Chapter 3

Concepts of Probability

21 August 2007

We introduce the basic concepts of probability and apply them to simple physical systems and everyday life. We will discover the universal nature of the central limit theorem and the Gaussian distribution for the sum of a large number of random variables and discuss its relation to why thermodynamics is possible. Because of the importance of probability in many contexts and the relatively little time it will take us to consider more advanced topics, our discussion goes beyond what we will need for the applications of statistical mechanics in these notes.

3.1 Probability in Everyday Life

One of our goals, which we will consider in Chapter 4 and subsequent chapters, is to relate the behavior of various macroscopic quantities to the underlying microscopic behavior of the individual atoms or other constituents. To do so, we will need to introduce some ideas from probability.

We all use the ideas of probability in everyday life. For example, every morning many of us decide what to wear based on the probability of rain. We cross streets knowing that the probability of being hit by a car is small. We can even make a rough estimate of the probability of being hit by a car. It must be less that one in a thousand, because you have crossed streets thousands of times and hopefully you have not been hit. You might be hit tomorrow, or you might have been hit the first time you tried to cross a street. These comments illustrate that we have some intuitive sense of probability, and because it is a useful concept for survival, we know how to estimate it. As expressed by Laplace (1819),

> Probability theory is nothing but common sense reduced to calculation.

Another interesting thought is due to Maxwell (1850): The true logic of this world is the calculus of probabilities ... That is, probability is a natural language for describing real world phenomena.

However, our intuition only takes us so far. Consider airplane travel. Is it safe to fly? Suppose that there is a one chance in 100,000 of a plane crashing on a given flight and that there are a 1000 flights a day. Then every 100 days or so there would be a reasonable likelihood of a plane crash.

This estimate is in rough accord with what we read. For a given flight, your chances of crashing are approximately one part in 10^5, and if you fly five times a year for 100 years, it seems that flying is not too much of a risk. Suppose that instead of living 100 years, you could live 20,000 years. In this case you would take 100,000 flights, and it would be much more risky to fly if you wished to live your full 20,000 years. Although this last statement seems reasonable, can you explain why?

Much of the motivation for the mathematical formulation of probability arose from the proficiency of professional gamblers in estimating betting odds and their desire to have more quantitative measures. Although games of chance have been played since history has been recorded, the first steps toward a mathematical formulation of games of chance began in the middle of the 17th century. Some of the important contributors over the following 150 years include Pascal, Fermat, Descartes, Leibnitz, Newton, Bernoulli, and Laplace, names that are probably familiar to you.

Given the long history of games of chance and the interest in estimating probability in a variety of contexts, it is remarkable that the theory of probability took so long to develop. One reason is that the idea of probability is subtle and is capable of many interpretations. An understanding of probability is elusive due in part to the fact that the probably depends on the status of the information that we have (a fact well known to poker players). Although the rules of probability are defined by simple mathematical rules, an understanding of probability is greatly aided by experience with real data and concrete problems. To test your current understanding of probability, try to solve Problems 3.1–3.6 before reading the rest of this chapter. Then in Problem 3.7 formulate the laws of probability as best as you can based on your solutions to these problems.

Problem 3.1. Marbles in a jar

A jar contains 2 orange, 5 blue, 3 red, and 4 yellow marbles. A marble is drawn at random from the jar. Find the probability that

(a) the marble is orange;

(b) the marble is red;

(c) the marble is orange or blue.

Problem 3.2. Piggy bank

A piggy bank contains one penny, one nickel, one dime, and one quarter. It is shaken until two coins fall out at random. What is the probability that at least $0.30 falls out?

Problem 3.3. Two dice

A girl tosses a pair of dice at the same time. Find the probability that

(a) both dice show the same number;

(b) both dice show a number less than 5;

(c) both dice show an even number;

(d) the product of the numbers is 12.

Problem 3.4. Free throws

A boy hits 16 free throws out of 25 attempts. What is the probability that he will make a free throw on his next attempt?

Problem 3.5. Toss of a die

Consider an experiment in which a die is tossed 150 times and the number of times each face is observed is counted. The value of A, the number of dots on the face of the die and the number of times that it appeared is shown in Table 3.1.

(a) What is the predicted average value of A assuming a fair die?

(b) What is the average value of A observed in this experiment?

value of A	frequency
1	23
2	28
3	30
4	21
5	23
6	25

Table 3.1: The number of times face A appeared in 150 tosses.

Problem 3.6. What's in your purse?

A coin is taken at random from a purse that contains one penny, two nickels, four dimes, and three quarters. If x equals the value of the coin, find the average value of x.

Problem 3.7. Rules of probability

Based on your solutions to the above problems, state the rules of probability as you understand them at this time.

The following problems are related to the use of probability in everyday life.

Problem 3.8. Choices

Suppose that you are offered the following choice:

(a) A certain prize of $50.

(b) You flip a (fair) coin and win $100 if you get a head, but $0 if you get a tail. Which choice would you make? Explain your reasoning. Would your choice change if the certain prize was $40?

Problem 3.9. More choices

Suppose that you are offered the following choices:

(a) A prize of $100 is awarded for each head found in ten flips of a coin, or

(b) a certain prize of $400. What choice would you make? Explain your reasoning.

Problem 3.10. Thinking about probability

(a) Suppose that you were to judge an event to be 99.9999% probable. Would you be willing to bet $999 999 against $1 that the event would occur? Discuss why probability assessments should be kept separate from decision issues.

(b) Suppose that someone gives you a dollar to play the lottery. What sequence of six numbers between 1 and 36 would you choose?

(c) Suppose you toss a coin 8 times and obtain heads each time. Estimate the probability that you will obtain heads on your ninth toss.

(d) What is the probability that it will rain tomorrow? What is the probability that the Dow Jones industrial average will increase tomorrow?

(e) Give several examples of the use of probability in everyday life. Distinguish between various types of probability.

3.2 The Rules of Probability

We now summarize the basic rules and ideas of probability.[1] Suppose that there is an operation or a process that has several distinct possible *outcomes*. The process might be the flip of a coin or the roll of a six-sided die.[2] We call each flip a *trial*. The list of all the possible *events* or *outcomes* is called the *sample space*. We assume that the events are *mutually exclusive*, that is, the occurrence of one event implies that the others cannot happen at the same time. We let n represent the number of events, and label the events by the index i which varies from 1 to n. For now we assume that the sample space is finite and discrete. For example, the flip of a coin results in one of two events that we refer to as heads and tails and the role of a die yields one of six possible events.

For each event i, we assign a probability $P(i)$ that satisfies the conditions

$$P(i) \geq 0, \tag{3.1}$$

and

$$\sum_i P(i) = 1. \tag{3.2}$$

$P(i) = 0$ implies that the event cannot occur, and $P(i) = 1$ implies that the event must occur. The normalization condition (3.2) says that the sum of the probabilities of all possible mutually exclusive outcomes is unity.

[1] In 1933 the Russian mathematician A. N. Kolmogorov formulated a complete set of axioms for the mathematical definition of probability.

[2] The earliest known six-sided dice have been found in the Middle East. A die made of baked clay was found in excavations of ancient Mesopotamia. The history of games of chance is discussed by Deborah J. Bennett, *Randomness*, Harvard University Press (1998).

Example 3.1. Let x be the number of points on the face of a die. What is the sample space of x?

Solution. The sample space or set of possible events is $x_i = \{1, 2, 3, 4, 5, 6\}$. These six outcomes are mutually exclusive.

The rules of probability will be summarized further in (3.3) and (3.5). These abstract rules must be supplemented by an *interpretation* of the term probability. As we will see, there are many different interpretations of probability because any interpretation that satisfies the rules of probability may be regarded as a kind of probability.

An interpretation of probability that is relatively easy to understand is based on *symmetry*. Suppose that we have a two-sided coin that shows heads and tails. Then there are two possible mutually exclusive outcomes, and if the coin is perfect, each outcome is equally likely.[3] If a die with six distinct faces (see Figure 3.1) is perfect, we can use symmetry arguments to argue that each outcome should be counted equally and $P(i) = 1/6$ for each of the six faces. For an actual die, we can estimate the probability *a posteriori*, that is, by the observation of the outcome of many throws. As is usual in physics, our intuition will lead us to the concepts.

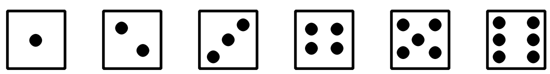

Figure 3.1: The six possible outcomes of the toss of a die.

Suppose that we know that the probability of rolling any face of a die in one throw is equal to 1/6, and we want to find the probability of finding face 3 *or* face 6 in one throw. In this case we wish to know the probability of a trial that is a combination of more elementary operations for which the probabilities are already known. That is, we want to know the probability of the outcome, i *or* j, where i is distinct from j. According to the rules of probability, the probability of event i *or* j is given by

$$P(i \text{ or } j) = P(i) + P(j). \qquad \text{(addition rule)} \tag{3.3}$$

The relation (3.3) is generalizable to more than two events. An important consequence of (3.3) is that if $P(i)$ is the probability of event i, then the probability of event i not occurring is $1 - P(i)$.

Example 3.2. What is the probability of throwing a three or a six with one throw of a die?

Solution. The probability that the face exhibits either 3 or 6 is $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Example 3.3. What is the probability of *not* throwing a six with one throw of die?

Solution. The answer is the probability of either "1 or 2 or 3 or 4 or 5." The addition rule gives that the probability $P(\text{not six})$ is

$$P(\text{not six}) = P(1) + P(2) + P(3) + P(4) + P(5) \tag{3.4a}$$

$$= 1 - P(6) = \frac{5}{6}, \tag{3.4b}$$

[3] Is the outcome of a coin toss really random? It appears that the randomness in a coin toss is introduced by sloppy humans. Each human-generated flip has a different height and speed and is caught at a different angle, giving different outcomes. See the references at the end of the chapter.

where the last relation follows from the fact that the sum of the probabilities for all outcomes sums to unity. It is very useful to take advantage of this property when solving many probability problems.

Another simple rule is for the probability of the joint occurrence of independent events. These events might be the probability of throwing a 3 on one die *and* the probability of throwing a 4 on a second die. If two events are independent, then the probability of both events occurring is the product of their probabilities

$$P(i \text{ and } j) = P(j)\,P(j). \qquad \text{(multiplication rule)} \tag{3.5}$$

Events are independent if the occurrence of one event does not change the probability for the occurrence of the other.

To understand the applicability of (3.5) and the meaning of the independence of events, consider the problem of determining the probability that a person chosen at random is a female over six feet tall. Suppose that we know that the probability of a person to be over six feet tall is $P(6^+) = \frac{1}{5}$, and the probability of being female is $P(\text{female}) = \frac{1}{2}$. We might conclude that the probability of being a tall female is $P(\text{female})P(6^+) = \frac{1}{2} \times \frac{1}{5} = \frac{1}{10}$. The same probability would hold for a tall male. However, this reasoning is incorrect, because the probability of being a tall female differs from the probability of being a tall male. The problem is that the two events – being over six feet tall and being female – are not independent. On the other hand, consider the probability that a person chosen at random is female and was born on September 6. We can reasonably assume equal likelihood of birthdays for all days of the year, and it is correct to conclude that this probability is $\frac{1}{2} \times \frac{1}{365}$ (not counting leap years). Being a woman and being born on September 6 are independent events.

Problem 3.11. Give an example from your solutions to Problems 3.1–3.6 where you used the addition rule or the multiplication rule or both.

Example 3.4. What is the probability of throwing an even number with one throw of a die?

Solution. We can use the addition rule to find that

$$P(\text{even}) = P(2) + P(4) + P(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}. \tag{3.6}$$

Example 3.5. What is the probability of the same face appearing on two successive throws of a die?

Solution. We know that the probability of any specific combination of outcomes, for example, (1,1), (2,2), ... (6,6) is $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$. Hence, by the addition rule

$$P(\text{same face}) = P(1,1) + P(2,2) + \ldots + P(6,6) = 6 \times \frac{1}{36} = \frac{1}{6}. \tag{3.7}$$

Example 3.6. What is the probability that in two throws of a die at least one six appears?

Solution. We have already established that

$$P(6) = \frac{1}{6} \qquad P(\text{not } 6) = \frac{5}{6}. \tag{3.8}$$

There are four possible outcomes $(6, 6)$, $(6, \text{not } 6)$, $(\text{not } 6, 6)$, $(\text{not } 6, \text{not } 6)$ with the probabilities

$$P(6, 6) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \tag{3.9a}$$

$$P(6, \text{not } 6) = P(\text{not } 6, 6) = \frac{1}{6} \times \frac{5}{6} = \frac{5}{36} \tag{3.9b}$$

$$P(\text{not } 6, \text{not } 6) = \frac{5}{6} \times \frac{5}{6} = \frac{25}{36}. \tag{3.9c}$$

All outcomes except the last have at least one six. Hence, the probability of obtaining at least one six is

$$P(\text{at least one } 6) = P(6, 6) + P(6, \text{not } 6) + P(\text{not } 6, 6) \tag{3.10a}$$

$$= \frac{1}{36} + \frac{5}{36} + \frac{5}{36} = \frac{11}{36}. \tag{3.10b}$$

A more direct way of obtaining this result is to use the normalization condition. That is,

$$P(\text{at least one six}) = 1 - P(\text{not } 6, \text{not } 6) = 1 - \frac{25}{36} = \frac{11}{36}. \tag{3.10c}$$

Example 3.7. What is the probability of obtaining at least one six in four throws of a die?

Solution. We know that in one throw of a die, there are two outcomes with $P(6) = \frac{1}{6}$ and $P(\text{not } 6) = \frac{5}{6}$. Hence, in four throws of a die there are sixteen possible outcomes, only one of which has no six. That is, in the fifteen mutually exclusive outcomes, there is at least one six. We can use the multiplication rule (3.3) to find that

$$P(\text{not } 6, \text{not } 6, \text{not } 6, \text{not } 6) = P(\text{not } 6)^4 = \Big(\frac{5}{6}\Big)^4, \tag{3.11}$$

and hence

$$P(\text{at least one six}) = 1 - P(\text{not } 6, \text{not } 6, \text{not } 6, \text{not } 6) \tag{3.12a}$$

$$= 1 - \Big(\frac{5}{6}\Big)^4 = \frac{671}{1296} \approx 0.517. \tag{3.12b}$$

Frequently we know the probabilities only up to a constant factor. For example, we might know $P(1) = 2P(2)$, but not $P(1)$ or $P(2)$ separately. Suppose we know that $P(i)$ is proportional to $f(i)$, where $f(i)$ is a known function. To obtain the normalized probabilities, we divide each function $f(i)$ by the sum of all the unnormalized probabilities. That is, if $P(i) \propto f(i)$ and $Z = \sum f(i)$, then $P(i) = f(i)/Z$. This procedure is called *normalization*.

Example 3.8. Suppose that in a given class it is three times as likely to receive a C as an A, twice as likely to obtain a B as an A, one-fourth as likely to be assigned a D as an A, and nobody fails the class. What are the probabilities of getting each grade?

Solution. We first assign the unnormalized probability of receiving an A as $f(A) = 1$. Then $f(B) = 2$, $f(C) = 3$, and $f(D) = 0.25$. Then $Z = \sum_i f(i) = 1 + 2 + 3 + 0.25 = 6.25$. Hence, $P(A) = f(A)/Z = 1/6.25 = 0.16$, $P(B) = 2/6.25 = 0.32$, $P(C) = 3/6.25 = 0.48$, and $P(D) = 0.25/6.25 = 0.04$.

The normalization procedure arises again and again in different contexts. We will see that much of the mathematics of statistical mechanics can be formulated in terms of the calculation of normalization constants.

Problem 3.12. Find the probability distribution $P(n)$ for throwing a sum n with two dice and plot $P(n)$ as a function of n.

Problem 3.13. What is the probability of obtaining at least one double six in twenty-four throws of a pair of dice?

Problem 3.14. Suppose that three dice are thrown at the same time. What is the probability that the sum of the three faces is 10 compared to 9?

Problem 3.15. What is the probability that the total number of spots shown on three dice thrown at the same time is 11? What is the probability that the total is 12? What is the fallacy in the following argument? The number 11 occurs in six ways: (1,4,6), (2,3,6), (1,5,5), (2,4,5), (3,3,5), (3,4,4). The number 12 also occurs in six ways: (1,5,6), (2,4,6), (3,3,6), (2,5,5), (3,4,5), (4,4,4) and hence the two numbers should be equally probable.

