ is
by more than t, and (

Laws of Large Nu
general comments ab
Let us state and pro
dom variable whose
Problems 9 and 13 we

(7.7) (proba

Let us choose the ar
where ϵ is now arbitr

(7.8) (probabil

or, when we divide th

(7.9) (probe

increases. This means that values of the ratio x/n tend to cluster about their most probable value, namely $np/n = p$. For example, if we toss a coin repeatedly, the difference "number of heads $-\frac{1}{2}$ number of tosses" is apt to be large and to increase with n (Figures 7.1 and 7.2), but the ratio "number of heads $\div$ number of tosses" is apt to be closer and closer to $\frac{1}{2}$ as n increases (Figures 7.4 and 7.5). It is for this reason that we can use experimentally determined values of x/n as a reasonable estimate of p.

Chebyshev's Inequality This is a simple but very general result which we will find useful. We consider a random variable x with probability function $f(x)$, and let μ be the mean value and σ the standard deviation of x. We are going to prove that if we select any number t, the probability that x differs from its mean value μ by more than t, is less than σ^2/t^2. This means that x is unlikely to differ from μ by more than a few standard deviations; for example, if t is twice the standard deviation σ, we find that the probability for x to differ from μ by more than 2σ is less than $\sigma^2/t^2 = \sigma^2/(2\sigma)^2 = \frac{1}{4}$. The proof is simple. By definition of σ, we have

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

where the sum is over all x. Then if we sum just over the values of x for which $|x-\mu| \geq t$, we get less than σ^2:

$$\sigma^2 > \sum_{|x-\mu|\geq t} (x-\mu)^2 f(x). \tag{7.5}$$

If we replace each $x-\mu$ by the number t in (7.5), the sum is decreased, so we have

$$\sigma^2 > \sum_{|x-\mu|\geq t} t^2 f(x) = t^2 \sum_{|x-\mu|\geq t} f(x) \qquad \text{or} \qquad \sum_{|x-\mu|\geq t} f(x) < \frac{\sigma^2}{t^2}. \tag{7.6}$$

But $\sum_{|x-\mu|\geq t} f(x)$ is just the sum of all probabilities of x values which differ from μ by more than t, and (7.6) says that this probability is less than σ^2/t^2, as we claimed.

Laws of Large Numbers Statements and proofs which make more precise our general comments about the effect of large n are known as *laws of large numbers.* Let us state and prove one such law. We apply Chebyshev's inequality to a random variable whose probability function is the binomial distribution (7.3). From Problems 9 and 13 we have $\mu = np$ and $\sigma = \sqrt{npq}$. Then by Chebyshev's inequality,

$$(\text{probability of } |x-np| \geq t) \qquad \text{is less than} \quad npq/t^2 \tag{7.7}$$

Let us choose the arbitrary value of t in (7.7) proportional to n, that is, $t = n\epsilon$ where ϵ is now arbitrary. Then (7.7) becomes

$$(\text{probability of } |x-np| \geq n\epsilon) \qquad \text{is less than} \quad npq/n^2\epsilon^2, \tag{7.8}$$

or, when we divide the first inequality by n,

$$\left(\text{probability of } \left|\frac{x}{n} - p\right| \geq \epsilon\right) \qquad \text{is less than} \quad \frac{pq}{n\epsilon^2}. \tag{7.9}$$

Recall that x/n is the relative number of successes; we intuitively expect x/n to be near p for large n. Now (7.9) says that, if ϵ is any small number, the probability is less than $pq/(n\epsilon^2)$ for x/n to differ from p by ϵ; that is, as n tends to infinity, this probability tends to zero. (Note, however, that x/n need not tend to p.) This is one form of the law of large numbers and it justifies our intuitive ideas.

▸ PROBLEMS, SECTION 7

For the values of n indicated in Problems 1 to 4:

(a) Write the probability density function $f(x)$ for the probability of x heads in n tosses of a coin and computer plot a graph of $f(x)$ as in Figures 7.1 and 7.2. Also computer plot a graph of the corresponding cumulative distribution function $F(x)$.

(b) Computer plot a graph of $nf(x)$ as a function of x/n as in Figures 7.4 and 7.5.

(c) Use your graphs and other calculations if necessary to answer these questions: What is the probability of exactly 7 heads? Of at most 7 heads? [Hint: Consider $F(x)$.] Of at least 7 heads? What is the most probable number of heads? The expected number of heads?

1. $n = 7$ **2.** $n = 12$ **3.** $n = 15$ **4.** $n = 18$

5. Write the formula for the binomial density function $f(x)$ for the case $n = 6, p = 1/6$, representing the probability of, say, x aces in 6 throws of a die. Computer plot $f(x)$ as in Figure (7.3). Also plot the cumulative distribution function $F(x)$. What is the probability of at least 2 aces out of 6 tosses of a die? *Hint:* Can you read the probability of at most one ace from one of your graphs?

For the given values of n and p in Problems 6 to 8, computer plot graphs of the binomial density function for the probability of x successes in n Bernoulli trials with probability p of success.

6. $n = 6$, $p = 5/6$ (Compare Problem 5)

7. $n = 50$, $p = 1/5$ **8.** $n = 50$, $p = 4/5$

9. Use the second method of Problem 5.11 to show that the expected number of successes in n Bernoulli trials with probability p of success is $\bar{x} = np$. *Hint:* What is the expected number of successes in one trial?

10. Show that the most probable number of heads in n tosses of a coin is $\frac{1}{2}n$ for even n [that is, $f(x)$ in (7.1) has its largest value for $x = n/2$] and that for odd n, there are two equal "largest" values of $f(x)$, namely for $x = \frac{1}{2}(n+1)$ and $x = \frac{1}{2}(n-1)$. *Hint:* Simplify the fraction $f(x+1)/f(x)$, and then find the values of x for which it is greater than 1 [that is, $f(x+1) > f(x)$], and less than or equal to 1 [that is, $f(x+1) \le f(x)$]. Remember that x must be an integer.

11. Use the method of Problem 10 to show that for the binomial distribution (7.3), the most probable value of x is approximately np (actually within 1 of this value).

12. Let x = number of heads in one toss of a coin. What are the possible values of x and their probabilities? What is μ_x? Hence show that $\text{Var}(x) =$ [average of $(x-\mu_x)^2$] $= \frac{1}{4}$, so the standard deviation is $\frac{1}{2}$. Now use the result from Problem 6.15 "variance of a sum of independent random variables = sum of their variances" to show that if x = number of heads in n tosses of a coin, $\text{Var}(x) = \frac{1}{4}n$ and the standard deviation $\sigma_x = \frac{1}{2}\sqrt{n}$.

13. Generalize Problem 12 to show that for the general binomial distribution (7.3), $\text{Var}(x) = npq$, and $\sigma = \sqrt{npq}$.

itively expect x/n to be
mber, the probability is
n tends to infinity, this
not tend to p.) This is
tuitive ideas.

ity of x heads in n tosses
1 and 7.2. Also computer
nction $F(x)$.
Figures 7.4 and 7.5.
er these questions: What
Hint: Consider $F(x)$.] Of
s? The expected number

4. $n = 18$

r the case $n = 6, p = 1/6$,
die. Computer plot $f(x)$
function $F(x)$. What is
Hint: Can you read the

t graphs of the binomial
trials with probability p

5

xpected number of suc-
$\bar{x} = np$. *Hint:* What is

a coin is $\frac{1}{2}n$ for even n
d that for odd n, there
$+ 1)$ and $x = \frac{1}{2}(n-1)$.
e values of x for which
or equal to 1 [that is,

l distribution (7.3), the
in 1 of this value).

possible values of x and
[average of $(x-\mu_x)^2$]
Problem 6.15 "variance
iances" to show that if
the standard deviation

ial distribution (7.3),

8. THE NORMAL OR GAUSSIAN DISTRIBUTION

The graph of the *normal* or *Gaussian distribution* is the bell-shaped curve you may know as the normal error curve (Figure 8.1). The normal distribution is used a great deal because, as we shall see, it is not only of interest in itself (see Problems 2 and 3), but also other distributions become almost normal when n (the number of trials or measurements) becomes large (see Figures 8.2 and 8.3).