3.3 Mean Values

The specification of the *probability distribution* $P(1)$, $P(2)$, $\ldots$ $P(n)$ for the n possible values of the variable x constitutes the most complete statistical description of the system. However, in many cases it is more convenient to describe the distribution of the possible values of x in a less detailed way. The most familiar way is to specify the *average* or *mean* value of x, which we will denote as $\overline{x}$. The definition of the mean value of x is

$$\overline{x} \equiv x_1 P(1) + x_2 P(2) + \ldots + x_n P(n) \tag{3.13a}$$

$$= \sum_{i=1}^{n} x_i P(i), \tag{3.13b}$$

where $P(i)$ is the probability of x_i. If $f(x)$ is a function of x, then the mean value of $f(x)$ is defined by

$$\overline{f(x)} = \sum_{i=1}^{n} f(x_i) P(i). \tag{3.14}$$

Example 3.9. Expected value

Lets reconsider the choices in Problem 3.8: A certain \$50 or \$100 if you flip a coin and get a head and \$0 if you get a tail. The *expected value* is

$$\text{expected value} = \sum_i P_i \times (\text{value of } i), \tag{3.15}$$

where the sum is over the expected outcomes and P_i is the probability of outcome i. In this case the expected value is $1/2 \times \$100 + 1/2 \times \$0 = \$50$. We see that the two choices are equivalent, and that the expected value is the same as the mean or average value. (Most people prefer the first choice because the outcome is "certain.")

If $f(x)$ and $g(x)$ are any two functions of x, then

$$\overline{f(x)+g(x)} = \sum_{i=1}^{n} [f(x_i)+g(x_i)]P(i) \tag{3.16a}$$

$$= \sum_{i=1}^{n} f(x_i)P(i) + \sum_{i=1}^{n} g(x_i)P(i), \tag{3.16b}$$

or

$$\overline{f(x)+g(x)} = \overline{f(x)} + \overline{g(x)}. \tag{3.16c}$$

Problem 3.16. Show that if c is a constant, then

$$\overline{cf(x)} = c\overline{f(x)}. \tag{3.17}$$

In general, we can define the mth *moment* of the probability distribution P as

$$\overline{x^m} \equiv \sum_{i=1}^{n} {x_i}^m P(i), \tag{3.18}$$

where we have let $f(x) = x^m$. The mean of x is the first moment of the probability distribution.

Problem 3.17. Suppose that the variable x takes on the values -2, -1, 0, 1, and 2 with probabilities 1/16, 4/16, 6/16, 4/16, and 1/16, respectively. Calculate the first two moments of x.

The mean value of x is a measure of the central value of x about which the various values of x_i are distributed. If we measure x from its mean, we have that

$$\Delta x \equiv x - \overline{x}, \tag{3.19}$$

and

$$\overline{\Delta x} = \overline{(x-\overline{x})} = \overline{x} - \overline{x} = 0. \tag{3.20}$$

That is, the average value of the deviation of x from its mean vanishes.

If only one outcome j were possible, we would have $P(i) = 1$ for $i = j$ and zero otherwise, that is, the probability distribution would have zero width. In general, there is more than one outcome and a possible measure of the width of the probability distribution is given by

$$\overline{\Delta x^2} \equiv \overline{\big(x-\overline{x}\big)^2}. \tag{3.21}$$

The quantity $\overline{\Delta x^2}$ is known as the *dispersion* or *variance* and its square root is called the *standard deviation*. It is easy to see that the larger the spread of values of x about $\overline{x}$, the larger the variance. The use of the square of $x - \overline{x}$ ensures that the contribution of x values that are smaller and larger than $\overline{x}$ enter with the same sign. A useful form for the variance can be found by letting

$$\overline{\big(x-\overline{x}\big)^2} = \overline{\big(x^2 - 2x\overline{x} + \overline{x}^2\big)} \tag{3.22a}$$

$$= \overline{x^2} - 2\overline{x}\;\overline{x} + \overline{x}^2, \tag{3.22b}$$

or

$$\overline{(x - \overline{x})^2} = \overline{x^2} - \overline{x}^2. \tag{3.23}$$

Because $\overline{\Delta x^2}$ is always nonnegative, it follows that $\overline{x^2} \geq \overline{x}^2$.

The variance is the mean value of $(x - \overline{x})^2$ and represents the square of a width. We will find that it is useful to interpret the width of the probability distribution in terms of the standard deviation σ, which is defined as the square root of the variance. The standard deviation of the probability distribution $P(x)$ is given by

$$\sigma_x = \sqrt{\overline{\Delta x^2}} = \sqrt{\big(\overline{x^2} - \overline{x}^2\big)}. \tag{3.24}$$

Example 3.10. Find the mean value $\overline{x}$, the variance $\overline{\Delta x^2}$, and the standard deviation σ_x for the value of a single throw of a die.

Solution. Because $P(i) = \frac{1}{6}$ for $i = 1, \ldots, 6$, we have that

$$\overline{x} = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2} = 3.5 \tag{3.25a}$$

$$\overline{x^2} = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{46}{3} \tag{3.25b}$$

$$\overline{\Delta x^2} = \overline{x^2} - \overline{x}^2 = \frac{46}{3} - \frac{49}{4} = \frac{37}{12} \approx 3.08 \tag{3.25c}$$

$$\sigma_x \approx \sqrt{3.08} = 1.76 \tag{3.25d}$$

Example 3.11. On the average, how many times must a die be thrown until a 6 appears?

Solution. Although it might seem obvious that the answer is six, it is instructive to confirm this answer. Let p be the probability of a six on a given throw. The probability of success for the first time on trial i is given in Table 3.2.

trial	probability of success on trial i
1	p
2	qp
3	q^2p
4	q^3p

Table 3.2: Probability of a head for the first time on trial i ($q = 1 - p$).

The sum of the probabilities is $p + qp + q^2p + \cdots = p(1 + q + q^2 + \cdots) = p/(1 - q) = p/p = 1$. The mean number of trials m is

$$m = p + 2pq + 3pq^2 + 4pq^3 + \cdots \tag{3.26a}$$

$$= p(1 + 2q + 3q^2 + \cdots) \tag{3.26b}$$

$$= p\frac{d}{dq}\big(1 + q + q^2 + q^3 + \cdots\big) \tag{3.26c}$$

$$= p\frac{d}{dq}\frac{1}{1 - q} = \frac{p}{(1 - q)^2} = \frac{1}{p} \tag{3.26d}$$

Another way to obtain this result is to note that if the first toss is a failure, then the mean number of tosses required is $1 + m$, and if the first toss is a success, the mean number is 1. Hence, $m = q(1 + m) + p(1)$ or $m = 1/p$.

3.4 The Meaning of Probability

How can we assign the probabilities of the various events? If we say that event E_1 is more probable than event E_2 ($P(E_1) > P(E_2)$), we mean that E_1 is more likely to occur than E_2. This statement of our intuitive understanding of probability illustrates that probability is a way of classifying the plausibility of events under conditions of uncertainty. Probability is related to our degree of belief in the occurrence of an event.

This definition of probability is not bound to a single evaluation rule and there are many ways to obtain $P(E_i)$. For example, we could use symmetry considerations as we have done, past frequencies, simulations, theoretical calculations, or as we will learn in Section 3.4.2, Bayesian inference. Probability assessments depend on who does the evaluation and the status of the information the evaluator has at the moment of the assessment. We always evaluate the conditional probability, that is, the probability of an event E given the information I, $P(E|I)$. Consequently, several people can have simultaneously different degrees of belief about the same event, as is well known to investors in the stock market.

If rational people have access to the same information, they should come to the same conclusion about the probability of an event. The idea of a *coherent bet* forces us to make probability assessments that correspond to our belief in the occurrence of an event. If we consider an event to be 50% probable, then we should be ready to place an even bet on the occurrence of the event or on its opposite. However, if someone wishes to place the bet in one direction but not in the other, it means that this person thinks that the preferred event is more probable than the other. In this case the 50% probability assessment is *incoherent* and this person's wish does not correspond to his or her belief.

A coherent bet has to be considered *virtual*. For example, a person might judge an event to be 99.9999% probable, but nevertheless refuse to bet \$999999 against \$1, if \$999999 is much more than the person's resources. Nevertheless, the person might be convinced that this bet would be fair if he/she had an infinite budget. Probability assessments should be kept separate from decision issues. Decisions depend not only on the probability of the event, but also on the subjective importance of a given amount of money (see for example, Problems 3.10 and 3.92).

Our discussion of probability as the degree of belief that an event will occur shows the inadequacy of the frequency definition of probability, which *defines* probability as the ratio of the number of desired outcomes to the total number of possible outcomes. This definition is inadequate because we would have to specify that each outcome has equal probability. Thus we would have to use the term probability in its own definition. If we do an experiment to measure the frequencies of various outcomes, then we need to make an additional assumption that the measured frequencies will be the same in the future as they were in the past. Also we have to make a large number of measurements to insure accuracy, and we have no way of knowing a priori how many measurements are sufficient. Thus, the definition of probability as a frequency really turns out to be a method for estimating probabilities with some hidden assumptions.

Our definition of probability as a measure of the degree of belief in the occurrence of an outcome implies that probability depends on our prior knowledge, because belief depends on prior knowledge. For example, if we toss a coin and obtain 100 tails in a row, we might use this knowledge as evidence that the coin or toss is biased, and thus estimate that the probability of throwing another tail is very high. However, if a careful physical analysis shows that there is no bias, then we would stick to our estimate of 1/2. The probability depends on what knowledge we bring to the problem. If we have no knowledge other than the possible outcomes, then the best estimate is to assume equal probability for all events. However, this assumption is not a definition, but an example of belief. As an example of the importance of prior knowledge, consider the following problem.

Problem 3.18. A couple with two children

(a) A couple has two children. What is the probability that at least one child is a girl?

(b) Suppose that you know that at least one child is a girl. What is the probability that the other child is a girl?

(c) Instead suppose that we know that the oldest child is a girl. What is the probability that the youngest is a girl?

We know that we can estimate probabilities empirically by sampling, that is, by making repeated measurements of the outcome of independent events. Intuitively we believe that if we perform more and more measurements, the calculated average will approach the exact mean of the quantity of interest. This idea is called *the law of large numbers.*

As an example, suppose that we flip a single coin M times and count the number of heads. Our result for the number of heads is shown in Table 3.3. We see that the fraction of heads approaches 1/2 as the number of measurements becomes larger.

Problem 3.19. Use the applet/application at <stp.clarku.edu/simulations/cointoss> to simulate multiple tosses of a single coin. What is the correspondence between this simulation of a coin being tossed many times and the actual physical tossing of a coin? If the coin is "fair," what do you think the ratio of the number of heads to the total number of tosses will be? Do you obtain this number after 100 tosses? 10,000 tosses?

Another way of estimating the probability is to perform a single measurement on many copies or replicas of the system of interest. For example, instead of flipping a single coin 100 times in succession, we collect 100 coins and flip all of them at the same time. The fraction of coins that show heads is an estimate of the probability of that event. The collection of identically prepared systems is called an *ensemble* and the probability of occurrence of a single event is estimated with respect to this ensemble. The ensemble consists of a large number M of identical systems, that is, systems that satisfy the same known conditions.

If the system of interest is not changing in time, it is reasonable to assume that an estimate of the probability by either a series of measurements on a single system at different times or similar measurements on many identical systems at the same time would give consistent results.

Note that we have *estimated* various probabilities by a frequency, but have not *defined* probability in terms of a frequency. As emphasized by D'Agostini, past frequency is experimental data.

heads	tosses	fraction of heads
4	10	0.4
29	50	0.58
49	100	0.49
101	200	0.505
235	500	0.470
518	1,000	0.518
4997	10,000	0.4997
50021	100,000	0.50021
249946	500,000	0.49999
500416	1,000,000	0.50042

Table 3.3: The number and fraction of heads in M tosses of a coin. We did not really toss a coin in the air 10^6 times. Instead we used a computer to generate a sequence of random numbers to simulate the tossing of a coin. Because you might not be familiar with such sequences, imagine a robot that can write the positive integers between 1 and 2^{31} on pieces of paper. Place these pieces in a hat, shake the hat, and then chose the pieces at random. If the number chosen is less than $\frac{1}{2} \times 2^{31}$, then we say that we found a head. Each piece is placed back in the hat after it is read.

This data happened with certainty so the concept of probability no longer applies. Probability is how much we believe that an event will occur taking into account all available information including past frequencies. Because probability quantifies the degree of belief at a given time, it is not measurable. If we make further measurements, they can only influence future assessments of the probability.

3.4.1 Information and uncertainty

Consider an experiment that has two outcomes E_1 and E_2 with probabilities P_1 and P_2. For example, the experiment could correspond to the toss of a coin. For one coin the probabilities are $P_1 = P_2 = 1/2$ and for the other (a bent coin) $P_1 = 1/5$ and $P_2 = 4/5$. Intuitively, we would say that the result of the first experiment is more uncertain.

Consider two additional experiments. In the third experiment there are four outcomes with $P_1 = P_2 = P_3 = P_4 = 1/4$ and in the fourth experiment there are six outcomes with $P_1 = P_2 = P_3 = P_4 = P_5 = P_6 = 1/6$. Intuitively the fourth experiment is the most uncertain because there are more outcomes and the first experiment is the least uncertain. You are probably not clear about how to rank the second and third experiments.

We will now introduce a mathematical measure that is consistent with our intuitive sense of uncertainty. Let us define the uncertainty function $S(P_1, P_2, \ldots, P_i, \ldots)$ where P_i is the probability of event i. We first consider the case where all the probabilities P_i are equal. Then $P_1 = P_2 = \ldots = P_i = 1/\Omega$, where Ω is the total number of outcomes. In this case we have $S = S(1/\Omega, 1/\Omega, \ldots)$ or simply $S(\Omega)$.

It is easy to see that $S(\Omega)$ has to satisfy some simple conditions. For only one outcome, $\Omega = 1$

and there is no uncertainty. Hence we must have

$$S(\Omega = 1) = 0. \tag{3.27}$$

We also have that

$$S(\Omega_1) > S(\Omega_2) \text{ if } \Omega_1 > \Omega_2. \tag{3.28}$$

That is, $S(\Omega)$ is a increasing function of Ω.

We next consider multiple events. For example, suppose that we throw a die with Ω_1 outcomes and flip a coin with Ω_2 equally probable outcomes. The total number of outcomes is $\Omega = \Omega_1\Omega_2$. If the result of the die is known, the uncertainty associated with the die is reduced to zero, but there still is uncertainty associated with the toss of the coin. Similarly, we can reduce the uncertainty in the reverse order, but the total uncertainty is still nonzero. These considerations suggest that

$$S(\Omega_1\Omega_2) = S(\Omega_1) + S(\Omega_2). \tag{3.29}$$

It is remarkable that there is an unique functional form that satisfies the three conditions (3.27)–(3.29). We can find this form by writing (3.29) in the form

$$S(xy) = S(x) + S(y), \tag{3.30}$$

and taking the variables x and y to be continuous. (The analysis can be done assuming that x and y are discrete variables, but the analysis is simpler if we assume that x and y are continuous. Given this assumption the functional form of S might already be clear.) This generalization is consistent with $S(\Omega)$ being a increasing function of Ω. First we take the partial derivative of $S(xy)$ with respect to x and then with respect to y. We let $z = xy$ and obtain

$$\frac{\partial S(z)}{\partial x} = \frac{\partial z}{\partial x}\frac{dS(z)}{dz} = y\frac{dS(z)}{dz} \tag{3.31a}$$

$$\frac{\partial S(z)}{\partial y} = \frac{\partial z}{\partial y}\frac{dS(z)}{dz} = x\frac{dS(z)}{dz}. \tag{3.31b}$$

From (3.30) we have

$$\frac{\partial S(z)}{\partial x} = \frac{dS(x)}{dx} \tag{3.32a}$$

$$\frac{\partial S(z)}{\partial y} = \frac{dS(y)}{dy}. \tag{3.32b}$$

By comparing the right-hand side of (3.31) and (3.32), we have

$$\frac{dS}{dx} = y\frac{dS}{dz} \tag{3.33a}$$

$$\frac{dS}{dy} = x\frac{dS}{dz}. \tag{3.33b}$$

If we multiply (3.33a) by x and (3.33b) by y, we obtain

$$x\frac{dS(x)}{dx} = y\frac{dS(y)}{dy} = z\frac{dS(z)}{dz}. \tag{3.34}$$

Note that the first term in (3.34) depends only on x and the second term depends only on y. Because x and y are independent variables, the three terms in (3.34) must be equal to a constant. Hence we have the desired condition

$$x\frac{dS(x)}{dx} = y\frac{dS(y)}{dy} = A, \tag{3.35}$$

where A is a constant. The differential equation in (3.35) can be integrated to give

$$S(x) = A\ln x + B. \tag{3.36}$$

The integration constant B must be equal to zero to satisfy the condition (3.27). The constant A is arbitrary so we choose $A = 1$. Hence for equal probabilities we have that

$$S(\Omega) = \ln\Omega. \tag{3.37}$$

What about the case where the probabilities for the various events are unequal? We will show in Section 3.12.1 that the general form of the uncertainty S is

$$S = -\sum_i P_i \ln P_i. \tag{3.38}$$

Note that if all the probabilities are equal, then

$$P_i = \frac{1}{\Omega} \tag{3.39}$$

for all i. In this case

$$S = -\sum_i \frac{1}{\Omega}\ln\frac{1}{\Omega} = \Omega\frac{1}{\Omega}\ln\Omega = \ln\Omega, \tag{3.40}$$

because there are Ω equal terms in the sum. Hence (3.38) reduces to (3.37) as required. We also see that if outcome j is certain, $P_j = 1$ and $P_i = 0$ if $i \neq j$ and $S = -1\ln 1 = 0$. That is, if the outcome is certain, the uncertainty is zero and there is no missing information.

We have shown that if the P_i are known, then the uncertainty or missing information S can be calculated. Usually the problem is the other way around, and we want to determine the probabilities. Suppose we flip a perfect coin for which there are two possibilities. We know intuitively that $P_1(\text{heads}) = P_2(\text{tails}) = 1/2$. That is, we would not assign a different probability to each outcome unless we had information to justify it. Intuitively we have adopted the principle of *least bias* or *maximum uncertainty*. Lets reconsider the toss of a coin. In this case S is given by

$$S = -\sum_i P_i \ln P_i = -(P_1\ln P_1 + P_2\ln P_2) \tag{3.41a}$$

$$= -(P_1\ln P_1 + (1-P_1)\ln(1-P_1), \tag{3.41b}$$

where we have used the fact that $P_1 + P_2 = 1$. To maximize S we take the derivative with respect to P_1:[4]

$$\frac{dS}{dP_1} = -[\ln P_1 + 1 - \ln(1-P_1) - 1] = -\ln\frac{P_1}{1-P_1} = 0. \tag{3.42}$$

[4]We have used the fact that $d(\ln x)/dx = 1/x$.

The solution of (3.42) satisfies

$$\frac{P_1}{1-P_1} = 1, \tag{3.43}$$

which is satisfied by $P_1 = 1/2$. We can check that this solution is a maximum by calculating the second derivative.

$$\frac{\partial^2 S}{\partial P_1^2} = -\Big[\frac{1}{P_1} + \frac{1}{1-P_1}\Big] = -4 < 0, \tag{3.44}$$

which is less then zero.

Problem 3.20. Uncertainty

(a) Consider the toss of a coin for which $P_1 = P_2 = 1/2$ for the two outcomes. What is the uncertainty in this case?

(b) What is the uncertainty for $P_1 = 1/3$ and $P_2 = 2/3$? How does the uncertainty in this case compare to that in part (a)?

(c) On page 111 we discussed four experiments with various outcomes. Compare the uncertainty S of the third and fourth experiments.

Example 3.12. The toss of a three-sided die yields events E_1, E_2, and E_3 with a face of one, two, and three points. As a result of tossing many dice, we learn that the mean number of points is $f = 1.9$, but we do not know the individual probabilities. What are the values of P_1, P_2, and P_3 that maximize the uncertainty?

Solution. We have

$$S = -\big[P_1 \ln P_1 + P_2 \ln P_2 + P_3 \ln P_3\big]. \tag{3.45}$$

We also know that

$$f = 1P_1 + 2P_2 + 3P_3, \tag{3.46}$$

and $P_1 + P_2 + P_3 = 1$. We use the latter condition to eliminate P_3 using $P_3 = 1 - P_1 - P_2$, and rewrite (3.46) as

$$f = P_1 + 2P_2 + 3(1 - P_1 - P_2) = 3 - 2P_1 - P_2. \tag{3.47}$$

We then use (3.47) to eliminate P_2 and P_3 from (3.45) using $P_2 = 3 - f - 2P_1$ and $P_3 = f - 2 + P_1$:

$$S = -[P_1 \ln P_1 + (3 - f - 2P_1)\ln(3 - f - 2P_1) + (f - 2 + P_1)\ln(f - 2 + P_1)]. \tag{3.48}$$

Because S in (3.48) depends on only P_1, we can differentiate S with respect P_1 to find its maximum value:

$$\frac{dS}{dP_1} = -\Big[\ln P_1 - 1 - 2[\ln(3 - f - 2P_1) - 1] + [\ln(f - 2 + P_1) - 1]\Big] \tag{3.49a}$$

$$= \ln\frac{P_1(f - 2 + P_1)}{(3 - f - 2P_1)^2} = 0. \tag{3.49b}$$

We see that for dS/dP_1 to be equal to zero, the argument of the logarithm must be one. The result is a quadratic equation for P_1 (see Problem 3.21).

Problem 3.21. Fill in the missing steps in Example 3.12 and solve for P_1, P_2, and P_3.

In Section 3.12.2 we maximize the uncertainty for a case for which there are more than three outcomes.

3.4.2 *Bayesian inference

Conditional probabilities are not especially important for the development of equilibrium statistical mechanics, so this section may be omitted for now. However, conditional probability and Bayes' theorem are very important for the analysis of data including spam filters for email, and in the more general context of statistical physics. Bayes' theorem gives us a way of understanding how the probability that a hypothesis is true is affected by new evidence.

Let us define $P(\mathrm{A}|\mathrm{B})$ as the probability of A occurring given that we know that B has occurred. We know that

$$P(\mathrm{A}) = P(\mathrm{A}|\mathrm{B}) + P(\mathrm{A}|{-\mathrm{B}}), \tag{3.50}$$

where $-\mathrm{B}$ means that B did not occur. We also know that

$$P(\mathrm{A\ and\ B}) = P(\mathrm{A}|\mathrm{B})P(\mathrm{B}) = P(\mathrm{B}|\mathrm{A})P(\mathrm{A}). \tag{3.51}$$

Equation (3.51) means that the probability that A and B occur equals the probability that A occurs given B times the probability that B occurs, which is the same as the probability that B occurs given A times the probability A that occurs. Note that $P(\mathrm{A\ and\ B})$ is the same as $P(\mathrm{B\ and\ A})$, but $P(\mathrm{A}|\mathrm{B})$ does not have the same meaning as $P(\mathrm{B}|\mathrm{A})$.

We can rearrange (3.51) to obtain Bayes' theorem

$$\boxed{P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{B}|\mathrm{A})P(\mathrm{A})}{P(\mathrm{B})}.} \qquad \text{(Bayes' theorem)} \tag{3.52}$$

We can generalize (3.52) for the case of multiple possible outcomes A_i for the same B. We rewrite (3.52) as

$$P(\mathrm{A}_i|\mathrm{B}) = \frac{P(\mathrm{B}|\mathrm{A}_i)P(\mathrm{A}_i)}{P(\mathrm{B})}. \tag{3.53}$$

If all the A_i are mutually exclusive and if at least one of the A_i must occur, then we can also write

$$P(\mathrm{B}) = \sum_i P(\mathrm{B}|\mathrm{A}_i)P(\mathrm{A}_i). \tag{3.54}$$

If we substitute (3.54) for $P(\mathrm{B})$ into (3.53), we obtain

$$P(\mathrm{A}_i|\mathrm{B}) = \frac{P(\mathrm{B}|\mathrm{A}_i)P(\mathrm{A}_i)}{\sum_i P(\mathrm{B}|\mathrm{A}_i)P(\mathrm{A}_i)}. \tag{3.55}$$

Bayes' theorem is very useful for finding the most probable explanation of a given data set. In this context A_i represents the possible explanation and B represents the data. As more data becomes available, the probabilities $P(\mathrm{B}|\mathrm{A}_i)P(\mathrm{A}_i)$ change.

Example 3.13. A chess program has two modes, expert (E) and novice (N). The expert mode beats you 75% of the time and the novice mode wins 50% of the time. You close your eyes and randomly choose one of the modes and play two games. The computer wins (W) both times. What is the probability that you chose the novice mode?

Solution. The probability of interest is $P(\text{N}|\text{WW})$, which is difficult to calculate directly. Bayes theorem lets you use the easy to calculate probability $P(\text{WW}|\text{N})$ to determine $P(\text{N}|\text{WW})$. We use (3.52) to write

$$P(\text{N}|\text{WW}) = \frac{P(\text{WW}|\text{N})P(\text{N})}{P(\text{WW})}. \tag{3.56}$$

We know that $P(\text{N}) = 1/2$ and $P(\text{WW}|\text{N}) = (1/2)^2 = 1/4$.

We next have to calculate $P(\text{WW})$. There are two ways that the program won the two games: (1) You chose the novice mode and it won twice, or (2) you chose the expert mode and it won twice. Because N and E are mutually exculsive, we have $P(\text{WW}) = P(\text{N and WW}) + P(\text{E and WW})$. From (3.51) we have

$$P(\text{WW}) = P(\text{WW}|\text{N})P(\text{N}) + P(\text{WW}|\text{E})P(\text{E}) \tag{3.57a}$$

$$= (1/2 \times 1/2 \times 1/2) + (3/4 \times 3/4 \times 1/2) = \frac{13}{32}. \tag{3.57b}$$

Hence

$$P(\text{N}|\text{WW}) = \frac{P(\text{WW}|\text{N})P(\text{N})}{P(\text{WW})} = \frac{(1/4 \times 1/2)}{\frac{13}{32}} = \frac{4}{13} \approx 0.31. \tag{3.58}$$

Note that the probability of choosing the novice mode has decreased from 50% to about 31% because you have the additional information that you lost twice and thus are more likely to have chosen the expert mode.

Example 3.14. Alice plants two types of flowers in her garden: 30% of type A and 70% of type B. Both types yield either red or yellow flowers, with $P(\text{red}|\text{A}) = 0.4$ and $P(\text{red}|\text{B}) = 0.3$.

(a) What is the percentage of red flowers that Alice will obtain?

Solution. We can use the total probability law (3.51) to write

$$P(\text{red}) = P(\text{red}|\text{A})P(\text{A}) + P(\text{red}|B)P(B) \tag{3.59a}$$

$$= (0.4 \times 0.3) + (0.3 \times 0.7) = 33/100. \tag{3.59b}$$

So Alice will find that one of three flowers will be red.

(b) Suppose a red flower is picked at random from Alice's garden. What is the probability of the flower being type A?

Solution. We apply Bayes' theorem and obtain

$$P(\text{A}|\text{red}) = \frac{P(\text{red}|\text{A})P(\text{A})}{P(\text{red}|\text{A})P(\text{A}) + P(\text{red}|\text{B})P(\text{B})} \tag{3.60a}$$

$$= \frac{0.4 \times 0.3}{(0.4 \times 0.3) + (0.3 \times 0.7)} = \frac{12}{33} = \frac{4}{11} \approx 0.36. \tag{3.60b}$$

We find that given that the flower is red, its probability of being type A increases to 0.36 because type A has a higher probability than type B of yielding red flowers.

Example 3.15. Do you have a fair coin?

Suppose that there are four coins of the same type in a bag. Three of them are fair, but the fourth is double-headed. You choose one coin at random from the bag and toss it five times. It comes up heads each time. What is the probability that you have chosen the double-headed coin?

Solution. If the coin were fair, the probability of five heads in a row (5H) would be $(1/2)^5 = 1/32 \approx 0.03$. This probability is small, so you would probably decide that you have not chosen a fair coin. But because you have more information, you can determine a better estimate of the probability.

We have

$$\begin{aligned} P(\text{H5}) &= P(\text{5H}|\text{fair})P(\text{fair}) + P(\text{5H}|\text{not fair})P(\text{not fair}) && (3.61\text{a})\\ &= [(1/2)^5 \times 3/4] + [1 \times 1/4] = 35/128 \approx 0.27. && (3.61\text{b})\\ P(\text{fair coin}|\text{5H}) &= P(\text{5H}|\text{fair coin})P(\text{fair coin})/P(5H) && (3.61\text{c})\\ &= \frac{[(1/2)^5 \times 3/4]}{35/128} = 3/35 = 0.12. && (3.61\text{d}) \end{aligned}$$

Thus the probability that the coin was fair is about a factor of four greater given that you tossed a coin five times.

Problem 3.22. More on choosing a fair coin

Suppose that you have two coins that look and feel identical, but one is double-headed and one is fair. The two coins are placed in a box and you choose one at random.

(a) What is the probability that you have chosen the fair coin?

(b) Suppose that you toss the chosen coin twice and obtain heads both times. What is the probability that you have chosen the fair coin? Why is this probability different than in Part 3.22a?

(c) Suppose that you toss the chosen coin four times and obtain four heads. What is the probability that you have chosen the fair coin?

(d) Suppose that there are ten coins in the box with nine fair and one double-headed. You toss the chosen twice and obtain two heads. What is the probability that you have chosen the fair coin?

(e) Now suppose that the biased coin is not double-headed, but has a probability of 0.98 of coming up heads. Also suppose that the probability of choosing the biased coin is 1 in 10^4. What is the probability of choosing the biased coin given that the first toss yields heads?

Example 3.16. Monty Hall problem

Consider the quandary known as the Monty Hall problem. In this former television show a contestant is shown three doors. Behind one door is an expensive prize such as a car and behind the other two doors are inexpensive gifts such as a tie. The contestant chooses a door. Suppose she chooses door 1. Then the host opens door 2 containing the tie knowing that the car is not behind door 2. The contestant now has a choice – should she stay with her original choice or switch to door 3? What would you do?[5]

Let us use Bayes' theorem to determine her best course of action. We want to calculate

$$P(A_1|B) = P(\text{car behind door 1}|\text{door 2 open after door 1 chosen}), \tag{3.62a}$$

and

$$P(A_3|B) = P(\text{car behind door 3}|\text{door 2 open after door 1 chosen}), \tag{3.62b}$$

where A_i denotes car behind door i. We know that all the $P(A_i)$ equal 1/3, because with no information we assume that the probability that the car is behind each door is the same. Because the host can open door 2 or 3 if the car is behind door 1, but can only open door 2 if the car is behind door 3 we have

$$P(\text{door 2 open after door 1 chosen}|\text{car behind 1}) = \frac{1}{2} \tag{3.63a}$$

$$P(\text{door 2 open after door 1 chosen}|\text{car behind 2}) = 0 \tag{3.63b}$$

$$P(\text{door 2 open after door 1 chosen}|\text{car behind 3}) = 1. \tag{3.63c}$$

From Bayes' theorem we have

$$P(\text{car behind 1}|\text{door 2 open after door 1 chosen}) = \frac{\frac{1}{2}\times\frac{1}{3}}{(\frac{1}{2}\times\frac{1}{3}) + (0\times\frac{1}{3}) + (1\times\frac{1}{3})} = \frac{1}{3} \tag{3.64a}$$

$$P(\text{car behind 3}|\text{door 2 open after door 1 chosen}) = \frac{1\times\frac{1}{3}}{(\frac{1}{2}\times\frac{1}{3}) + (0\times\frac{1}{3}) + (1\times\frac{1}{3})} = \frac{2}{3}. \tag{3.64b}$$

The results in (3.64) suggest the contestant has a higher probability of winning the car if she switches doors and chooses door 3. The same logic suggests that she should always switch doors independently of which door she originally chose.[6]

Problem 3.23. What does the host know?

The point of Bayesian statistics is that it approaches a given data set with a particular model in mind. In the Monte Hall problem the model we have used is that the host knows where the car is.

(a) Suppose that the host doesn't know where the car is, but chooses door 2 at random and there is no car. What is the probability that the car is behind door 1?

(b) Is the probability that you found in Part 3.23a the same as found in Example 3.16? Why or why not? Discuss why the probability that the car is behind door 1 depends on what the host knows.

Example 3.17. Bayes theorem and the problem of false positives

Even though you have no symptoms, your doctor wishes to test you for a rare disease that only 1 in 10,000 people of your age contract. The test is 98% accurate, which means that if you have the

[5] This question was posed on the TV game show, "Let's Make A Deal," hosted by Monty Hall.