The probability density function $f(x)$ and the cumulative distribution function $F(x)$ for the normal or Gaussian distribution are given by

$$(8.1)\qquad \begin{aligned} f(x) &= \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/(2\sigma^2)},\\ F(x) &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-(t-\mu)^2/(2\sigma^2)}\,dt. \end{aligned}\qquad \text{Normal distribution}$$

It is straightforward to show (Problem 1) that if x is a random variable with probability density $f(x)$ in (8.1), then the mean of x is μ and the standard deviation is σ. Also we can show that the integral of $f(x)$ from $-\infty$ to ∞ is equal to 1 as it must be for a probability function. Then the probability that a normally distributed random variable x lies between x_1 and x_2 is the area under the $f(x)$ curve between x_1 and x_2 which is

$$(8.2)\qquad F(x_2) - F(x_1) = \text{probability that } x_1 \le x \le x_2.$$

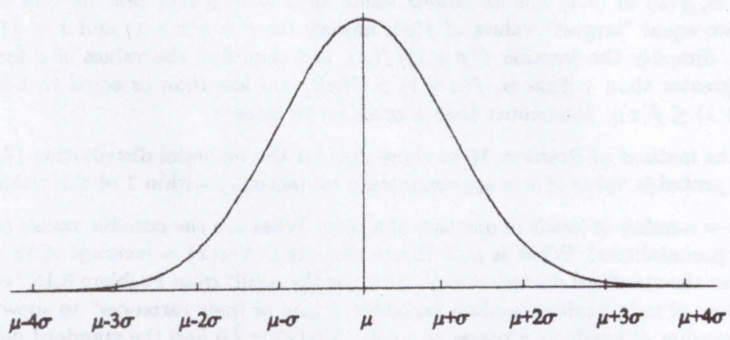


Figure 8.1

A normal density function graph (Figure 8.1) has its peak at $x = \mu$ and is symmetric with respect to the line $x = \mu$. Since the area from $-\infty$ to ∞ is 1, the area from $-\infty$ to μ is $\frac{1}{2}$ (that is, $F(\mu) = \frac{1}{2}$), and similarly the area from μ to ∞ is $\frac{1}{2}$. A change in μ merely translates the graph with no change in shape. An increase in σ widens and flattens the graph so that the area remains 1, and similarly a decrease in σ makes the graph taller and narrower. (Problems 4 to 6). The area from $\mu - \sigma$ to $\mu + \sigma$ is 0.6827, that is, the probability that x differs from its mean value by 1 standard deviation or less, is just over 68%. The probability that $|x - \mu| \le 2\sigma$

is over 95% and the probability that $|x - \mu| \le 3\sigma$ is over 99.7%. Note that these probabilities are independent of the values of μ and σ (Problem 7).

Normal Approximation to the Binomial Distribution As an example of approximating another distribution by a normal distribution, let's consider the binomial distribution (7.3). For large n and large np, we can use Stirling's formula (Chapter 11, Section 11) to approximate the factorials in $C(n, x)$ in (7.3) and make other approximations to find

$$(8.3) \qquad f(x) = C(n,x)p^x q^{n-x} \sim \frac{1}{\sqrt{2\pi npq}} e^{-(x-np)^2/(2npq)}.$$

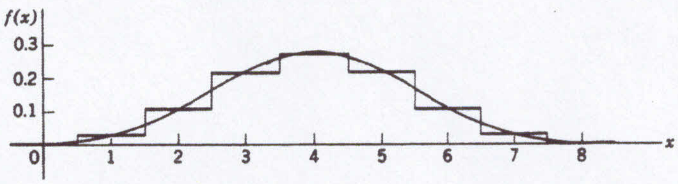


Figure 8.2 Binomial distribution for $n = 8$, $p = \frac{1}{2}$, and the normal approximation.

The sign $\sim$ means (as in Chapter 11, Section 11) that the ratio of the exact binomial distribution (7.3) and the right-hand side of (8.3) tends to 1 as $n \to \infty$. An outline of a derivation of (8.3) is given in Problem 8, but you may be more impressed by doing some computer plotting of graphs like Figures 8.2 and 8.3 (Problems 9 and 10). Although we have said that equation (8.3) gives an approximation valid for large n, the agreement is quite good even for fairly small values of n. Figure 8.2 shows this for the case $n = 8$. The binomial distribution $f(x)$ is defined only for integral x; you should compare the values of $f(x)$ with the values of the approximating normal curve at integral values of x. When n is very large (Figure 8.3), a graph of the exact binomial distribution is very close to the normal approximation (Problem 9).

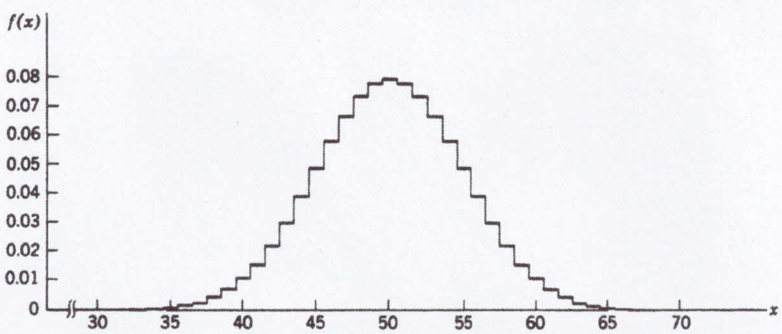


Figure 8.3 Binomial distribution for $n = 100$, $p = \frac{1}{2}$.

In (8.3), the left-hand side is the exact binomial distribution and the right-hand side is a normal distribution with $\mu = np$ and $\sigma = \sqrt{npq}$ as we see by comparing (8.3) and (8.1). Recall from Problems 7.9 and 7.13 that the mean value

μ and standard devi
binomial distributio

(8.4) For the b
$\mu = np.$

We can expect this
the normal approxim

Example 1. Find the pro
binomial distributio
See Figure (8.3)
$n = 100, p = \frac{1}{2}$. We
you could also read
For the normal
$\sigma = \sqrt{npq} = \sqrt{100}$
$\sigma = 5$, we find by co

Example 2. Find the prob
coin, that is $45 \le x$
As in Example 1
cumulative binomial
terms; we want the
can find $F(55)$, the b
the probability of 55
of 44 heads or less.
0.72875.
For the normal
normal $F(55)$ − nor
integrating from 44.
under the exact bino
and $x = 55$. This giv

Standard Normal
for the special case μ
and the correspondin

(8.5) $\phi(z) =$
$\Phi(z) =$

The cumulative distr
ter 11, Section 9).

μ and standard deviation σ for a random variable whose probability function is the binomial distribution (7.3) are also $\mu = np$ and $\sigma = \sqrt{npq}$.

(8.4) For the binomial distribution and its normal approximation,

$$\mu = np, \qquad \sigma = \sqrt{npq}\,.$$

We can expect this in general; whatever the μ and σ are for a given distribution, the normal approximation will have the same μ and σ.