[6] A search for Monty Hall will bring many sites, including <en.wikipedia.org/wiki/Monty_Hall_problem>, that discuss the problem in detail.

disease, 98% of the times the test will come out positive, and 2% negative. We also assume that if you do not have the disease, the test will come out negative 98% of the time and positive 2% of the time. You take the test and it comes out positive. What is the probability that you have the disease?

Solution. Let $P(p|D) = 0.98$ represent the probability of testing positive and having the disease, $-D$ represent the probability of not having the disease, and n represent testing negative. Then we are given that $P(n|D) = 0.02$, $P(n|-D) = 0.98$, $P(p|-D) = 0.02$, $P(D) = 0.0001$, and $P(-D) = 0.9999$. From Bayes' theorem we have

$$P(D|p) = \frac{P(p|D)P(D)}{P(p|D)P(D) + P(p|-D)P(-D)} \tag{3.65a}$$

$$= \frac{(0.98)(0.0001)}{(0.98)(0.0001) + (0.02)(0.9999)} \tag{3.65b}$$

$$= 0.0047 = 0.47\%. \tag{3.65c}$$

Is this test useful?

Because of the problem of false positives, some tests might actually reduce your life span and thus are not recommended. Suppose that a certain type of cancer occurs in 1 in 1000 people who are less than 50 years old. The death rate from this cancer is 25% in 10 years. The probability of having cancer if the test is positive is 1 in 20. Because people who test positive become worried, 90% of the patients who test positive have surgery to remove the cancer. As a result of surgery, 2% die due to complications, and the rest are cured.

We have that

$$P(\text{death rate due to cancer}) = P(\text{death}|\text{cancer})P(\text{cancer}) \tag{3.66a}$$

$$= 0.25 \times 0.001 = 0.00025 \tag{3.66b}$$

$$P(\text{death due to test}) = P(\text{die}|\text{surgery})P(\text{surgery}|\text{positive})P(\text{test}|\text{posiive}) \tag{3.66c}$$

$$= 0.02 \times 0.90 \times 0.02 = 0.00036. \tag{3.66d}$$

Hence, the probability of dring from surgery is greater than dying from the cancer.

Problem 3.24. Imagine that you have a sack of 3 balls that can be either red or green. There are four hypotheses for the distribution of colors for the balls: (1) all are red, (2) 2 are red, (3) 1 is red, and (4) all are green. Initially, you have no information about which hypothesis is correct, and thus you assume that they are equally probable. Suppose that you pick one ball out of the sack and it is green. Use Bayes' theorem to determine the new probabilities for each hypothesis.

Problem 3.25. Make a table that determines the accuracy necessary for a test to give the probability of having a disease if tested positive equal to at least 50% for diseases that occur in 1 in 100, 1 in 1000, 1 in 10,000, and 1 in 100,000 people.

We have emphasized that the definition of probability as a frequency is inadequate. If you are interesting in learning more about Bayesian inference, see in particular the paper by D'Agostini.

3.5 Bernoulli Processes and the Binomial Distribution

Because most physicists spend little time gambling,[7] we will have to develop our intuitive understanding of probability in other ways. Our strategy will be to first consider some physical systems for which we can calculate the probability distribution by analytical methods. Then we will use the computer to generate more data to analyze.

Noninteracting magnetic moments

Consider a system of N noninteracting magnetic moments of spin $\frac{1}{2}$, each having a magnetic moment μ in an external magnetic field B. The field B is in the up $(+z)$ direction. Spin $\frac{1}{2}$ implies that a spin can point either up (parallel to B) or down (antiparallel to B). The energy of interaction of each spin with the magnetic field is $E = \mp\mu B$, according to the orientation of the magnetic moment. As discussed in Section 1.10, this model is a simplification of more realistic magnetic systems.

We will take p to be the probability that the spin (magnetic moment) is up and q the probability that the spin is down. Because there are no other possible outcomes,we have $p+q = 1$ or $q = 1-p$. If $B = 0$, there is no preferred spatial direction and $p = q = 1/2$. For $B \neq 0$ we do not yet know how to calculate p and for now we will assume that p is a known parameter. In Section 4.8 we will learn how to calculate p and q when the system is in equilibrium at temperature T.

We associate with each spin a random variable s_i which has the values ± 1 with probability p and q, respectively. One of the quantities of interest is the magnetization M, which is the net magnetic moment of the system. For a system of N spins the magnetization is given by

$$M = \mu(s_1 + s_2 + \ldots + s_N) = \mu \sum_{i=1}^{N} s_i. \tag{3.67}$$

In the following, we will take $\mu = 1$ for convenience whenever it will not cause confusion. Alternatively, we can interpret M as the net number of up spins.

We will first calculate the mean value of M, then its variance, and finally the probability distribution $P(M)$ that the system has magnetization M. To compute the mean value of M, we need to take the mean values of both sides of (3.67). If we use (3.16c), we can interchange the sum and the average and write

$$\overline{M} = \overline{\Big(\sum_{i=1}^{N} s_i\Big)} = \sum_{i=1}^{N} \overline{s}_i. \tag{3.68}$$

Because the probability that any spin has the value ± 1 is the same for each spin, the mean value of each spin is the same, that is, $\overline{s}_1 = \overline{s}_2 = \ldots = \overline{s}_N \equiv \overline{s}$. Therefore the sum in (3.68) consists of N equal terms and can be written as

$$\overline{M} = N\overline{s}. \tag{3.69}$$

The meaning of (3.69) is that the mean magnetization is N times the mean magnetization of a single spin. Because $\overline{s} = (1 \times p) + (-1 \times q) = p - q$, we have that

$$\overline{M} = N(p - q). \tag{3.70}$$

[7]After a Las Vegas hotel hosted a meeting of the American Physical Society in March, 1986, the physicists were asked never to return.

Now let us calculate the variance of M, that is, $\overline{(M - \overline{M})^2}$. We write

$$\Delta M = M - \overline{M} = \sum_{i=1}^{N} \Delta s_i, \tag{3.71}$$

where

$$\Delta s_i \equiv s_i - \overline{s}. \tag{3.72}$$

As an example, let us calculate $\overline{(\Delta M)^2}$ for $N = 3$ spins. In this case $(\Delta M)^2$ is given by

$$(\Delta M)^2 = (\Delta s_1 + \Delta s_2 + \Delta s_3)(\Delta s_1 + \Delta s_2 + \Delta s_3) \tag{3.73a}$$

$$= \big[(\Delta s_1)^2 + (\Delta s_2)^2 + (\Delta s_3)^2\big] + 2\big[\Delta s_1 \Delta s_2 + \Delta s_1 \Delta s_3 + \Delta s_2 \Delta s_3\big]. \tag{3.73b}$$

We take the mean value of (3.73b), interchange the order of the sums and averages, and write

$$\overline{(\Delta M)^2} = \big[\overline{(\Delta s_1)^2} + \overline{(\Delta s_2)^2} + \overline{(\Delta s_3)^2}\big] + 2\big[\overline{\Delta s_1 \Delta s_2} + \overline{\Delta s_1 \Delta s_3} + \overline{\Delta s_2 \Delta s_3}\big]. \tag{3.74}$$

The first term on the right of (3.74) represents the three terms in the sum that are multiplied by themselves. The second term represents all the cross terms arising from different terms in the sum, that is, the products in the second sum refer to different spins. Because different spins are statistically independent (the spins do not interact), we have that

$$\overline{\Delta s_i \, \Delta s_j} = \overline{\Delta s_i} \; \overline{\Delta s_j} = 0, \qquad (i \neq j) \tag{3.75}$$

because $\overline{\Delta s_i} = 0$. That is, each cross term vanishes on the average. Hence (3.75) reduces to a sum of squared terms

$$\overline{(\Delta M)^2} = \big[\overline{(\Delta s_1)^2} + \overline{(\Delta s_2)^2} + \overline{(\Delta s_3)^2}\big]. \tag{3.76}$$

Because each spin is equivalent on the average, each term in (3.76) is equal. Hence, we obtain the desired result

$$\overline{(\Delta M)^2} = 3\overline{(\Delta s)^2}. \tag{3.77}$$

The variance of M is 3 times the variance of a single spin, that is, the variance is additive.

We can evaluate $\overline{(\Delta M)^2}$ further by finding an explicit expression for $\overline{(\Delta s)^2}$. We have that $\overline{s^2} = [1^2 \times p] + [(-1)^2 \times q] = p + q = 1$. Hence, we have

$$\overline{(\Delta s)^2} = \overline{s^2} - \overline{s}^2 = 1 - (p - q)^2 = 1 - (2p - 1)^2 \tag{3.78a}$$

$$= 1 - 4p^2 + 4p - 1 = 4p(1 - p) = 4pq, \tag{3.78b}$$

and our desired result for $\overline{(\Delta M)^2}$ is

$$\overline{(\Delta M)^2} = 3(4pq). \tag{3.79}$$

Problem 3.26. Use similar considerations to show that for $N = 3$ that

$$\overline{n} = 3p \tag{3.80}$$

and

$$\overline{(n - \overline{n})^2} = 3pq, \tag{3.81}$$

where n is the number of up spins. Explain the difference between (3.70) and (3.80) for $N = 3$, and the difference between (3.79) and (3.81).

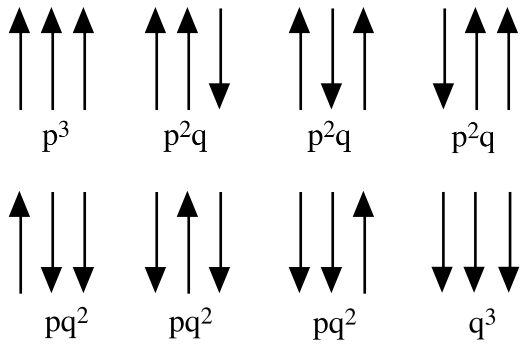

Figure 3.2: An ensemble of $N = 3$ spins. The arrow indicates the direction of the magnetic moment of a spin. The probability of each member of the ensemble is shown.

Problem 3.27. In the text we showed that $\overline{(\Delta M)^2} = 3\overline{(\Delta s)^2}$ for $N = 3$ spins (see (3.77) and (3.79)). Use similar considerations for N noninteracting spins to show that

$$\overline{(\Delta M)^2} = N(4pq). \tag{3.82}$$

Because of the simplicity of a system of noninteracting spins, we can calculate the probability distribution itself and not just the first few moments. As an example, let us consider the statistical properties of a system of $N = 3$ noninteracting spins. Because each spin can be in one of two states, there are $2^{N=3} = 8$ distinct outcomes (see Figure 3.2). Because each spin is independent of the other spins, we can use the multiplication rule (3.5) to calculate the probabilities of each outcome as shown in Figure 3.2. Although each outcome is distinct, several of the configurations have the same number of up spins. One quantity of interest is the probability $P_N(n)$ that n spins are up out a total of N spins. For example, there are three states with $n = 2$, each with probability p^2q so the probability that two spins are up is equal to $3p^2q$. For $N = 3$ we see from Figure 3.2 that

$$P_3(n = 3) = p^3 \tag{3.83a}$$
$$P_3(n = 2) = 3p^2q \tag{3.83b}$$
$$P_3(n = 1) = 3pq^2 \tag{3.83c}$$
$$P_3(n = 0) = q^3. \tag{3.83d}$$

Example 3.18. Find the first two moments of $P_3(n)$.

Solution. The first moment $\overline{n}$ of the distribution is given by

$$\overline{n} = 0 \times q^3 + 1 \times 3pq^2 + 2 \times 3p^2q + 3 \times p^3 \tag{3.84a}$$
$$= 3p\,(q^2 + 2pq + p^2) = 3p\,(q + p)^2 = 3p. \tag{3.84b}$$

Similarly, the second moment $\overline{n^2}$ of the distribution is given by

$$\overline{n^2} = 0 \times q^3 + 1 \times 3pq^2 + 4 \times 3p^2q + 9 \times p^3 \tag{3.84c}$$
$$= 3p\,(q^2 + 4pq + 3p^2) = 3p(q + 3p)(q + p) \tag{3.84d}$$
$$= 3p\,(q + 3p) = (3p)^2 + 3pq. \tag{3.84e}$$

Hence

$$\overline{(n - \overline{n})^2} = \overline{n^2} - \overline{n}^2 = 3pq. \tag{3.84f}$$

The mean magnetization M or the mean of the net number of up spins is given by the difference between the mean number of spins pointing up minus the mean number of spins pointing down: $\overline{M} = [\overline{n} - (3 - \overline{n}]$, or $\overline{M} = 3(2p - 1) = 3(p - q)$.

Problem 3.28. Coins and random walks

The outcome of N coins is identical to N noninteracting spins, if we associate the number of coins with N, the number of heads with n, and the number of tails with $N - n$. For a fair coin the probability p of a head is $p = \frac{1}{2}$ and the probability of a tail is $q = 1 - p = \frac{1}{2}$. What is the probability that in three tosses of a coin, there will be two heads?

Problem 3.29. One-dimensional random walk

The original statement of the *random walk* problem was posed by Pearson in 1905. If a drunkard begins at a lamp post and takes N steps of equal length in random directions, how far will the drunkard be from the lamp post? We will consider an idealized example of a random walk for which the steps of the walker are restricted to a line (a one-dimensional random walk). Each step is of equal length a, and at each interval of time, the walker either takes a step to the right with probability p or a step to the left with probability $q = 1 - p$. The direction of each step is independent of the preceding one. Let n be the number of steps to the right, and n' the number of steps to the left. The total number of steps $N = n + n'$. What is the probability that a random walker in one dimension has taken three steps to the right out of four steps?

From the above examples and problems, we see that the probability distributions of noninteracting magnetic moments, the flip of a coin, and a random walk are identical. These examples have two characteristics in common. First, in each trial there are only *two* outcomes, for example, up or down, heads or tails, and right or left. Second, the result of each trial is independent of all previous trials, for example, the drunken sailor has no memory of his or her previous steps. This type of process is called a *Bernoulli* process (after the mathematician Jacob Bernoulli, 1654–1705).

Because of the importance of magnetic systems, we will cast our discussion of Bernoulli processes in terms of the noninteracting magnetic moments of spin $\frac{1}{2}$. The main quantity of interest is the probability $P_N(n)$ which we now calculate for arbitrary N and n. We know that a particular outcome with n up spins and n' down spins occurs with probability $p^n q^{n'}$. We write the probability $P_N(n)$ as

$$P_N(n) = W_N(n, n')s\, p^n q^{n'}, \tag{3.85}$$

where $n' = N - n$ and $W_N(n, n')$ is the number of distinct configurations of N spins with n up spins and n' down spins. From our discussion of $N = 3$ noninteracting spins, we already know the first several values of $W_N(n, n')$.

We can determine the general form of $W_N(n, n')$ by obtaining a recursion relation between W_N and W_{N-1}. A total of n up spins and n' down spins out of N total spins can be found by adding one spin to $N - 1$ spins. The additional spin is either

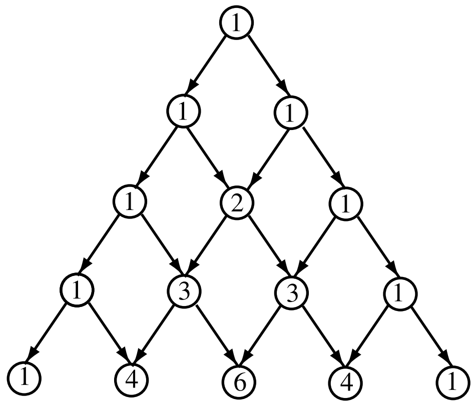

Figure 3.3: The values of the first few coefficients $W_N(n,n')$. Each number is the sum of the two numbers to the left and right above it. This construction is called a Pascal triangle.

(a) up if there are $(n-1)$ up spins and n' down spins, or

(b) down if there are n up spins and n' down spins.

Because there are $W_N(n-1,n')$ ways of reaching the first case and $W_N(n,n'-1)$ ways in the second case, we obtain the recursion relation

$$W_N(n,n') = W_{N-1}(n-1,n') + W_{N-1}(n,n'-1). \tag{3.86}$$

If we begin with the known values $W_0(0,0)=1$, $W_1(1,0)=W_1(0,1)=1$, we can use the recursion relation (3.86) to construct $W_N(n,n')$ for any desired N. For example,

$$W_2(2,0) = W_1(1,0) + W_1(2,-1) = 1+0 = 1. \tag{3.87a}$$
$$W_2(1,1) = W_1(0,1) + W_1(1,0) = 1+1 = 2. \tag{3.87b}$$
$$W_2(0,2) = W_1(-1,2) + W_1(0,1) = 0+1. \tag{3.87c}$$

In Figure 3.3 we show that $W_N(n,n')$ forms a pyramid or (a Pascal) triangle.

It is straightforward to show by induction that the expression

$$W_N(n,n') = \frac{N!}{n!\,n'!} = \frac{N!}{n!(N-n)!} \tag{3.88}$$

satisfies the relation (3.86). Note the convention $0! = 1$. We can combine (3.85) and (3.88) to find the desired result

$$\boxed{P_N(n) = \frac{N!}{n!\,(N-n)!}\,p^n q^{N-n}}. \qquad \text{(binomial distribution)} \tag{3.89}$$

The form (3.89) is called the *binomial distribution.* Note that for $p=q=1/2$, $P_N(n)$ reduces to

$$P_N(n) = \frac{N!}{n!\,(N-n)!}2^{-N}. \tag{3.90}$$

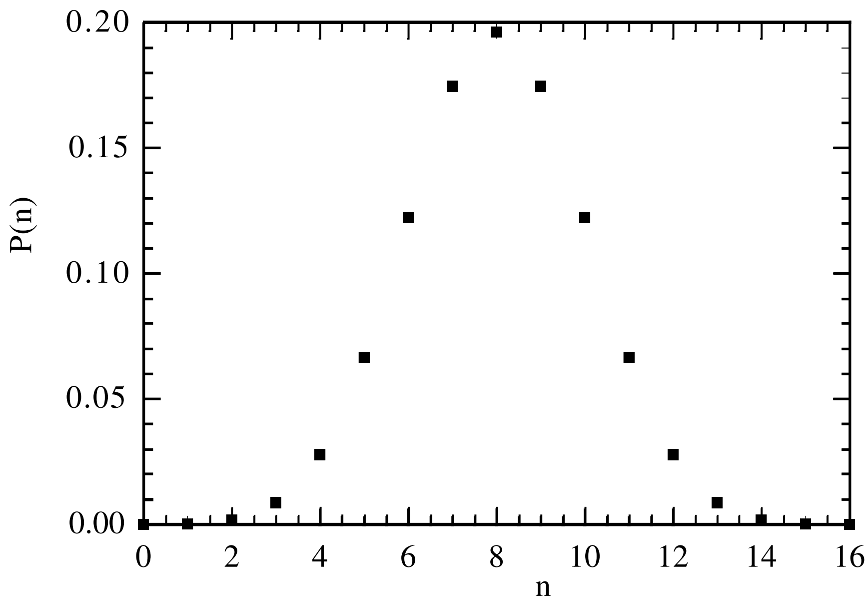

Figure 3.4: The binomial distribution $P_{16}(n)$ for $p = q = 1/2$ and $N = 16$. What is your visual estimate for the width of the distribution?

The probability $P_N(n)$ is shown in Figure 3.4 for $N = 16$.

Problem 3.30. Binomial distribution

(a) Calculate the distribution $P_N(n)$ that n spins are up out of a total of N for $N = 4$ and $N = 16$ and put your results in the form of a table. Calculate the mean values of n and n^2 using your tabulated values of $P_N(n)$. It is possible to do the calculation for general p and q, but choose $p = q = 1/2$ if necessary. Although it is better to first do the calculation of $P_N(n)$ by hand, you can use the applet/application at <stp.clarku.edu/simulations/binomial>.

(b) Plot your tabulated results for $P_N(n)$ (see Figure 3.4) or use the applet mentioned in part (a). Assume that $p = q = 1/2$. Visually estimate the width of the distribution for each value of N. Then use the applet/application for larger values of N. What is the qualitative dependence of the width on N? Also compare the relative heights of the maximum of P_N.

(c) Plot $P_N(n)$ as a function of $n/\overline{n}$ for $N = 4$ and $N = 16$ on the same graph as in part (b). Visually estimate the relative width of the distribution for each value of N.

(d) The applet/application plots $P_N(n)$ for various values of N in the same size window. Does the width of the distribution appear to become larger or smaller as N is increased?

(e) Plot $\ln P_N(n)$ versus $n/\overline{n}$ for $N = 16$. (Choose Log Axes under the Views menu.) Describe the behavior of $\ln P_N(n)$. Can $\ln P_N(n)$ be fitted to a parabola of the form $A + B(n - \overline{n})^2$, where A and B are fit parameters?

Problem 3.31. Asymmetrical distribution

(a) Plot $P_N(n)$ versus n for $N = 16$ and $p = 2/3$. For what value of n is $P_N(n)$ a maximum? How does the width of the distribution compare to what you found in Problem 3.30?

(b) For what value of p and q do you think the width is a maximum for a given N?

Example 3.19. Show that the expression (3.89) for $P_N(n)$ satisfies the normalization condition (3.2).

Solution. The reason that (3.89) is called the binomial distribution is that its form represents a typical term in the expansion of $(p+q)^N$. By the binomial theorem we have

$$(p+q)^N = \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!}\, p^n\, q^{N-n}. \tag{3.91}$$

We use (3.89) and write

$$\sum_{n=0}^{N} P_N(n) = \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n} = (p+q)^N = 1^N = 1, \tag{3.92}$$

where we have used (3.91) and the fact that $p + q = 1$.

Problem 3.32. Monte Carlo simulation of a one-dimensional random walk

The applet/application at <stp.clarku.edu/simulations/randomwalks/fixedlength1d.html> simulates a random walk in one dimension. As described in the text, a walker starts at the origin and takes N steps. At each step the walker goes to the right with probability p or to the left with probability $(1-p)$. Each step is the same length and independent of the previous steps. What is the displacement of the walker after N steps? Are some displacements more likely than others?

We can simulate a N-step walk by the following pseudocode:

```
do istep = 1,N
  if (rnd <= p) then
    x = x + 1
  else
    x = x - 1
  end if
end do
```

The function `rnd` generates a random number between zero and one. The quantity x is the net displacement assuming that the steps are of unit length.

We average over many walkers (trials), where each trial consists of a N step walk and construct a histogram for the number of times that the displacement x is found for a given number of walkers. The probability that the walker is a distance x from the origin after N steps is proportional to the corresponding value of the histogram. This procedure is called *Monte Carlo* sampling.[8]

[8] The name "Monte Carlo" was coined by Nicolas Metropolis in 1949.

(a) Is the value of x for one trial of any interest? Why do we have to average over many trials?

(b) Will we obtain the exact answer for the probability distribution?

(c) Choose $N = 4$ and $p = 1/2$. How does the histogram change, if at all, as the number of walk increases for fixed N?

(d) Describe the qualitative changes of the histogram for larger values of N and $p = 1/2$.

(e) What is the most probable value of x for $p = 1/2$ and $N = 16$ and $N = 32$? What is the approximate width of the distribution? Define the width visually. One way to do so is to determine the value of x at which the value of the histogram is one-half of its maximum value. How does the width change as a function of N for fixed p?

Calculation of the mean value

We now find an analytical expression for the dependence of $\overline{n}$ on N and p. From the definition (3.13) and (3.89) we have

$$\overline{n} = \sum_{n=0}^{N} n\, P_N(n) = \sum_{n=0}^{N} n\, \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n}. \tag{3.93}$$

We evaluate the sum in (3.93) by using a technique that is useful in a variety of contexts.[9] The technique is based on the fact that

$$p\,\frac{d}{dp} p^n = n p^n. \tag{3.94}$$

We use (3.94) to rewrite (3.93) as

$$\overline{n} = \sum_{n=0}^{N} n \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n} \tag{3.95a}$$

$$= \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!} \big(p \frac{\partial}{\partial p} p^n\big) q^{N-n}. \tag{3.95b}$$

We have used a partial derivative in (3.95b) to remind us that the derivative operator does not act on q. We interchange the order of summation and differentiation in (3.95b) and write

$$\overline{n} = p \frac{\partial}{\partial p} \Big[\sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n} \Big] \tag{3.96a}$$

$$= p \frac{\partial}{\partial p}\, (p+q)^N, \tag{3.96b}$$

where we have temporarily assumed that p and q are independent variables. Because the operator acts only on p, we have

$$\overline{n} = pN(p+q)^{N-1}. \tag{3.97}$$

[9]The integral $\int_0^\infty x^n\, e^{-ax^2}$ for $a > 0$ is evaluated in Appendix A using a similar technique.

The result (3.97) is valid for arbitrary p and q, and hence it is applicable for $p+q=1$. Thus our desired result is

$$\overline{n} = pN. \tag{3.98}$$

The dependence of $\overline{n}$ on N and p should be intuitively clear. Compare the general result (3.98) to the result (3.84b) for $N=3$. What is the dependence of $\overline{n'}$ on N and p?

Calculation of the relative fluctuations

To determine $\overline{\Delta n^2}$ we need to know $\overline{n^2}$ (see the relation (3.23)). The average value of n^2 can be calculated in a manner similar to that for $\overline{n}$. We write

$$\overline{n^2} = \sum_{n=0}^{N} n^2 \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n} \tag{3.99a}$$

$$= \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!} \big(p\,\frac{\partial}{\partial p}\big)^2\, p^n q^{N-n} \tag{3.99b}$$

$$= \big(p\,\frac{\partial}{\partial p}\big)^2 \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n} = \big(p\,\frac{\partial}{\partial p}\big)^2\,(p+q)^N \tag{3.99c}$$

$$= \big(p\,\frac{\partial}{\partial p}\big)\big[pN(p+q)^{N-1}\big] \tag{3.99d}$$

$$= p\big[N(p+q)^{N-1} + pN(N-1)(p+q)^{N-2}\big]. \tag{3.99e}$$

Because we are interested in the case $p+q=1$, we have

$$\overline{n^2} = p\,[N + pN(N-1)] \tag{3.100a}$$

$$= p\,[pN^2 + N(1-p)] = (pN)^2 + p\,(1-p)N \tag{3.100b}$$

$$= \overline{n}^2 + pqN, \tag{3.100c}$$

where we have used (3.98) and let $q = 1-p$. Hence, from (3.100c) we find that the variance of n is given by

$${\sigma_n}^2 = \overline{(\Delta n)^2} = \overline{n^2} - \overline{n}^2 = pqN. \tag{3.101}$$

Problem 3.33. Compare the calculated values of σ_n from (3.101) with your estimates in Problem 3.30 and to the exact result (3.84f) for $N=3$.

The relative width of the probability distribution of n is given by (3.98) and (3.101)

$$\frac{\sigma_n}{\overline{n}} = \frac{\sqrt{pqN}}{pN} = \Big(\frac{q}{p}\Big)^{1/2} \frac{1}{\sqrt{N}}. \tag{3.102}$$

We see that the relative width goes to zero as $1/\sqrt{N}$.

Frequently we need to evaluate $\ln N!$ for $N \gg 1$. A simple approximation for $\ln N!$ known as *Stirling's approximation* is

$$\ln N! \approx N \ln N - N. \qquad \text{(Stirling's approximation)} \tag{3.103}$$

A more accurate approximation is given by

$$\ln N! \approx N \ln N - N + \frac{1}{2} \ln(2\pi N). \tag{3.104}$$

A simple derivation of Stirling's approximation is given in Appendix A.

Problem 3.34. Applicability of Stirling's approximation

(a) What is the largest value of $\ln N!$ that you can calculate exactly using a typical hand calculator?

(b) Compare the approximations (3.103) and (3.104) to each other and to the exact value of $\ln N!$ for $N = 5$, 10, 20, and 50. If necessary, compute $\ln N!$ directly using the relation

$$\ln N! = \sum_{m=1}^{N} \ln m. \tag{3.105}$$

(c) Use the simple form of Stirling's approximation to show that

$$\frac{d}{dx} \ln x! = \ln x \text{ for } x \gg 1. \tag{3.106}$$

Problem 3.35. Consider the binomial distribution $P_N(n)$ for $N = 16$ and $p = q = 1/2$. What is the value of $P_N(n)$ for $n = \sigma_n/2$? What is the value of the product $P_N(n = \overline{n})\sigma_n$?

Problem 3.36. Density fluctuations

A container of volume V contains N molecules of a gas. We assume that the gas is dilute so that the position of any one molecule is independent of all other molecules. Although the density will be uniform on the average, there are fluctuations in the density. Divide the volume V into two parts V_1 and V_2, where $V = V_1 + V_2$.

(a) What is the probability p that a particular molecule is in each part?

(b) What is the probability that N_1 molecules are in V_1 and N_2 molecules are in V_2?

(c) What is the average number of molecules in each part?

(d) What are the relative fluctuations of the number of particles in each part?

Problem 3.37. Random walk

Suppose that a random walker takes n steps to the right and n' steps to the left. Each step is of equal length a and the probability of a step to the right is p. Denote x as the net displacement of a walker. What is the mean value $\overline{x}$ for a N-step random walk? What is the analogous expression for the variance $\overline{(\Delta x)^2}$?

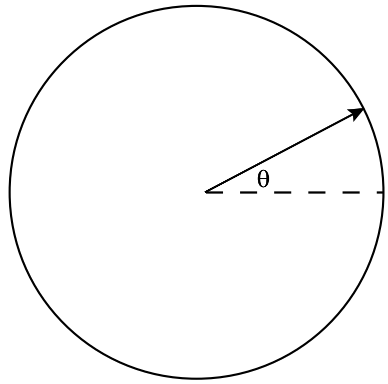

Figure 3.5: The angle θ is an example of a continuous random variable.

3.6 Continuous Probability Distributions

In many cases of physical interest the random variables have continuous values. Examples of continuous variables are the position of the holes in a dart board, the position and velocity of a classical particle, and the angle of a compass needle.

As an example, consider a spinner, the equivalent of a wheel of fortune,[10] with an arrow that spins around and stops at some angle at random (see Figure 3.5). In this case the variable θ is a continuous random variable that takes all values in the interval $[0, 2\pi]$. What is the probability that θ has a particular value? Because there are an infinite number of possible values of θ in the interval $[0, 2\pi]$, the probability of obtaining any particular value of θ is zero. We say that the values of θ are not countable. Instead, we have to reformulate the question and ask for the probability that the value of θ is between θ and $\theta + \Delta theta$. In other words, we have to ask for the probability that θ is in a particular bin of width $d\theta$. For example, the probability that θ is between, say, 0 and π or between $\pi/4$ and $\pi/2$ is clear.

Another example of a continuous random variable is the displacement from the origin of a one-dimensional random walker that steps at random to the right with probability p, but with a step length that is chosen at random between zero and the maximum step length a. The continuous nature of the step length means that the displacement x of the walker is a continuous variable. If we do a simulation of this random walk, we can record the number of times $H(x)$ that the displacement of the walker from the origin after N steps is in a bin of width Δx between x and $x + \Delta x$. A plot of $H(x)$ as a function of x for a given bin width Δx is shown in Figure 3.6). If the number of walkers that is sampled is sufficiently large, we would find that $H(x)$ is proportional to the estimated probability that a walker is in a bin of width Δx a distance x from the origin after N steps. To obtain the probability, we divide $H(x)$ by the total number of walkers.

In practice, the choice of the bin width is a compromise. If Δx is too big, the features of the histogram would be lost. If Δx is too small, many of the bins would be empty for a given number of walkers, and our estimate of the number of walkers in each bin would be less accurate.

[10]The Wheel of Fortune is an American television game show that involves three contestants. The name of the show comes from the large spinning wheel that determines the dollar amounts and prizes won by the contestants.

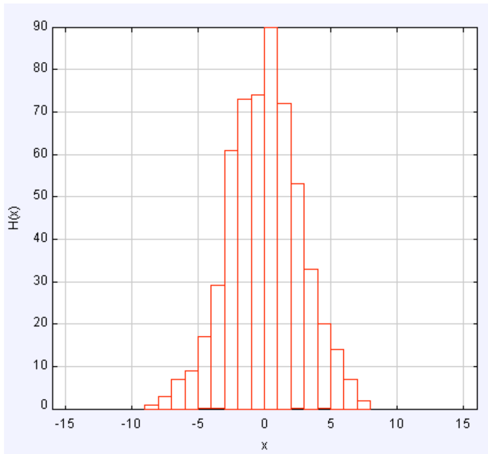

Figure 3.6: Histogram of the number of times that the displacement of a one-dimensional random walker is between x and $x + \Delta x$ after $N = 16$ steps. The data was generated by simulating 565 walkers, a relatively small number in this case. The length of each step was chosen at random between zero and unity and the bin width is $\Delta x = 1$.