Example 1. Find the probability of exactly 52 heads in 100 tosses of a coin using the binomial distribution and using the normal approximation.

See Figure (8.3) which is a plot of the binomial probability density function with $n = 100, p = \frac{1}{2}$. We find by computer for $x = 52$, binomial $f(52) = 0.07353$, which you could also read approximately from Figure (8.3).

For the normal approximation, we find from (8.4), $\mu = np = 100 \cdot \frac{1}{2} = 50$, $\sigma = \sqrt{npq} = \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 5$. Then for the normal approximation with $\mu = 50$, $\sigma = 5$, we find by computer for $x = 52$, normal $f(52) = 0.07365$.

Example 2. Find the probability $P(45, 55)$ of between 45 and 55 heads in 100 tosses of a coin, that is $45 \le x \le 55$.

As in Example 1, for the binomial distribution we have $n = 100, p = \frac{1}{2}$. The cumulative binomial distribution function $F(x)$ in (7.4) gives $P(45, 55)$ as a sum of terms; we want the sum of the 11 terms with $x = 45, 46, \cdots 55$. By computer, we can find $F(55)$, the binomial cumulative distribution function with $x = 55$, which is the probability of 55 heads or less, and then find and subtract $F(44)$, the probability of 44 heads or less. Thus we find $P(45, 55) =$ binomial $F(55)$ − binomial $F(44) =$ 0.72875.

For the normal approximation, we find by computer from (8.2), $P(45, 55) =$ normal $F(55)$ − normal $F(45) = 0.68269$. We can get a better approximation by integrating from 44.5 to 55.5; this corresponds more closely to the appropriate area under the exact binomial graph in Figure 8.3 by including the whole steps at $x = 45$ and $x = 55$. This gives $P(44.5, 55.5) =$ normal $F(55.5)$ − normal $F(44.5) = 0.72867$.

Standard Normal Distribution This is just the normal distribution in (8.1) for the special case $\mu = 0$ and $\sigma = 1$. The density function is often denoted by $\phi(z)$, and the corresponding cumulative distribution function by $\Phi(z)$:

(8.5) Standard normal distribution

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

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

The cumulative distribution function $\Phi(z)$ is related to the error function (see Chapter 11, Section 9).

It is sometimes convenient to write the functions in (8.1) in terms of $\phi(z)$ and $\Phi(z)$. We can do this by making the change of variables $z = (x - \mu)/\sigma$. The result is (Problem 21)

$$\text{(8.6)} \qquad \begin{aligned} f(x) &= \frac{1}{\sigma}\phi(z), \\ F(x) &= \Phi(z), \end{aligned} \qquad \text{where } z = \frac{(x-\mu)}{\sigma}.$$

The functions $\phi(z)$ and $\Phi(z)$ [or sometimes $\Phi(z) - \frac{1}{2}$] are tabulated so you can use either tables or computer to do problems.

Example 3. Find the number r such that the area under the normal distribution curve $y = f(x)$ from $\mu - r$ to $\mu + r$ is equal to 1/2.

Look at Figure 8.1 and recall that the area from $-\infty$ to ∞ is 1 and that the graph is symmetric about $x = \mu$. Then the integral from $-\infty$ to $\mu - r$ and the integral from $\mu + r$ to ∞ are equal to each other and so each is equal to 1/4. Thus the integral from $-\infty$ to $\mu + r$ must be 3/4, that is $F(\mu + r) = 3/4$. By (8.6) this is $\Phi(z) = 3/4$ where $z = (\mu + r - \mu)/\sigma = r/\sigma$. By computer or tables we find that if $\Phi(z) = 3/4$, then $z = 0.6745$. Thus $r = 0.6745\sigma$.

Example 4. You have taken a test (academic like the SAT, or medical like a bone density test) and a report gives your z-score as 1.14. What percent of your peers scored higher than you?

If we call the actual test scores x, and their average is μ and standard deviation σ, then the term z-score means the value of $z = (x - \mu)/\sigma$ as in (8.6). (In words, the z-score is the difference between x and its average, measured in units of the standard deviation.) Now we want the area $1 - F(x) = 1 - \Phi(z)$ by (8.6). By computer (or tables) we find $\Phi(1.14) = 0.87$; then $1 - 0.87 = 0.13$, so 13% of your peers scored higher than you. If your z-score is negative, then you are below average—bad if it's a physics test, good if it's your cholesterol! For example, if $z = -0.25$, then $\Phi(z) = 0.40$, so 60% of your peers scored higher than you.

Example 5. Suppose that boxes of a certain kind of cereal have an average weight of 16 ounces and it is known that 70% of the boxes weigh within 1 ounce of the average. What is the probability that the box you buy weighs less than 14 ounces?

If x represents the weight of a box, then we are given that the probability of $15 < x < 17$ is 0.7. Assuming a normal distribution, the area under the $f(x)$ curve up to $x = \mu = 16$ is $\frac{1}{2}$ and the area from $x = 16$ to $x = 17$ is half of 0.7 (by symmetry; see Figure 8.1). Thus $F(17) = 0.5 + 0.35 = 0.85$. We want to find the probability that $x < 14$; this is $F(14)$. Using (8.6), $x = 17$ gives $z = (17 - 16)/\sigma = 1/\sigma$, and similarly $x = 14$ gives $z = -2/\sigma$. So we are given $\Phi(1/\sigma) = 0.85$, and we want to find $\Phi(-2/\sigma)$. By computer (or tables) we find that if $\Phi(1/\sigma) = 0.85$, then $1/\sigma = 1.0364$, so $2/\sigma = 2.0728$, and $\Phi(-2/\sigma) = 0.019$. So there is almost a 2% chance that we would get a box weighing less than 14 ounces.

Note that in Examples 4 and 5 we assumed a normal distribution with no obvious justification. It is a very interesting and useful fact that such an assumption is

reasonable if the nu
at the end of Sectio

PROBLEMS, SECTION 8

1. Verify that for
the mean value
$-\infty$ to ∞ is 1
the integrals
(6.3), and (6.4)

2. Do Problem 6.4

3. The probability
of an ideal gas
velocity, m is t
the Boltzmann
deviation of v,

4. Computer plot
$\sigma = 1$, and with

5. Computer plot
2, and 5. Label

6. Do Problem 5 f

7. By computer fin
$\mu + 2\sigma$, $\mu + 3\sigma$,
μ and σ. Find
its mean value
See Figure (8.1)
value is the area
$\frac{1}{2}$ (that is the ar
result.

8. Carry through th
an approximatio
by Stirling's form

Show that if $\delta =$
for x and $n - x$
(ignore the squa
that

and a similar form
of $\delta/(np)$, collect

ln

Hence

reasonable if the number of measurements is very large. We will discuss this further at the end of Section 10.

PROBLEMS, SECTION 8

1. Verify that for a random variable x with normal density function $f(x)$ as in (8.1), the mean value of x is μ , the standard deviation is σ, and the integral of $f(x)$ from $-\infty$ to ∞ is 1 as it must be for a probability function. *Hint:* Write and evaluate the integrals $\int_{-\infty}^{\infty} f(x)\,dx$, $\int_{-\infty}^{\infty} xf(x)\,dx$, $\int_{-\infty}^{\infty}(x-\mu)^2 f(x)\,dx$. See equations (6.2), (6.3), and (6.4).

2. Do Problem 6.4 by comparing e^{-ax^2} with $f(x)$ in (8.1).

3. The probability density function for the x component of the velocity of a molecule of an ideal gas is proportional to $e^{-mv^2/(2kT)}$ where v is the x component of the velocity, m is the mass of the molecule, T is the temperature of the gas and k is the Boltzmann constant. By comparing this with (8.1), find the mean and standard deviation of v, and write the probability density function $f(v)$.

4. Computer plot on the same axes the normal probability density functions with $\mu = 0$, $\sigma = 1$, and with $\mu = 3$, $\sigma = 1$ to note that they are identical except for a translation.

5. Computer plot on the same axes the normal density functions with $\mu = 0$ and $\sigma = 1$, 2, and 5. Label each curve with its σ.

6. Do Problem 5 for $\sigma = \frac{1}{6}, \frac{1}{3}, 1$.

7. By computer find the value of the normal cumulative distribution function at $\mu+\sigma$, $\mu+2\sigma$, $\mu+3\sigma$, and satisfy yourself that these are independent of your choices for μ and σ. Find the probabilities that x is within 1, 2, or 3 standard deviations of its mean value μ to verify the results stated in the paragraph following (8.2). *Hint:* See Figure (8.1). The probability that x is within 1 standard deviation of its mean value is the area from $\mu-\sigma$ to $\mu+\sigma$; this is twice the area from μ to $\mu+\sigma$. Subtract $\frac{1}{2}$ (that is the area from $-\infty$ to μ) from your value of $F(\mu+\sigma)$ and then double the result.

8. Carry through the following details of a derivation of (8.3). Start with (7.3); we want an approximation to (7.3) for large n. First approximate the factorials in $C(n,x)$ by Stirling's formula (Chapter 11, Section 11) and simplify to get

$$f(x) \sim \left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{n-x} \sqrt{\frac{n}{2\pi x(n-x)}}.$$

Show that if $\delta = x - np$, then $x = np + \delta$ and $n - x = nq - \delta$. Make these substitutions for x and $n-x$ in the approximate $f(x)$. To evaluate the first two factors in $f(x)$ (ignore the square root for now): Take the logarithm of the first two factors; show that

$$\ln\frac{np}{x} = -\ln\left(1+\frac{\delta}{np}\right)$$

and a similar formula for $\ln[nq/(n-x)]$; expand the logarithms in a series of powers of $\delta/(np)$, collect terms and simplify to get

$$\ln\left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{n-x} \sim -\frac{\delta^2}{2npq}\left(1+\text{powers of }\frac{\delta}{n}\right).$$

Hence

$$\left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{n-x} \sim e^{-\delta^2/(2npq)}$$

for large n. [We really want δ/n small, that is, x near enough to its average value np so that $\delta/n = (x - np)/n$ is small. This means that our approximation is valid for the central part of the graph (see Figures 7.1 to 7.3) around $x = np$ where $f(x)$ is large. Since $f(x)$ is negligibly small anyway for x far from np, we ignore the fact that our approximation may not be good there. For more detail on this point, see Feller, p. 192]. Returning to the square root factor in $f(x)$, approximate x by np and $n - x$ by nq (assuming $\delta \ll np$ or nq) and obtain (8.3).

9. Computer plot a graph like Figure 8.3 of the binomial distribution with $n = 1000$, $p = \frac{1}{2}$, and observe that you have practically the corresponding normal approximation.

10. Computer plot graphs like Figure 8.2 but with $p \neq \frac{1}{2}$ to see that as n increases, the normal approximation becomes good (at least in the region around $x = \mu$ where the probabilities are large) even though the binomial graph is not symmetric (see Figure 7.3).

As in Examples 1 and 2, use (a) the binomial distribution; (b) the corresponding normal approximation, to find the probabilities of each of the following:

11. Exactly 50 heads in 100 tosses of a coin.

12. Exactly 120 aces in 720 tosses of a die.

13. Between 100 and 140 aces in 720 tosses of a die.

14. Between 499,000 and 501,000 heads in 10^6 tosses of a coin.

15. Exactly 195 tails in 400 tosses of a coin.

16. Between 195 and 205 tails in 400 tosses of a coin.

17. Exactly 31 4's in 180 tosses of a die.

18. Between 29 and 33 4's in 180 tosses of a die.

19. Exactly 21 successes in 100 Bernoulli trials with probability $\frac{1}{5}$ of success.

20. Between 17 and 21 successes in 100 Bernoulli trials with probability $\frac{1}{5}$ of success.

21. Verify equations (8.6). *Hints:* In $F(x)$, let $u = (t - \mu)/\sigma$; note that $dt = \sigma du$. What is u when $t = -\infty$? When $t = x$? Remember that by definition $z = (x - \mu)/\sigma$.

22. Using (8.6), do Problem 7.

23. Using (8.6), find h such that 90% of the area under a normal $f(x)$ lies between $\mu - h$ and $\mu + h$. Repeat for 95%. *Hint:* See Example 3.

24. Write out a proof of Chebyshev's inequality (see end of Section 7) for the case of a continuous probability function $f(x)$.

25. An instructor who grades "on the curve" computes the mean and standard deviation of the grades, and then, assuming a normal distribution with this μ and σ, sets the border lines between the grades at: C from $\mu - \frac{1}{2}\sigma$ to $\mu + \frac{1}{2}\sigma$, B from $\mu + \frac{1}{2}\sigma$ to $\mu + \frac{3}{2}\sigma$, A from $\mu + \frac{3}{2}\sigma$ up, etc. Find the percentages of the students receiving the various grades. Where should the border lines be set to give the percentages A and F, 10%; B and D, 20%; C, 40%?

9. THE POISSON DISTRI

The Poisson distribut
of some occurrence is
is also a good approx
np is small even thou

Let's derive the P
Suppose we observe a
radioactive substance
the half-life of the su
during the experimen
small time interval Δt
of two particles durin
observing exactly n c
the probability of obs
the sum of the probal
none in Δt" and "(n -

(9.1) P_n

Now $P_1(\Delta t)$ is the pr
Then the probability
these values into (9.1)

(9.2) P_n

or,

(9.3)

Letting $\Delta t \to 0$, we ha

(9.4)

For $n = 0$, (9.1) simp
particles in Δt," and (

(9.5)

Then, since $P_0(0) =$
interval" = 1, integrat

(9.6)

Substituting (9.6) into
solution (Problem 1) is
$P_2, P_3, \cdots, P_n$, we obt

(9.7)

ugh to its average value np
approximation is valid for
und $x = np$ where $f(x)$ is
om np, we ignore the fact
e detail on this point, see
(x), approximate x by np
3).

stribution with $n = 1000$,
nding normal approxima-

e that as n increases, the
ion around $x = \mu$ where
ph is not symmetric (see

he corresponding normal

$\frac{1}{5}$ of success.

obability $\frac{1}{5}$ of success.

e that $dt = \sigma du$. What
ion $z = (x - \mu)/\sigma$.

$f(x)$ lies between $\mu - h$

ion 7) for the case of a

and standard deviation
this μ and σ, sets the
$\frac{1}{2}\sigma$. B from $\mu + \frac{1}{2}\sigma$ to
the students receiving
give the percentages:

9. THE POISSON DISTRIBUTION

The Poisson distribution is useful in a variety of problems in which the probability of some occurrence is small and constant. (See Example 1 and Problems 3 to 9.) It is also a good approximation to the binomial distribution when p is so small that np is small even though n is large (see Example 2).