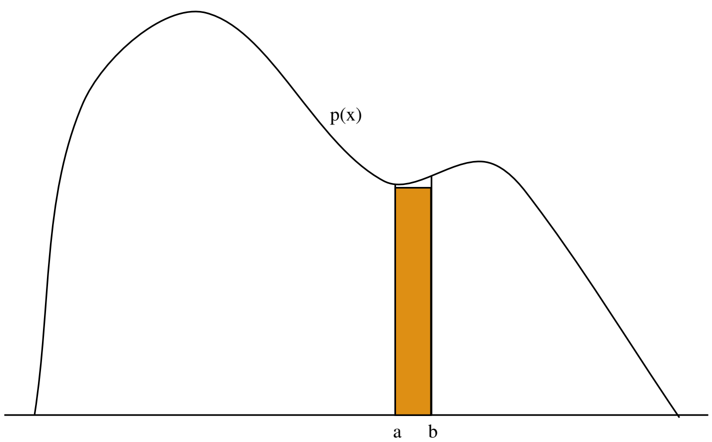

Figure 3.7: The probability that x is between a and b is equal to the shaded area.

Because we expect the number of walkers in a particular bin to be proportional to the width of the bin, we can write $H(x) = p(x)\Delta x$. The quantity $p(x)$ is called the *probability density*. In the limit that $\Delta x \to 0$, $H(x)$ becomes a continuous function of x, and we can write the probability that the displacement of the walker is between a and b as (see Figure 3.7).

$$P(a \text{ to } b) = \int_a^b p(x)\, dx. \tag{3.107}$$

Note that the probability density $p(x)$ is nonnegative and has units of one over the dimension of x.

The formal properties of the probability density $p(x)$ can be generalized from the discrete case. For example, the normalization condition is given by

$$\int_{-\infty}^{\infty} p(x)\, dx = 1. \tag{3.108}$$

The mean value of the function $f(x)$ in the interval a to b is given by

$$\overline{f} = \int_a^b f(x)\, p(x)\, dx. \tag{3.109}$$

Problem 3.38. Simulation of a one-dimensional random walk with variable step length

The applet/application at `<stp.clarku.edu/simulations/randomwalks/continuous1d.html>` simulates a random walk in one dimension with a variable jump length.

(a) The simulation uses a step length with a uniform probability between 0 and 1. Calculate the mean step length and its variance.

(b) How does the variance found in the simulation depend on the variance of the step length that you calculated in part (a)?

(c) Does the qualitative features of the histogram change as the number of walkers increases?

(d) Explore how the histogram changes with the bin width. What is a reasonable choice of the bin width for $N = 100$?

Problem 3.39. Exponential probability density

The random variable x has the probability density

$$p(x) = \begin{cases} A\, e^{-\lambda x} & \text{if } 0 \leq x \leq \infty \\ 0 & x < 0. \end{cases} \tag{3.110}$$

(a) Determine the normalization constant A in terms of λ.

(b) What is the mean value of x? What is the most probable value of x?

(c) What is the mean value of x^2?

(d) Choose $\lambda = 1$ and determine the probability that a measurement of x yields a value between 1 and 2.

(e) Choose $\lambda = 1$ and determine the probability that a measurement of x yields a value less than 0.3.

Problem 3.40. Probability density for velocity

Consider the probability density function $p(\mathbf{v}) = (a/\pi)^{3/2}\, e^{-av^2}$ for the velocity $\mathbf{v}$ of a particle, where $\mathbf{v} = |\mathbf{v}|$ and $v^2 = v_x^2 + v_y^2 + v_z^2$. Each of the three velocity components can range from $-\infty$ to $+\infty$ and a is a constant.

(a) What is the probability that a particle has a velocity between v_x and $v_x + dv_x$, v_y and $v_y + dv_y$, and v_z and $v_z + dv_z$?

(b) Show that $p(\mathbf{v})$ is normalized to unity. Use the fact that

$$\int_0^\infty e^{-au^2}\,du = \frac{1}{2}\sqrt{\frac{\pi}{a}}. \tag{3.111}$$

Note that this calculation involves doing three similar integrals that can be evaluated separately.

(c) What is the probability that $v_x \geq 0$, $v_y \geq 0$, $v_z \geq 0$ simultaneously?

Problem 3.41. Gaussian probability density

(a) Find the first four moments of the Gaussian probability density

$$p(x) = (2\pi)^{-\frac{1}{2}} e^{-x^2/2}. \qquad (-\infty < x < \infty) \tag{3.112}$$

Guess the dependence of the kth moment on k for k even. What are the odd moments of $p(x)$?

(b) Calculate the value of C_4, the fourth-order cumulant, defined by

$$C_4 = \overline{x^4} - 4\overline{x^3}\,\overline{x} - 3\,\overline{x^2}^2 + 12\,\overline{x^2}\,\overline{x}^2 - 6\,\overline{x}^4. \tag{3.113}$$

Problem 3.42. Uniform probability distribution
Consider the probability density given by

$$p(x) = \begin{cases} (2a)^{-1} & \text{for } |x| \leq a \\ 0 & \text{for } |x| > a \end{cases} \tag{3.114}$$

(a) Sketch the dependence of $p(x)$ on x.

(b) Find the first four moments of $p(x)$.

(c) Calculate the value of the fourth-order cumulant C_4 defined in (3.113)). What is C_4 for the probability density in (3.114)? Compare your result to the corresponding result for C_4 for the Gaussian distribution.

Problem 3.43. Cauchy probability distribution
Not all probability densities have a finite variance.

(a) Sketch the *Lorentz* or *Cauchy distribution* given by

$$p(x) = \frac{1}{\pi}\frac{\gamma}{(x-a)^2 + \gamma^2}. \qquad (-\infty < x < \infty) \tag{3.115}$$

Choose $a = 0$ and $\gamma = 1$ and compare the form of $p(x)$ in (3.115) to the Gaussian distribution given by (3.112).

(b) Give a simple argument for the existence of the first moment of the Lorentz distribution. Does the second moment exist?

3.7 The Gaussian Distribution as a Limit of the Binomial Distribution

In Problem 3.30 we found that for large N, the binomial distribution has a well-defined maximum at $n = pN$ and can be approximated by a smooth, continuous function even though only integer values of n are physically possible. We now find the form of this function of n.

The first step is to realize that for $N \gg 1$, $P_N(n)$ is a rapidly varying function of n near $n = pN$, and for this reason we do not want to approximate $P_N(n)$ directly. Because the logarithm of $P_N(n)$ is a slowly varying function (see Problem 3.30), we expect that the power series expansion of $\ln P_N(n)$ to converge. Hence, we expand $\ln P_N(n)$ in a Taylor series about the value of $n = \tilde{n}$ at which $\ln P_N(n)$ reaches its maximum value. We will write $p(n)$ instead of $P_N(n)$ because we will treat n as a continuous variable and hence $p(n)$ is a probability density. We find

$$\ln p(n) = \ln p(n = \tilde{n}) + (n - \tilde{n})\frac{d\ln p(n)}{dn}\Big|_{n=\tilde{n}} + \frac{1}{2}(n - \tilde{n})^2\frac{d^2\ln p(n)}{d^2n}\Big|_{n=\tilde{n}} + \cdots \tag{3.116}$$

Because we have assumed that the expansion (3.116) is about the maximum $n = \tilde{n}$, the first derivative $d\ln p(n)/dn\big|_{n=\tilde{n}}$ must be zero. For the same reason the second derivative $d^2\ln p(n)/dn^2\big|_{n=\tilde{n}}$ must be negative. We assume that the higher terms in (3.116) can be neglected and adopt the notation

$$\ln A = \ln p(n = \tilde{n}), \tag{3.117}$$

and

$$B = -\frac{d^2\ln p(n)}{dn^2}\Big|_{n=\tilde{n}}. \tag{3.118}$$

The approximation (3.116) and the notation in (3.117) and (3.118) allows us to write

$$\ln p(n) \approx \ln A - \frac{1}{2}B(n - \tilde{n})^2, \tag{3.119}$$

or

$$p(n) \approx A\, e^{-\frac{1}{2}B(n-\tilde{n})^2}. \tag{3.120}$$

We next use Stirling's approximation to evaluate the first two derivatives of $\ln p(n)$ and the value of $\ln p(n)$ at its maximum to find the parameters A, B, and $\tilde{n}$. We write

$$\ln p(n) = \ln N! - \ln n! - \ln(N - n)! + n\ln p + (N - n)\ln q. \tag{3.121}$$

It is straightforward to use the relation (3.106) to obtain

$$\frac{d(\ln p(n))}{dn} = -\ln n + \ln(N - n) + \ln p - \ln q. \tag{3.122}$$

The most probable value of n is found by finding the value of n that satisfies the condition $d\ln p/dn = 0$. We find

$$\frac{N - \tilde{n}}{\tilde{n}} = \frac{q}{p}, \tag{3.123}$$

or $(N-\tilde{n})p = \tilde{n}q$. If we use the relation $p+q=1$, we obtain

$$\tilde{n} = pN. \tag{3.124}$$

Note that $\tilde{n} = \overline{n}$, that is, the value of n for which $p(n)$ is a maximum is also the mean value of n.

The second derivative can be found from (3.122). We have

$$\frac{d^2(\ln p(n))}{dn^2} = -\frac{1}{n} - \frac{1}{N-n}. \tag{3.125}$$

Hence, the coefficient B defined in (3.118) is given by

$$B = -\frac{d^2 \ln p(n)}{dn^2}\Big|_{n=\tilde{n}} = \frac{1}{\tilde{n}} + \frac{1}{N-\tilde{n}} = \frac{1}{Npq}. \tag{3.126}$$

From the relation (3.101) we see that

$$B = \frac{1}{\sigma^2}, \tag{3.127}$$

where σ^2 is the variance of n.

If we use the simple form of Stirling's approximation (3.103) to find the normalization constant A from the relation $\ln A = \ln p(n=\tilde{n})$, we would find that $\ln A = 0$. Instead, we have to use the more accurate form of Stirling's approximation (3.104). The result is

$$A = \frac{1}{(2\pi Npq)^{1/2}} = \frac{1}{(2\pi\sigma^2)^{1/2}}. \tag{3.128}$$

Problem 3.44. Derive (3.128) using the more accurate form of Stirling's approximation (3.104) with $n = pN$ and $N - n = qN$.

If we substitute our results for $\tilde{n}$, B, and A into (3.120), we find the standard form for the Gaussian probability distribution

$$\boxed{p(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(n-\overline{n})^2/2\sigma^2}}. \qquad \text{(Gaussian probability density)} \tag{3.129}$$

An alternative derivation of the parameters A, B, and $\tilde{n}$ is given in Problem 3.75.

From our derivation we see that (3.129) is valid for large values of N and for values of n near $\overline{n}$. Even for relatively small values of N, the Gaussian approximation is a good approximation for most values of n. A comparison of the Gaussian approximation to the binomial distribution is given in Table 3.4.

The most important feature of the Gaussian probability distribution is that its relative width, $\sigma_n/\overline{n}$, decreases as $N^{-1/2}$. Of course, the binomial distribution shares this feature.

n	$P_{10}(n)$	Gaussian approximation
0	0.000977	0.001700
1	0.009766	0.010285
2	0.043945	0.041707
3	0.117188	0.113372
4	0.205078	0.206577
5	0.246094	0.252313

Table 3.4: Comparison of the exact values of $P_{10}(n)$ with the Gaussian probability distribution (3.129) for $p = q = 1/2$.

3.8 The Central Limit Theorem or Why is Thermodynamics Possible?

We have discussed how to estimate probabilities empirically by sampling, that is, by making repeated measurements of the outcome of independent events. Intuitively we believe that if we perform more and more measurements, the calculated average will approach the exact mean of the quantity of interest. This idea is called *the law of large numbers*. However, we can go further and find the form of the probability distribution that a particular measurement differs from the exact mean. The form of this probability distribution is given by the *central limit theorem*. We first illustrate this theorem by considering a simple measurement.

Suppose that we wish to estimate the probability of obtaining face 1 in one throw of a die. The answer of $\frac{1}{6}$ means that if we perform N measurements, face 1 will appear approximately $N/6$ times. What is the meaning of approximately? Let S be the total number of times that face one appears in N measurements. We write

$$S = \sum_{i=1}^{N} s_i, \tag{3.130}$$

where

$$s_i = \begin{cases} 1, & \text{if the } i\text{th throw gives 1} \\ 0 & \text{otherwise.} \end{cases} \tag{3.131}$$

If N is large, then S/N approaches $1/6$. How does this ratio approach the limit? We can empirically answer this question by repeating the measurement M times. (Each measurement of S consists of N throws of a die.) Because S itself is a random variable, we know that the measured values of S will not be identical. In Figure 3.8 we show the results of $M = 10,000$ measurements of S for $N = 100$ and $N = 800$. We see that the approximate form of the distribution of values of S is a Gaussian. In Problem 3.45 we calculate the absolute and relative width of the distributions.

Problem 3.45. Estimate the absolute width and the relative width of the distributions shown in Figure 3.8 for $N = 100$ and $N = 800$. Does the error of any one measurement of S decreases with increasing N as expected? How would the plot change if M were increased to $M = 10,000$?

In Section 3.12.3 we show that in the limit of large N, the probability density $p(S)$ is given

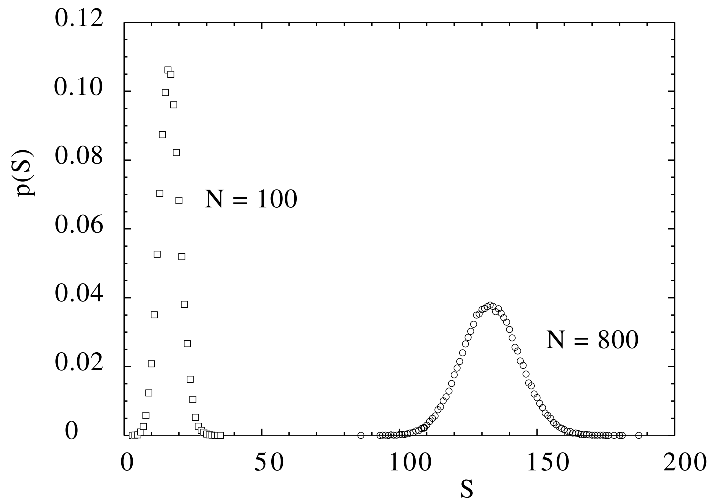

Figure 3.8: The distribution of the measured values of $M = 10,000$ different measurements of the sum S for $N = 100$ and $N = 800$ terms in the sum. The quantity S is the number of times that face 1 appears in N throws of a die. For $N = 100$, the measured values are $\overline{S} = 16.67$, $\overline{S^2} = 291.96$, and $\sigma_S = 3.74$. For $N = 800$, the measured values are $\overline{S} = 133.31$, $\overline{S^2} = 17881.2$, and $\sigma_S = 10.52$. What are the estimated values of the relative width for each case?

by

$$p(S) = \frac{1}{\sqrt{2\pi\sigma_S^2}} e^{-(S-\overline{S})^2/2\sigma_S^2}, \tag{3.132}$$

where

$$\overline{S} = N\overline{s} \tag{3.133}$$

$$\sigma_S^2 = N\sigma^2, \tag{3.134}$$

with $\sigma^2 = \overline{s^2} - \overline{s}^2$. The quantity $p(S)\Delta S$ is the probability that the value of $\sum_{i=1}^N s_i$ is between S and $S + \Delta S$. Equation (3.132) is equivalent to the central limit theorem. Note that the Gaussian form in (3.132) holds only for large N and for values of S near its most probable (mean) value. The latter restriction is the reason that the theorem is called the *central* limit theorem; the requirement that N be large is the reason for the term *limit*.

The central limit theorem is one of the most remarkable results of the theory of probability. In its simplest form, the theorem states that the probability o the sum of a large number of random variables approximates a Gaussian distribution. Moreover, the approximation steadily improves as the number of variables in the sum increases.