Let's derive the Poisson distribution by considering the following experiment. Suppose we observe and count the number of particles emitted per unit time by a radioactive substance. We assume that our period of observation is much less than the half-life of the substance, so that the average counting rate does not decrease during the experiment. Then the probability that one particle is emitted during a small time interval Δt is $\mu\Delta t$, $\mu =$const., if Δt is short enough so that the probability of two particles during Δt is negligible. We want to find the probability $P_n(t)$ of observing exactly n counts during a time interval t. The probability $P_n(t+\Delta t)$ is the probability of observing n counts in the time interval $t + \Delta t$. For $n > 0$, this is the sum of the probabilities of the two mutually exclusive events, "n particles in t, none in Δt" and "$(n-1)$ particles in t, one in Δt"; in symbols,

$$P_n(t+\Delta t) = P_n(t)P_0(\Delta t) + P_{n-1}(t)P_1(\Delta t). \tag{9.1}$$

Now $P_1(\Delta t)$ is the probability of one particle in Δt; this, by assumption, is $\mu\Delta t$. Then the probability of no particles in Δt is $1 - P_1(\Delta t) = 1 - \mu\Delta t$. Substituting these values into (9.1), we get

$$P_n(t+\Delta t) = P_n(t)(1-\mu\Delta t) + P_{n-1}(t)\mu\Delta t, \tag{9.2}$$

or,

$$\frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = \mu P_{n-1}(t) - \mu P_n(t). \tag{9.3}$$

Letting $\Delta t \to 0$, we have

$$\frac{dP_n(t)}{dt} = \mu P_{n-1}(t) - \mu P_n(t). \tag{9.4}$$

For $n = 0$, (9.1) simplifies since the only possible event is "no particles in t, no particles in Δt," and (9.4) becomes, for $n = 0$,

$$\frac{dP_0(t)}{dt} = -\mu P_0(t). \tag{9.5}$$

Then, since $P_0(0) =$ "probability that no particle is emitted during a zero time interval" $= 1$, integration of (9.5) gives

$$P_0 = e^{-\mu t}. \tag{9.6}$$

Substituting (9.6) into (9.4) with $n = 1$ gives a differential equation for $P_1(t)$; its solution (Problem 1) is $P_1(t) = \mu t e^{-\mu t}$. Solving (9.4) successively (Problem 1) for $P_2, P_3, \cdots, P_n$, we obtain

$$P_n(t) = \frac{(\mu t)^n}{n!}e^{-\mu t}. \tag{9.7}$$

Putting $t = 1$, we get for the probability of exactly n counts per unit time

$$P_n = \frac{\mu^n}{n!} e^{-\mu}. \qquad \text{Poisson distribution} \tag{9.8}$$

The probability density function (9.8) is called the *Poisson distribution* or the *Poisson probability density function.* You can show (Problem 2) that for the random variable n, the mean (that is the average number of counts per unit time) is μ, and the variance is also μ so the standard deviation is $\sqrt{\mu}$.

Example 1. The number of particles emitted each minute by a radioactive source is recorded for a period of 10 hours; a total of 1800 counts are registered. During how many 1-minute intervals should we expect to observe no particles; exactly one; etc.?

The average number of counts per minute is $1800/(10\cdot 60) = 3$ counts per minute; this is the value of μ. Then by (9.8), the probability of n counts per minute is

$$P_n = \frac{3^n}{n!} e^{-3}.$$

A graph of this probability function is shown in Figure 9.1. For $n = 0$, we find $P_0 = e^{-3} = 0.05$; then we should expect to observe no particles in about 5% of the 600 1-minute intervals, that is, during 30 1-minute intervals. Similarly we could compute the expected number of 1-minute intervals during which $1, 2, \cdots,$ particles would be observed.

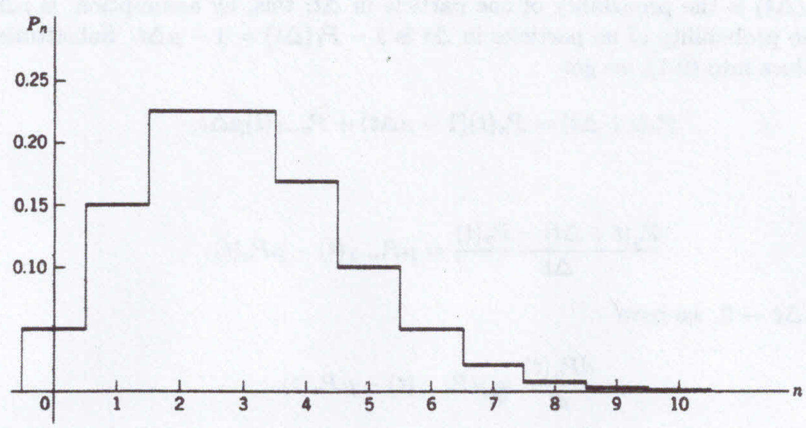


Figure 9.1 Poisson distribution $\mu = 3$.

Poisson Approximation of the Binomial Distribution In Section 8, we discussed the fact that the binomial distribution can be approximated by the normal distribution for large n and large np. If p is very small so that np is very much less than n (say, for example, $p = 10^{-3}, n = 2000, np = 2$), the normal approximation is not good. In this case you can show (Problem 10) that the Poisson distribution gives a good approximation to the binomial distribution (7.3), that is, that

(9.9) $C(n,$

[The exact meaning o
proaches 1 as $n \to \infty$

Example 2. If 1500 people
the probability that 2
The answer is give
$x = 2$. This is

$C(n, x$

(Or from your compu
$p = 1/500$, $x = 2$, is 0
the Poisson approxim
0.2240. (Or from yo
$\mu = 3$, $x = 2$, is 0.224
the same axes the bin
distribution with $\mu =$
(Problem 12).

Approximations by
distributions can be a
are both large, and
Poisson distribution
distribution as in (9.1

(9.10)

Note that the normal
Poisson distribution
variance). It is usef
distribution and their

PROBLEMS, SECTION 9

1. Solve the sequen
in (9.5) and (9.6

2. Show that the a
the Poisson distr
deviation of the
differentiate it a
differentiate the

$$(9.9) \qquad C(n,x)p^x q^{n-x} \sim \frac{(np)^x e^{-np}}{x!}, \qquad \text{Large } n, \text{ small } p.$$

[The exact meaning of (9.9) is that, for any fixed x, the ratio of the two sides approaches 1 as $n \to \infty$ and $p \to 0$ with np remaining constant.]

Example 2. If 1500 people each select a number at random between 1 and 500, what is the probability that 2 people selected the number 29?

The answer is given by the binomial distribution (7.3) with $n = 1500$, $p = 1/500$, $x = 2$. This is

$$C(n,x)p^x q^{n-x} = \frac{1500!}{2!1498!}\left(\frac{1}{500}\right)^2\left(\frac{499}{500}\right)^{998} = 0.2241.$$

(Or from your computer: the binomial probability density function with $n = 1500$, $p = 1/500$, $x = 2$, is 0.2241 to four decimal places.). A simpler formula from (9.9) is the Poisson approximation with $\mu = np = 3$, $x = 2$, namely $\mu^x e^{-x}/x! = 3^2 e^{-2}/2! = 0.2240$. (Or from your computer, the Poisson probability density function with $\mu = 3$, $x = 2$, is 0.2240 to four decimal places.) It is interesting to computer plot on the same axes the binomial distribution with $n = 1500$, $p = 1/500$, and the Poisson distribution with $\mu = 3$ as in Figure 9.1 to discover that they are almost identical (Problem 12).

Approximations by the Normal Distribution We have commented that many distributions can be approximated by the normal distribution when n and $\mu = np$ are both large, and have shown this for the binomial distribution in (8.1). The Poisson distribution when μ is large is also fairly well approximated by the normal distribution as in (9.10).

$$(9.10) \qquad \frac{\mu^x e^{-\mu}}{x!} \cong \frac{1}{\sqrt{2\pi\mu}} e^{-(x-\mu)^2/(2\mu)}, \qquad \mu \text{ large.}$$

Note that the normal distribution in (9.10) has the same mean and variance as the Poisson distribution it is approximating (see Problem 2 for the Poisson mean and variance). It is useful to computer plot on the same axes graphs of the Poisson distribution and their normal approximations (Problem 13).

PROBLEMS, SECTION 9

1. Solve the sequence of differential equations (9.4) for successive n values [as started in (9.5) and (9.6)] to obtain (9.7).

2. Show that the average value of a random variable n whose probability function is the Poisson distribution (9.8) is the number μ in (9.8). Also show that the standard deviation of the random variable is $\sqrt{\mu}$. *Hint:* Write the infinite series for e^x, differentiate it and multiply by x to get $xe^x = \sum(nx^n/n!)$; put $x = \mu$. To find σ^2 differentiate the xe^x series again, etc.

3. In an alpha-particle counting experiment the number of alpha particles is recorded each minute for 50 hours. The total number of particles is 6000. In how many 1-minute intervals would you expect no particles? Exactly n particles, for $n = 1, 2, 3, 4, 5$? Plot the Poisson distribution.

4. Suppose you receive an average of 4 phone calls per day. What is the probability that on a given day you receive no phone calls? Just one call? Exactly 4 calls?

5. Suppose that you have 5 exams during the 5 days of exam week. Find the probability that on a given day you have no exams; just 1 exam; 2 exams; 3 exams.

6. If you receive, on the average, 5 email messages per day, in how many days out of a 365-day year would you expect to receive exactly 5 messages? Fewer than 5? Exactly 10? More than 10? Just 1? None at all?

7. In a club with 500 members, what is the probability that exactly two people have birthdays on July 4?

8. If there are 100 misprints in a magazine of 40 pages, on how many pages would you expect to find no misprints? Two misprints? Five misprints?

9. If there are, on the average, 7 defects in a new car, what is the probability that your new car has only 2 defects? That it has 6 or 7? That it has more than 10?

10. Derive equation (9.9) as follows: In $C(n, x)$, show that $n!/(n-x)! \sim n^x$ for fixed x and large n [write $n!/(n-x)!$ as a product of x factors, divide by n^x, and show that the limit is 1 as $n \to \infty$]. Then write $q^{n-x} = (1-p)^{n-x}$ as $(1-p)^n(1-p)^{-x} = (1-np/n)^n(1-p)^{-x}$; evaluate the limit of the first factor as $n \to \infty$, np fixed; the limit of the second factor as $p \to 0$ is 1. Collect your results to obtain equation (9.9).

11. Suppose 520 people each have a shuffled deck of cards and draw one card from the deck. What is the probability that exactly 13 of the 520 cards will be aces of spades? Write the binomial formula and approximate it. Which is best, the normal or the Poisson approximation? Although you only need values at one x to answer the question, you might like to computer plot on the same axes graphs of the three distributions for the given n and p.

12. Computer plot on the same axes graphs of the binomial distribution in Example 2 and the Poisson and normal approximations.

13. Computer plot on the same axes a graph of the Poisson distribution and the corresponding normal approximation for the cases $\mu = 1, 5, 10, 20, 30$.

10. STATISTICS AND EXPERIMENTAL MEASUREMENTS

Statistics uses probability theory to consider sets of data and draw reasonable conclusions from them. So far in this chapter, we have been discussing problems for which we could write down a density function formula (normal, Poisson, etc.). Suppose that, instead, we have only a table of data, say a set of laboratory measurements of some physical quantity. Presumably, if we spent more time, we could enlarge this table of data as much as we liked. We can then imagine an infinite set of measurements of which we have only a sample. The infinite set is called the *parent population* or *universe*. What we would really like to know is the probability function for the parent population, or at least the average value μ (often thought of as the "true" value of the quantity being measured) and the standard deviation σ of the parent population. We must content ourselves with the best estimates we can make of these quantities using our available sample, that is, the set of measurements which we have made.

Estimate of Popul
median of our measur
and smaller measurer
times, that is the mo
of μ is, however, the
sample mean $\bar{x} = (1/$

(10.1) Estim

For a large set of mea
lem 1). Assuming tha
density function $f(x$
to show (Problem 2)
of $\bar{x}$ is $\sigma/\sqrt{n}$. Now
variable is unlikely t
deviations. For our
than a few multiples
an increasingly good
Note that this just s
that the average of a
than the average of
be too large, but it's

Estimate of Popul
be $s^2 = (1/n)\sum_{i=1}^{n}($
find the expected val
with mean μ and va
We conclude that a

(10.2) Estim

(*Caution:* The term
ence books, compute
check the definition
term.)

The quantity σ w
parent population wh
measurement x. The
the different possible
the value we are ap
roughly the spread o
measurement, it is of

Estimate of Population Average As a quick estimate of μ we might take the median of our measurements x_i (a value such that there are equal numbers of larger and smaller measurements), or the mode (the measurement we obtained the most times, that is the most probable measurement). The most frequently used estimate of μ is, however, the arithmetic mean (or average) of the measurements, that is the sample mean $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$. Thus we have

$$(10.1) \qquad \text{Estimate of population mean is } \mu \simeq \bar{x} = (1/n)\sum_{i=1}^{n} x_i.$$

For a large set of measurements we can justify this choice as follows (also see Problem 1). Assuming that the parent population for our measurements has probability density function $f(x)$ with expected value μ and standard deviation σ, it is easy to show (Problem 2) that the expected value of $\bar{x}$ is μ and the standard deviation of $\bar{x}$ is $\sigma/\sqrt{n}$. Now Chebyshev's inequality (end of Section 7) says that a random variable is unlikely to differ from its expected value by more than a few standard deviations. For our problem this says that $\bar{x}$ is unlikely to differ from μ by more than a few multiples of $\sigma/\sqrt{n}$, which becomes small as n increases. Thus $\bar{x}$ becomes an increasingly good estimate of μ as we increase the number n of measurements. Note that this just says mathematically what you would assume from experience, that the average of a large number of measurements is more likely to be accurate than the average of a small number. For example, two measurements might both be too large, but it's unlikely that 20 would all be too large.

Estimate of Population Variance Our first guess for an estimate of σ^2 might be $s^2 = (1/n)\sum_{i=1}^{n}(x_i - \bar{x})^2$, but we would be wrong. To see what is reasonable, we find the expected value of s^2 assuming that our measurements are from a population with mean μ and variance σ^2. The result is (Problem 3), $E(s^2) = [(n-1)/n]\sigma^2$. We conclude that a reasonable estimate of σ^2 is $\frac{n}{n-1}s^2$.

$$(10.2) \qquad \text{Estimate of population variance is } \sigma^2 \simeq \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

(*Caution:* The term "sample variance" is used in various references—texts, reference books, computer programs—to mean either our s^2 or our estimate of σ^2, so check the definition carefully in any reference you use. We shall avoid using the term.)