For the throw of a die, $\overline{s} = \frac{1}{6}$, $\overline{s^2} = \frac{1}{6}$, and $\sigma^2 = \overline{s^2} - \overline{s}^2 = \frac{1}{6} - \frac{1}{36} = \frac{5}{36}$. For N throws of a die, we have $\overline{S} = N/6$ and $\sigma_S^2 = 5N/36$. Hence, we see that in this case the most probable relative error in any one measurement of S decreases as $\sigma_S/\overline{S} = \sqrt{5/N}$.

Note that if we let S represent the displacement of a walker after N steps, and let σ^2 equal the mean square displacement for a single step, then the result (3.132)–(3.134) is equivalent to our results for random walks in the limit of large N. Or we can let S represent the magnetization of a system of noninteracting spins and obtain similar results. That is, a random walk and its equivalents are examples of an *additive* random process.

The central limit theorem shows why the Gaussian probability density is ubiquitous in nature. If a random process is related to a sum of a large number of microscopic processes, the sum will be distributed according to the Gaussian distribution *independently* of the nature of the distribution of the microscopic processes.

The central limit theorem implies that macroscopic bodies have well defined macroscopic properties even though their constituent parts are changing rapidly. For example in a gas or liquid, the particle positions and velocities are continuously changing at a rate much faster than a typical measurement time. For this reason we expect that during a measurement of the pressure of a gas or a liquid, there are many collisions with the wall and hence the pressure has a well defined average. We also expect that the probability that the measured pressure deviates from its average value is proportional to $N^{-1/2}$, where N is the number of particles. Similarly, the vibrations of the molecules in a solid have a time scale much smaller than that of macroscopic measurements, and hence the pressure of a solid also is a well-defined quantity.

Problem 3.46. Random walks and the central limit theorem Use the central limit theorem to show that the probability that a one-dimensional random walker has a displacement between x and $x + dx$. (There is no need to derive the central limit theorem.)

Problem 3.47. Central limit theorem

Use the applet/application at `<stp.clarku.edu/simulations/centralLimitTheorem.html>` to test the applicability of the central limit theorem.

(a) First assume that the variable s_i is uniformly distributed between 0 and 1. Calculate the mean and standard deviation of s and compare your numerical results with your analytical calculation.

(b) Use the default value of N, the number of terms in the sum, and describe the qualitative form of $p(S)$, where $p(S)\Delta S$ is the probability that the sum S is between S and $S + \Delta S$. Does the qualitative form of $p(S)$ change as the number of measurements (trials) of S is increased for a given value of N?

(c) What is the approximate width of $p(S)$ for $N = 12$? Describe the changes, if any, of the width of $p(S)$ as N is increased. Increase N by at least a factor of 4. Do your results depend strongly on the number of measurements?

(d) To determine the generality of your results, consider the probability density $f(s) = e^{-s}$ for $s \geq 0$ and answer the same questions as in parts (a)–(c).

(e) Consider the Lorentz distribution $f(s) = (1/\pi)(1/(s^2 + 1)$, where $-\infty \leq s \leq \infty$. What is the mean value and variance of s? Is the form of $p(S)$ consistent with the results that you found in parts (b)–(d)?

(f) Each value of S can be considered to be a measurement. The sample variance $\tilde{\sigma}_S^2$ is a measure of the square of the differences of the result of each measurement and is given by

$$\tilde{\sigma}_S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(S_i - \overline{S})^2. \tag{3.135}$$

The reason for the factor of $N-1$ rather than N in the definition of $\tilde{\sigma}_S^2$ is that to compute it, we need to use the N values of s to compute the mean of S, and thus, loosely speaking, we have only $N-1$ independent values of s remaining to calculate $\tilde{\sigma}_S^2$. Show that if $N \gg 1$, then $\tilde{\sigma}_S \approx \sigma_S$, where the standard deviation σ_S is given by $\sigma_S^2 = \overline{S^2} - \overline{S}^2$.

(g) The quantity $\tilde{\sigma}_S$ is known as the standard deviation of the means. That is, $\tilde{\sigma}_S$ is a measure of how much variation we expect to find if we make repeated measurements of S. How does the value of $\tilde{\sigma}_S$ compare to your estimated width of the probability density $p(S)$?

3.9 The Poisson distribution and Should You Fly in Airplanes?

We now return to the question of whether or not it is safe to fly. If the probability of a plane crashing is $p = 10^{-5}$, then $1-p$ is the probability of surviving a single flight. The probability of surviving N flights is then $P_N = (1-p)^N$. For $N = 400$, $P_N \approx 0.996$, and for $N = 10^5$, $P_N \approx 0.365$. Thus, our intuition is verified that if we took 400 flights, we would have only a small chance of crashing.

This type of reasoning is typical when the probability of an individual event is small, but there are very many attempts. Suppose we are interested in the probability of the occurrence of n events out of N attempts given that the probability p of the event for each attempt is very small. The resulting probability is called the *Poisson distribution*, a distribution that is important in the analysis of experimental data. We discuss it here because of its intrinsic interest.

To derive the Poisson distribution, we begin with the binomial distribution:

$$P(n) = \frac{N!}{n!\,(N-n)!}\,p^n(1-p)^{N-n}. \tag{3.136}$$

(As before, we suppress the N dependence of P.) As in Section (3.7, we will approximate $\ln P(n)$ rather than $P(n)$ directly. We first use Stirling's approximation to write

$$\ln\frac{N!}{(N-n)!} = \ln N! - \ln(N-n)! \tag{3.137a}$$

$$\approx N\ln N - (N-n)\ln(N-n) \tag{3.137b}$$

$$\approx N\ln N - (N-n)\ln N \tag{3.137c}$$

$$= N\ln N - N\ln N + n\ln N \tag{3.137d}$$

$$= n\ln N. \tag{3.137e}$$

From (3.137e) we obtain

$$\frac{N!}{(N-n)!} \approx e^{n \ln N} = N^n. \tag{3.138}$$

For $p \ll 1$, we have $\ln(1-p) \approx -p$, $e^{\ln(1-p)} = 1 - p \approx e^{-p}$, and $(1-p)^{N-n} \approx e^{-p(N-n)} \approx e^{-pN}$. If we use the above approximations, we find

$$P(n) \approx \frac{N^n}{n!}\, p^n e^{-pN} = \frac{(Np)^n}{n!} e^{-pN}, \tag{3.139}$$

or

$$\boxed{P(n) = \frac{\overline{n}^n}{n!} e^{-\overline{n}}}, \qquad \text{(Poisson distribution)} \tag{3.140}$$

where $\overline{n} = pN$. The form (3.140) is the Poisson distribution.

Let us apply the Poisson distribution to the airplane survival problem. We want to know the probability of never crashing, that is, $P(n = 0)$. The mean $\overline{N} = pN$ equals $10^{-5} \times 400 = 0.004$ for $N = 400$ flights and $\overline{N} = 1$ for $N = 10^5$ flights. Thus, the survival probability is $P(0) = e^{-\overline{N}} \approx 0.996$ for $N = 400$ and $P(0) \approx 0.368$ for $N = 10^5$ as we calculated previously. We see that if we fly 100,000 times, we have a much larger probability of dying in a plane crash.

Problem 3.48. Poisson distribution

(a) Show that the Poisson distribution is properly normalized, and calculate the mean and variance of n. Because $P(n)$ for $n > N$ is negligibly small, you can sum $P(n)$ from $n = 0$ to $n = \infty$ even though the maximum value of n is N.

(b) Plot the Poisson distribution $P(n)$ as a function of n for $p = 0.01$ and $N = 100$.

3.10 *Traffic Flow and the Exponential Distribution

The Poisson distribution is closely related to the exponential distribution as we will see in the following. Consider a sequence of similar random events and let t_1, t_2, ... be the time at which each successive event occurs. Examples of such sequences are the successive times when a phone call is received and the times when a Geiger counter registers a decay of a radioactive nucleus. Suppose that we determine the sequence over a very long time T that is much greater than any of the intervals $t_i - t_{i-1}$. We also suppose that the average number of events is λ per unit time so that in a time interval t, the mean number of events is λt.

Assume that the events occur at random and are independent of each other. Given λ, the mean number of events per unit time, we wish to find the probability distribution $w(t)$ of the interval t between the events. We know that if an event occurred at time $t = 0$, the probability that another event occurs within the interval $[0, t]$ is

$$\int_0^t w(t)\Delta t, \tag{3.141}$$

and the probability that no event occurs in the interval t is

$$1 - \int_0^t w(t)\Delta t. \tag{3.142}$$

Thus the probability that the duration of the interval between the two events is between t and $t + \Delta t$ is given by

$$\begin{aligned} w(t)\Delta t &= \text{probability that no event occurs in the interval } [0,t] \\ &\quad \times \text{ probability that an event occurs in interval } [t, t+\Delta t] \\ &= \big[1 - \int_0^t w(t)dt\big]\lambda\Delta t. \end{aligned} \tag{3.143}$$

If we cancel Δt from each side of (3.143) and differentiate both sides with respect to t, we find

$$\frac{dw}{dt} = -\lambda w, \tag{3.144}$$

so that

$$w(t) = Ae^{-\lambda t}. \tag{3.145}$$

The constant of integration A is determined from the normalization condition:

$$\int_0^\infty w(t)\,dt = 1 = A\int_0^\infty e^{-\lambda t}dt = A/\lambda. \tag{3.146}$$

Hence, $w(t)$ is the exponential function

$$w(t) = \lambda e^{-\lambda t}. \tag{3.147}$$

These results for the exponential distribution lead naturally to the Poisson distribution. Let us divide a long time interval T into n smaller intervals $t = T/n$. What is the probability that 0, 1, 2, 3, … events occur in the time interval t, given λ, the mean number of events per unit time? We will show that the probability that n events occur in the time interval t is given by the Poisson distribution:

$$P_n(t) = \frac{(\lambda t)^n}{n!}e^{-\lambda t}. \tag{3.148}$$

We first consider the case $n = 0$. If $n = 0$, the probability that no event occurs in the interval t is (see (3.143))

$$P_{n=0}(t) = 1 - \lambda\int_0^t e^{-\lambda t'}dt' = e^{-\lambda t}. \tag{3.149}$$

For the case $n = 1$, there is exactly one event in time interval t. This event must occur at some time t' which may occur with equal probability in the interval $[0, t]$. Because no event can occur in the interval $[t', t]$, we have

$$P_{n=1}(t) = \int_0^t \lambda e^{-\lambda t'}e^{-\lambda(t'-t)}dt', \tag{3.150}$$

where we have used (3.149) with $t \to (t' - t)$. Hence,

$$P_{n=1}(t) = \int_0^t \lambda e^{-\lambda t} dt = (\lambda t)e^{-\lambda t}. \tag{3.151}$$

In general, if n events are to occur in the interval $[0, t]$, the first must occur at some time t' and exactly $(n-1)$ must occur in the time $(t - t')$. Hence,

$$P_n(t) = \int_0^t \lambda e^{-\lambda t'} P_{n-1}(t - t'). \tag{3.152}$$

Equation (3.152) is a recurrence formula that can be used to derive (3.148) by induction. It is easy to see that (3.148) satisfies (3.152) for $n = 0$ and 1. As is usual when solving recursion formula by induction, we assume that (3.148) is correct for $(n-1)$. We substitute this result into (3.152) and find

$$P_n(t) = \lambda^n e^{-\lambda t} \int_0^t (t - t')^{n-1} dt'/(n-1)! = \frac{(\lambda t)^b}{n!} e^{-\lambda t}. \tag{3.153}$$

An application of the Poisson distribution is given in Problem 3.49.

N	frequency
0	1
1	7
2	14
2	25
4	31
5	26
6	27
7	14
8	8
9	3
10	4
11	3
12	1
13	0
14	1
> 15	0

Table 3.5: Observed distribution of vehicles passing a marker on a highway in thirty second intervals. The data was taken from page 98 of Montroll and Badger.

*__Problem 3.49.__ Analysis of traffic data

In Table 3.5 we show the number of vehicles passing a marker during a thirty second interval. The observations were made on a single lane of a six lane divided highway. Assume that the traffic density is so low that passing occurs easily and no platoons of cars develop.

(a) Is the distribution of the number of vehicles consistent with the Poisson distribution? If so, what is the value of the parameter λ?

(b) As the traffic density increases, the flow reaches a regime where the vehicles are very close to one another so that they are no longer mutually independent. Make arguments for the form of the probability distribution of the number of vehicles passing a given point in this regime.

3.11 *Are All Probability Distributions Gaussian?

We have discussed the properties of *random additive processes* and found that the probability distribution for their sum is a Gaussian for a sufficiently large number of terms. As an example of such a process, we discussed a one-dimensional random walk on a lattice for which the displacement x is the sum of N random steps.

We now discuss *random multiplicative processes*. Examples of such processes include the distributions of incomes, rainfall, and fragment sizes in rock crushing processes.[11] Consider the latter for which we begin with a rock of size w. We strike the rock with a hammer and generate two fragments whose sizes are pw and qw, where $q = 1 - p$. In the next step the possible sizes of the fragments are p^2w, pqw, qpw, and q^2w. What is the distribution of the fragment sizes after N blows of the hammer?

To answer this question, we consider a binary sequence in which the numbers x_1 and x_2 appear independently with probabilities p and q respectively. If there are N elements in the product Π, we can ask what is $\overline{\Pi}$, the mean value of Π? To calculate $\overline{\Pi}$ we define $P(n)$ to be the probability that the product of N independent factors of x_1 and x_2 has the value ${x_1}^n {x_2}^{N-n}$. This probability is given by the number of sequences where x_1 appears n times multiplied by the probability of choosing a specific sequence with x_1 appearing n times. This probability is the familiar binomial distribution:

$$P_N(n) = \frac{N!}{n!\,(N-n)!} p^n q^{N-n}. \tag{3.154}$$

We average over all possible outcomes of the product to obtain its mean value

$$\overline{\Pi} = \sum_{n=0}^{N} P(n) {x_1}^n {x_2}^{N-n} = (px_1 + qx_2)^N. \tag{3.155}$$

The most probable event in the product contains Np factors of x_1 and Nq factors of x_2. Hence, the most probable value of the product is

$$\widetilde{\Pi} = ({x_1}^p {x_2}^q)^N. \tag{3.156}$$

To obtain a better feeling for these results, we first consider some special cases. For $x_1 = 2$, $x_2 = 1/2$, and $p = q = 1/2$ we have $\overline{\Pi} = (1/4)[x_2^2 + 2x_1x_2 + x_2^2] = (1/4)[4 + 2 + 1/4] = 25/16$ for $N = 2$; for general N we have $\overline{\Pi} = (5/4)^N$. In contrast, the most probable value for $N = 2$ is given by $\widetilde{\Pi} = 2^{1/2} \times (1/2)^{1/2} = 1$; the same result holds for any N. For $p = 1/3$ and $q = 2/3$ and the same values of x_1 and x_2 we find $\overline{\Pi} = 1$ for all N and $\widetilde{\Pi} = (\frac{1}{2} \times \frac{1}{2} \times 2)^{2/3} = 2^{-2/3}$ for $N = 2$ and $2^{-N/3}$ for any N. We see that $\widetilde{\Pi} \neq \overline{\Pi}$ for a random multiplicative process. In contrast, the

[11] The following discussion is based on an article by Sidney Redner (see references).

most probable event is a good approximation to the mean value of the sum orf a random additive process (and is identical for $p = q$).

The reason for the large discrepancy between $\overline{\Pi}$ and $\widetilde{\Pi}$ is the important role played by rare events. For example, a sequence of N factors of $x_1 = 2$ occurs with a very small probability, but the value of this product is very large in comparison to the most probable value. Hence, this extreme event makes a finite contribution to $\overline{\Pi}$ and a dominant contribution to the higher moments $\overline{\Pi^m}$.

*__Problem 3.50.__ (a) Confirm the general result in (3.155) for $N = 4$ by showing explicitly all the possible values of the product.

(b) Consider the case $x_1 = 2$, $x_2 = 1/2$, $p = 1/4$, and $q = 3/4$, and calculate $\overline{\Pi}$ and $\widetilde{\Pi}$.

(c) Show that $\overline{\Pi^m}$, the mean value of the mth moment, reduces to $(px_1^m)^N$ as $m \to \infty$. This result implies that the mth moment is determined solely by the most extreme event for $m \gg 1$.

(d) Based on the Gaussian approximation for the probability of a random additive process, what is a reasonable guess for the continuum approximation to the probability of a random multiplicative process? Such a distribution is called the *log-normal* distribution. Discuss why or why not you expect the log-normal distribution to be a good approximation for $N \gg 1$.