The quantity σ which we have just estimated is the standard deviation for the parent population whose probability function we call $f(x)$. Consider just a single measurement x. The function $f(x)$ (if we knew it) would give us the probabilities of the different possible values of x, the population mean μ would tell us approximately the value we are apt to find for x, and the standard deviation σ would tell us roughly the spread of x values about μ. Since σ tells us something about a single measurement, it is often called the *standard deviation of a single measurement.*

Standard Deviation of the Mean; Standard Error Instead of a single measurement, let us consider $\overline{x}$, the average (mean) of a set of n measurements. (The mean, $\overline{x}$, will be what we will use or report as the result of an experiment.) Just as we originally imagined obtaining the probability function $f(x)$ by making a large number of single measurements, so we can imagine obtaining a probability function $g(\overline{x})$ by making a large number of *sets* of n measurements with each set giving us a value of $\overline{x}$. The function $g(\overline{x})$ (if we knew it) would give us the probability of different values of $\overline{x}$. We have seen (Problem 2) that $\mathrm{Var}(\overline{x}) = \sigma^2/n$, so the *standard deviation of the mean* (that is, of $\overline{x}$) is

$$\sigma_m = \sqrt{\mathrm{Var}(\overline{x})} = \frac{\sigma}{\sqrt{n}}. \tag{10.3}$$

The quantity σ_m is also called the *standard error*; it gives us an estimate of the spread of values of $\overline{x}$ about μ. We see that the new probability function $g(\overline{x})$ must be much more peaked than $f(x)$ about the value μ because the standard deviation $\sigma/\sqrt{n}$ is much smaller than σ. Collecting formulas (10.2) and (10.3), we have

$$\sigma_m \cong \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n(n-1)}} \tag{10.4}$$

▸ **Example 1.** To illustrate our discussion, let's consider the following set of measurements: {7.2, 7.1, 6.7, 7.0, 6.8, 7.0, 6.9, 7.4, 7.0, 6.9}. [Note that, to show methods but minimize computation, we consider unrealistically small sets of measurements.]

From (10.1) we find $\mu \simeq \overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{70}{10} = 7.0.$

From (10.2) we find $\sigma^2 \simeq \frac{1}{9}\sum_{i=1}^{10}(x_i - 7)^2 = \frac{0.36}{9} = 0.04, \sigma \simeq 0.2.$

From (10.4), the standard error is $\sigma_m \simeq \sqrt{\frac{0.36}{10 \cdot 9}} = 0.0632.$

Combination of Measurements We have discussed how we can use a set of measurements x_i to estimate μ (the population average) by $\overline{x}$ (the sample average) and to estimate the standard error $\sigma_{mx} = \sqrt{\mathrm{Var}(\overline{x})}$ [equation (10.4)]. Now suppose we have done this for two quantities, x and y, and we want to use a known formula $w = w(x, y)$ to estimate a value for w and the standard error in w. First we consider the simple example $w = x + y$. Then, by Problem 6.13,

$$E(w) = E(x) + E(y) = \mu_x + \mu_y \tag{10.5}$$

where μ_x and μ_y are population averages. As discussed above, we estimate μ_x and μ_y by $\overline{x}$ and $\overline{y}$ and conclude that a reasonable estimate of w is

$$\overline{w} = \overline{x} + \overline{y}. \tag{10.6}$$

Now let us assume t
Problem 6.15,

(10.7)

Next consider the
find $\overline{w} = 4 - 2\overline{x} + 3\overline{y}$
$\mathrm{Var}(Kx) = K^2\,\mathrm{Var}(x$

(10.8) $\mathrm{Var}(\overline{w})$

(10.9)

We can now see
be approximated by
namely (see Chapter

(10.10) $w(x,$

where the partial deri
[Practically speaking,
near zero—we can't e
the higher derivatives
point (μ_x, μ_y).] Assum
derivatives are consta

(10.11) $E[w(x,y)]$

Since we have agreed t
able estimate of w is

(10.12)

(This may look obviou
Then, putting $x =$
(10.11), we find as in (

$\mathrm{Var}(\overline{w})$

(10.13)

We can use (10.12) and
measured quantities x

Now let us assume that x and y are independently measured quantities. Then by Problem 6.15,

$$\text{(10.7)}\qquad \begin{aligned}\operatorname{Var}(\overline{w}) &= \operatorname{Var}(\overline{x}) + \operatorname{Var}(\overline{y}) = \sigma_{mx}^2 + \sigma_{my}^2,\\ \sigma_{mw} &= \sqrt{\sigma_{mx}^2 + \sigma_{my}^2}.\end{aligned}$$

Next consider the case $w = 4 - 2x + 3y$. As in equations (10.5) and (10.6), we find $\overline{w} = 4 - 2\overline{x} + 3\overline{y}$. Now by Problem 5.13, we have $\operatorname{Var}(x+K) = \operatorname{Var}(x)$, and $\operatorname{Var}(Kx) = K^2 \operatorname{Var}(x)$, where K is a constant. Thus,

$$\text{(10.8)}\qquad \begin{aligned}\operatorname{Var}(\overline{w}) &= \operatorname{Var}(4 - 2\overline{x} + 3\overline{y}) = \operatorname{Var}(-2\overline{x} + 3\overline{y})\\ &= (-2)^2 \operatorname{Var}(\overline{x}) + (3)^2 \operatorname{Var}(\overline{y}) = 4\sigma_{mx}^2 + 9\sigma_{my}^2,\end{aligned}$$

$$\text{(10.9)}\qquad \sigma_{mw} = \sqrt{4\sigma_{mx}^2 + 9\sigma_{my}^2}.$$

We can now see how to find $\overline{w}$ and σ_{mw} for any function $w(x, y)$ which can be approximated by the linear terms of its Taylor series about the point (μ_x, μ_y), namely (see Chapter 4, Section 2)

$$\text{(10.10)}\qquad w(x,y) \cong w(\mu_x, \mu_y) + \left(\frac{\partial w}{\partial x}\right)(x - \mu_x) + \left(\frac{\partial w}{\partial y}\right)(y - \mu_y)$$

where the partial derivatives are evaluated at $x = \mu_x$, $y = \mu_y$, and so are constants. [Practically speaking, this means that the first partial derivatives should not be near zero—we can't expect good results near a maximum or minimum of w—and the higher derivatives should not be large, that is, w should be "smooth" near the point (μ_x, μ_y).] Assuming (10.10), and remembering that $w(\mu_x, \mu_y)$ and the partial derivatives are constants, we find

$$\text{(10.11)}\qquad \begin{aligned}E[w(x,y)] &\cong w(\mu_x, \mu_y) + \left(\frac{\partial w}{\partial x}\right)[E(x) - \mu_x] + \left(\frac{\partial w}{\partial y}\right)[E(y) - \mu_y]\\ &= w(\mu_x, \mu_y).\end{aligned}$$

Since we have agreed to estimate μ_x and μ_y by $\overline{x}$ and $\overline{y}$, we conclude that a reasonable estimate of w is

$$\text{(10.12)}\qquad \overline{w} = w(\overline{x}, \overline{y}).$$

(This may look obvious, but see Problem 7.)

Then, putting $x = \overline{x}, y = \overline{y}$ in (10.10) and remembering the comment just before (10.11), we find as in (10.8)

$$\begin{aligned}\operatorname{Var}(\overline{w}) &= \operatorname{Var}[w(\overline{x}, \overline{y})]\\ &= \operatorname{Var}\left[w(\mu_x, \mu_y) + \left(\frac{\partial w}{\partial x}\right)(\overline{x} - \mu_x) + \left(\frac{\partial w}{\partial y}\right)(\overline{y} - \mu_y)\right]\\ &= \left(\frac{\partial w}{\partial x}\right)^2 \sigma_{mx}^2 + \left(\frac{\partial w}{\partial y}\right)^2 \sigma_{my}^2,\end{aligned}$$

$$\text{(10.13)}\qquad \sigma_{mw} = \sqrt{\left(\frac{\partial w}{\partial x}\right)^2 \sigma_{mx}^2 + \left(\frac{\partial w}{\partial y}\right)^2 \sigma_{my}^2}.$$

We can use (10.12) and (10.13) to estimate the value of a given function w of two measured quantities x and y and to find the standard error in w.

▸ **Example 2.** From Example 1 we have $\overline{x} = 7$ and $\sigma_{mx} = 0.0632$. Suppose we have also found from measurements that $\overline{y} = 5$ and $\sigma_{my} = 0.0591$. If $w = x/y$, find $\overline{w}$ and σ_{mw}. From (10.12) we have $\overline{w} = \overline{x}/\overline{y} = 7/5 = 1.4$. From (10.13) we find

$$\sigma_{mw} = \sqrt{\left(\frac{1}{\overline{y}}\right)^2 \sigma_{mx}^2 + \left(\frac{-\overline{x}}{\overline{y}^2}\right)^2 \sigma_{my}^2} = \sqrt{\left(\frac{1}{5}\right)^2 (0.0632)^2 + \left(\frac{-7}{25}\right)^2 (0.0591)^2}$$
$$= 0.0208.$$

Central Limit Theorem So far we have not assumed any special form (such as normal, etc.) for the density function $f(x)$ of the parent population, so that our results for computation of approximate values of μ, σ, and σ_m from a set of measurements apply whether or not the parent distribution is normal. (And, in fact, it may not be; for example, Poisson distributions are quite common.) You will find, however, that most discussions of experimental errors are based on an assumed normal distribution. Let us discuss the justification for this. We have seen above that we can think of the sample average $\overline{x}$ as a random variable with average μ and standard deviation $\sigma/\sqrt{n}$. We have said that we might think of a density function $g(\overline{x})$ for $\overline{x}$ and that it would be more strongly peaked about μ than the density function $f(x)$ for a single measurement, but we have not said anything so far about the form of $g(\overline{x})$. There is a basic theorem in probability (which we shall quote without proof) which gives us some information about the probability function for $\overline{x}$. The *central limit theorem* says that no matter what the parent probability function $f(x)$ is (provided μ and σ exist), the probability function for $\overline{x}$ is approximately the normal distribution with standard deviation $\sigma/\sqrt{n}$ if n is large.

Confidence Intervals, Probable Error If we assume that the probability function for $\overline{x}$ is normal (a reasonable assumption if n is large), then we can give a more specific meaning to σ_m (standard deviation of the mean) than our vague statement that it gives us an estimate of the spread of $\overline{x}$ values about μ. Since the probability for a normally distributed random variable to have values between $\mu - \sigma$ and $\mu + \sigma$ is 0.6827 (see Section 8 and Problem 8.7), we can say that the probability is about 68% for a measurement of $\overline{x}$ to lie between $\mu - \sigma_m$ and $\mu + \sigma_m$. This interval is called the 68% *confidence interval.* Similarly we can find an interval $\mu \pm r$ such that the probability is $\frac{1}{2}$ that a new measurement would fall in this interval (and so also the probability is $\frac{1}{2}$ that it would fall outside!), that is, a 50% confidence interval. From Section 8, Example 3, this is $r = 0.6745\sigma_m$. The number r is called the *probable error.* When we have found σ_m as in Examples 1 and 2, we just have to multiply it by 0.6745 to find the corresponding probable error. Similarly we can find the error corresponding to other choices of confidence interval (see Problem 4).

▸ PROBLEMS, SECTION 10

1. Let $m_1, m_2, \cdots, m_n$ be a set of measurements, and define the values of x_i by $x_1 = m_1 - a, x_2 = m_2 - a, \cdots, x_n = m_n - a$, where a is some number (as yet unspecified, but the same for all x_i). Show that in order to minimize $\sum_{i=1}^n x_i^2$, we should choose $a = (1/n)\sum_{i=1}^n m_i$. *Hint:* Differentiate $\sum_{i=1}^n x_i^2$ with respect to a. You have shown that the arithmetic mean is the "best" average in the least squares sense, that is, that if the sum of the squares of the deviations of the measurements from their

"average" is a minimum, the "average" is the arithmetic mean (rather than, say, the median or mode).

2. Let $x_1, x_2, \cdots, x_n$ be independent random variables, each with density function $f(x)$, expected value μ, and variance σ^2. Define the sample mean by $\bar{x} = \sum_{i=1}^{n} x_i$. Show that $E(\bar{x}) = \mu$, and $\mathrm{Var}(\bar{x}) = \sigma^2/n$. (See Problems 5.9, 5.13, and 6.15.)

3. Define s by the equation $s^2 = (1/n)\sum_{i=1}^{n}(x_i - \bar{x})^2$. Show that the expected value of s^2 is $[(n-1)/n]\sigma^2$. *Hints:* Write

$$\begin{aligned}(x_i - \bar{x})^2 &= [(x_i - \mu) - (\bar{x} - \mu)]^2 \\ &= (x_i - \mu)^2 - 2(x_i - \mu)(\bar{x} - \mu) + (\bar{x} - \mu)^2.\end{aligned}$$

Find the average value of the first term from the definition of σ^2 and the average value of the third term from Problem 2. To find the average value of the middle term write

$$(\bar{x} - \mu) = \left(\frac{x_1 + x_2 + \cdots + x_n}{n} - \mu\right) = \frac{1}{n}[(x_1 - \mu) + (x_2 - \mu) + \cdots + (x_n - \mu)].$$

Show by Problem 6.14 that

$$E[(x_i - \mu)(x_j - \mu)] = E(x_i - \mu)E(x_j - \mu) = 0 \qquad \text{for } i \neq j,$$

and evaluate $E[(x_i - \mu)^2]$ (same as the first term). Collect terms to find

$$E(s^2) = \frac{n-1}{n}\sigma^2.$$

4. Assuming a normal distribution, find the limits $\mu \pm h$ for a 90% confidence interval; for a 95% confidence interval; for a 99% confidence interval. What percent confidence interval is $\mu \pm 1.3\sigma$? *Hints:* See Section 8, Example 3, and Problems 8.7, 8.22, and 8.23.

5. Show that if $w = xy$ or $w = x/y$, then (10.14) gives the convenient formula for relative error

$$\frac{r_w}{w} = \sqrt{\left(\frac{r_x}{x}\right)^2 + \left(\frac{r_y}{y}\right)^2}.$$

6. By expanding $w(x, y, z)$ in a three-variable power series similar to (10.10), show that

$$r_w = \sqrt{\left(\frac{\partial w}{\partial x}\right)^2 r_x^2 + \left(\frac{\partial w}{\partial y}\right)^2 r_y^2 + \left(\frac{\partial w}{\partial z}\right)^2 r_z^2}.$$

7. Equation (10.12) is only an approximation (but usually satisfactory). Show, however, that if you keep the second order terms in (10.10), then

$$\bar{w} = w(\bar{x}, \bar{y}) + \frac{1}{2}\left(\frac{\partial^2 w}{\partial x^2}\right)\sigma_x^2 + \frac{1}{2}\left(\frac{\partial^2 w}{\partial y^2}\right)\sigma_y^2.$$

8. The following measurements of x and y have been made.

x : 5.1, 4.9, 5.0, 5.2, 4.9, 5.0, 4.8, 5.1

y : 1.03, 1.05, 0.96, 1.00, 1.02, 0.95, 0.99, 1.01, 1.00, 0.99

Find the mean value and the probable error of x, y, $x + y$, xy, $x^3 \sin y$, and $\ln x$. *Hint:* See Examples 1 and 2 and the last paragraph of this section.