*__Problem 3.51.__ Simulation of multiplicative process

(a) Run the applet/application at `<stp.clarku.edu/simulations/productprocess.html>` which simulates the distribution of values of the product ${x_1}^n {x_2}^{N-n}$. Choose $x_1 = 2$, $x_2 = 1/2$, and $p = q = 1/2$. First choose $N = 4$ and estimate $\overline{\Pi}$ and $\widetilde{\Pi}$. Do your estimated values converge more or less uniformly to the analytical values as the number of measurements becomes large? Do a similar simulation for $N = 40$. Compare your results with a similar simulation of a random walk and discuss the importance of extreme events for random multiplicative processes.

(b) The true average value of a product of random variables is governed by rare events that are at the tail of the distribution. However, the most probable events will likely dominate in a simulation of a multiplicative process. As the number of trials increase, there will be an increase in the number of rare events that are sampled, and we expect that the observed averages will fluctuate greatly. As the number of trials is increased still further, the number of rare events will be more accurately sampled, and the observed averages will eventually converge to their true values. Redner has estimated that the minimum number of trials for this crossover to occur is given by

$$\ln T^* = \frac{N}{2pq}\Big(p - \frac{pq(x_1/x_2)^m}{q + p(x_1/x_2)^m}\Big)^2, \tag{3.157}$$

where T is the number of trials and m is the moment of the distribution that we wish to estimate. How does the estimate of T^* in (3.157) compare with the results you observe in the simulation?

3.12 *Supplementary Notes

3.12.1 The uncertainty for unequal probabilities

Consider a loaded die for which the probabilities P_j are not equal. We wish to motivate the form (3.38) for S. Imagine that we roll the die a large number of times N. Then each outcome would occur $N_j = NP_j$ times and there would be $N_j = NP_1$ outcomes of face 1, NP_2 outcomes of face 2, ... These outcomes could occur in many different orders. Thus the original uncertainty about the outcome of one roll of a die is converted into an uncertainty about order. Because all the possible orders that can occur in an experiment of N rolls are equally likely, we can use (3.37) for the associated uncertainty S_N:

$$S_N = \ln \Omega = \ln \Big[\frac{N!}{\prod_j N_j!} \Big], \tag{3.158}$$

The right-hand side of (3.158) equals the total number of possible sequences.

To understand the form (3.158) suppose that we know that if we toss a coin four times, we will obtain 2 heads and 2 tails. What we don't know is the sequence. In Table 3.6 we show the six possible sequences. It is easy to see that this number is given by

$$M = \frac{N!}{\prod_j N_j} = \frac{4!}{2!\,2!} = 6. \tag{3.159}$$

H	H	T	T
H	T	H	T
H	T	T	H
T	T	H	H
T	H	T	H
T	H	H	T

Table 3.6: Possible sequences of tossing a coin four times such that two heads and two tails are obtained.

Now that we understand the form of S_N in (3.158), we can find the desired form of S. The uncertainty S_N in (3.158) is the uncertainty associated with all N rolls. The uncertainty associated with one roll is

$$S = \lim_{N\to\infty} \frac{1}{N} S_N = \lim_{N\to\infty} \frac{1}{N} \ln \Big[\frac{N!}{\prod_j N_j!} \Big] = \lim_{N\to\infty} \frac{1}{N} \big[\ln N! - \sum_j \ln N_j! \big]. \tag{3.160}$$

We can reduce (3.160) to a simpler form by using Stirling's approximation, $\ln N! \approx N \ln N - N$

for large N and substituting $N_j = NP_j$:

$$S = \lim_{N\to\infty} \frac{1}{N}\big[N\ln N - N - \sum_j (NP_j)\ln(NP_j) + \sum_j (NP_j)\big] \tag{3.161a}$$

$$= \lim_{N\to\infty} \frac{1}{N}\big[N\ln N - N - N\ln N\sum_j P_j - N\sum_j P_j\ln P_j + N\sum_j P_j\big] \tag{3.161b}$$

$$= -\sum_j P_j \ln P_j, \tag{3.161c}$$

where we used the fact that $\sum_j P_j = 1$.

3.12.2 Method of undetermined multipliers

Suppose that we want to maximize the function $f(x,y) = xy^2$ subject to the constraint that $x^2 + y^2 = 1$. One way would be to substitute $y^2 = 1 - x^2$ and maximize $f(x) = x(1-x^2)$. However, this approach works only if f can be reduced to a function of one variable. However we first consider this simple case as a way of introducing the general method of undetermined multipliers.

We wish to maximize $f(x,y)$ subject to the constraint that $g(x,y) = x^2 + y^2 - 1 = 0$. In the method of undetermined multipliers, this problem can be reduced to solving the equation

$$df - \lambda dg = 0, \tag{3.162}$$

where $df = y^2dx + 2xydy = 0$ at the maximum of f and $dg = 2xdx + 2ydy = 0$. If we substitute df and dg in (3.162), we have

$$(y^2 - 2\lambda x)dx + 2(xy - \lambda y)dy = 0. \tag{3.163}$$

We can choose $\lambda = y^2/2x$ so that the first term is zero. Because this term is zero, the second term must also be zero; that is, $x = \lambda = y^2/2x$, so $x = \pm y/\sqrt{2}$. Hence, from the constraint $g(x,y) = 0$, we obtain $x = \sqrt{1/3}$ and $\lambda = 2$.

In general, we wish to maximize the function $f(x_1, x_2, \ldots, x_N)$ subject to the constraints $g_j(x_1, x_2, \ldots, x_N) = 0$ where $j = 1, 2, \ldots, M$ with $M < N$. The maximum of f is given by

$$df = \sum_{i=1}^{N}\Big(\frac{\partial f}{\partial x_i}\Big)dx_i = 0, \tag{3.164}$$

and the constraint can be expressed as

$$dg = \sum_{i=1}^{N}\Big(\frac{\partial g_j}{\partial x_i}\Big)dx_i = 0. \tag{3.165}$$

As in our example, we can combine (3.164) and (3.165) and write $df - \sum_{j=1}^{M}\lambda_j dg_j = 0$ or

$$\sum_{i=1}^{N}\Big[\Big(\frac{\partial f}{\partial x_i}\Big) - \sum_{i=1}^{M}\lambda_j\Big(\frac{\partial g_j}{\partial x_i}\Big)\Big]dx_i = 0. \tag{3.166}$$

We are free to choose all M values of α_j such that the first M terms in the square brackets are zero. For the remaining $N-M$ terms, the dx_i can be independently varied because the constraints have been satisfied. Hence, the remaining terms in square brackets must be independently zero and we are left with $N - M$ equations of the form

$$\Big(\frac{\partial f}{\partial x_i}\Big) - \sum_{i=1}^{M} \lambda_j \Big(\frac{\partial g_j}{\partial x_i}\Big) = 0. \tag{3.167}$$

In Example 3.12 we were able to obtain the probabilities by reducing the uncertainty S to a function of a single variable P_1 and then maximizing $S(P_1)$. We now consider a more general problem where there are more outcomes, the case of a loaded die for which there are six outcomes. Suppose that we know that the average number of points on the face of a die if f. Then we wish to determine P_1, P_2, ..., P_6 subject to the constraints

$$\sum_{j=1}^{6} P_j = 1, \tag{3.168}$$

$$\sum_{j=1}^{6} jP_j = f. \tag{3.169}$$

For a perfect die $f = 3.5$. Equation (3.167) becomes

$$\sum_{j=1}^{6} \Big[(1 + \ln P_j) + \alpha + \beta j]dP_j = 0, \tag{3.170}$$

where we have used $dS = -\sum_{j=1}^{6} d(P_j \ln P_j) = -\sum_{j=1}^{6}(1 + \ln P_j)dP_j$; α and β are the undetermined (Lagrange) multipliers. We choose α and β so that the first two terms in the brackets (with $j = 1$ and $j = 2$ are independently zero. We write

$$\alpha = \ln P_2 - 2\ln P_1 - 1 \tag{3.171a}$$

$$\beta = \ln P_1 - \ln P_2. \tag{3.171b}$$

We can solve (3.171b) for $\ln P_2 = \ln P_1 - \beta$ and use (3.171a) to find $\ln P_1 = -1 - \alpha - \beta$ and use this result to write $P_2 = -1 - \alpha - \beta 2$. We can independently vary $dP_3, \ldots dP_6$ because the two constraints are satisfied by the values of P_1 and P_2. We let

$$\ln P_j = -1 - \alpha - j\beta, \tag{3.172}$$

or

$$P_j = e^{-1-\alpha} e^{-\beta j}. \tag{3.173}$$

We can eliminate the constant α by the normalization condition (3.168):

$$P_j = \frac{e^{-\beta j}}{\sum_j e^{-\beta j}}. \tag{3.174}$$

The constant β is determined by the constraint (3.47):

$$f = \frac{e^{-\beta} + 2e^{-\beta 2} + 3e^{-\beta 3} + 4e^{-\beta 4} + 5e^{-\beta 5} + 6e^{-\beta 6}}{e^{-\beta} + e^{-\beta 2} + e^{-\beta 3} + e^{-\beta 4} + e^{-\beta 5} + e^{-\beta 6}}. \tag{3.175}$$

In general, (3.175) must be solved numerically.

Problem 3.52. Show that the solution to (3.175) is $\beta = 0$ for $f = 7/2$, $\beta = +\infty$ for $f = 2$, $\beta = -\infty$ for $f = 6$, and $\beta = -0.1746$ for $f = 4$.

3.12.3 Derivation of the central limit theorem

To discuss the derivation of the central limit theorem, it is convenient to introduce the *characteristic function* $\phi(k)$ of the probability density $p(x)$. The main utility of the characteristic function is that it simplifies the analysis of the sums of independent random variables. We define $\phi(k)$ as the Fourier transform of $p(x)$:

$$\phi(k) = \overline{e^{ikx}} = \int_{-\infty}^{\infty} dx\, e^{ikx} p(x). \tag{3.176}$$

Because $p(x)$ is normalized, it follows that $\phi(k = 0) = 1$. The main property of the Fourier transform that we need is that if $\phi(k)$ is known, we can find $p(x)$ by calculating the inverse Fourier transform:

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, e^{-ikx} \phi(k). \tag{3.177}$$

Problem 3.53. Calculate the characteristic function of the Gaussian probability density.

One useful property of $\phi(k)$ is that its power series expansion yields the moments of $p(x)$:

$$\phi(k) = \sum_{n=0}^{\infty} \frac{k^n}{n!} \frac{d^n \phi(k)}{dk^n}\Big|_{k=0}, \tag{3.178}$$

$$= \overline{e^{ikx}} = \sum_{n=0}^{\infty} \frac{(ik)^n \overline{x^n}}{n!}. \tag{3.179}$$

By comparing coefficients of k^n in (3.178) and (3.179), we see that

$$\overline{x} = -i \frac{d\phi}{dk}\Big|_{k=0}. \tag{3.180}$$

In Problem 3.54 we show that

$$\overline{x^2} - \overline{x}^2 = -\frac{d^2}{dk^2} \ln \phi(k)\Big|_{k=0} \tag{3.181}$$

and that certain convenient combinations of the moments are related to the power series expansion of the logarithm of the characteristic function.

Problem 3.54. The characteristic function generates the *cumulants* C_m defined by

$$\ln \phi(k) = \sum_{m=1}^{\infty} \frac{(ik)^m}{m!} C_m. \tag{3.182}$$

Show that the cumulants are combinations of the moments of x and are given by

$$C_1 = \overline{x} \tag{3.183a}$$
$$C_2 = \sigma^2 = \overline{x^2} - \overline{x}^2 \tag{3.183b}$$
$$C_3 = \overline{x^3} - 3\,\overline{x^2}\,\overline{x} + 2\,\overline{x}^3 \tag{3.183c}$$
$$C_4 = \overline{x^4} - 4\,\overline{x^3}\,\overline{x} - 3\,\overline{x^2}^2 + 12\,\overline{x^2}\,\overline{x}^2 - 6\,\overline{x}^4. \tag{3.183d}$$

Now let us consider the properties of the characteristic function for the sums of independent variables. For example, let $p_1(x)$ be the probability density for the weight x of adult males and let $p_2(y)$ be the probability density for the weight of adult females. If we assume that people marry one another independently of weight, what is the probability density $p(z)$ for the weight z of an adult couple? We have that

$$z = x + y. \tag{3.184}$$

How do the probability densities combine? The answer is

$$p(z) = \int dx\,dy\,p_1(x)p_2(y)\,\delta(z - x - y). \tag{3.185}$$

The integral in (3.185) represents all the possible ways of obtaining the combined weight z as determined by the probability density $p_1(x)p_2(y)$ for the combination of x and y that sums to z. The form (3.185) of the integrand is known as a *convolution*. An important property of a convolution is that its Fourier transform is a simple product. We have

$$\phi_z(k) = \int dz\,e^{ikz} p(z) \tag{3.186a}$$
$$= \int dz \int dx \int dy\,e^{ikz} p_1(x)p_2(y)\,\delta(z - x - y) \tag{3.186b}$$
$$= \int dx\,e^{ikx} p_1(x) \int dy\,e^{iky} p_2(y) \tag{3.186c}$$
$$= \phi_1(k)\phi_2(k). \tag{3.186d}$$

It is straightforward to generalize this result to a sum of N random variables. We write

$$z = x_1 + x_2 + \ldots + x_N. \tag{3.187}$$

Then

$$\phi_z(k) = \prod_{i=1}^{N} \phi_i(k). \tag{3.188}$$

That is, *for independent variables the characteristic function of the sum is the product of the individual characteristic functions.* If we take the logarithm of both sides of (3.188), we obtain

$$\ln \phi_z(k) = \sum_{i=1}^{N} \ln \phi_i(k). \tag{3.189}$$

Each side of (3.189) can be expanded as a power series and compared order by order in powers of ik. The result is that when random variables are added, their associated cumulants also add. That is, the nth order cumulants satisfy the relation:

$$C_n^z = C_n^1 + C_n^2 + \ldots + C_n^N. \tag{3.190}$$

We conclude see that if the random variables x_i are independent (uncorrelated), their cumulants and in particular, their variances, add.

If we denote the mean and standard deviation of the weight of an adult male as $\overline{x}$ and σ respectively, then from (3.183a) and (3.190) we find that the mean weight of N adult males is given by $N\overline{x}$. Similarly from (3.183b) we see that the standard deviation of the weight of N adult males is given by $\sigma_N^2 = N\sigma^2$, or $\sigma_N = \sqrt{N}\sigma$. Hence, we find the now familiar result that the sum of N random variables scales as N while the standard deviation scales as $\sqrt{N}$.

We are now in a position to derive the central limit theorem. Let $x_1, x_2, \ldots, x_N$ be N mutually independent variables. For simplicity, we assume that each variable has the same probability density $p(x)$. The only condition is that the variance σ_x^2 of the probability density $p(x)$ must be finite. For simplicity, we make the additional assumption that $\overline{x} = 0$, a condition that always can be satisfied by measuring x from its mean. The central limit theorem states that the sum S has the probability density

$$p(S) = \frac{1}{\sqrt{2\pi N {\sigma_x}^2}}\, e^{-S^2/2N{\sigma_x}^2} \tag{3.191}$$

From (3.183b) we see that $\overline{S^2} = N\sigma_x^2$, and hence the variance of S grows linearly with N. However, the distribution of the values of the arithmetic mean S/N becomes narrower with increasing N:

$$\overline{\left(\frac{x_1 + x_2 + \ldots x_N}{N}\right)^2} = \frac{N\sigma_x^2}{N^2} = \frac{\sigma_x^2}{N}. \tag{3.192}$$

From (3.192) we see that it is useful to define a scaled sum:

$$z = \frac{1}{\sqrt{N}}(x_1 + x_2 + \ldots + x_N), \tag{3.193}$$

and to write the central limit theorem in the form

$$p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2}. \tag{3.194}$$

To obtain the result (3.194), we write the characteristic function of z as

$$\phi_z(k) = \int dx\, e^{ikz} \int dx_1 \int dx_2 \cdots \int dx_N\, \delta\big(z - [\frac{x_1 + x_2 + \ldots + x_N}{N^{1/2}}]\big) \tag{3.195a}$$

$$\times\ p(x_1)\, p(x_2) \ldots p(x_N) \tag{3.195b}$$

$$= \int dx_1 \int dx_2 \ldots \int dx_N\, e^{ik(x_1+x_2+\ldots x_N)/N^{1/2}} p(x_1)\, p(x_2) \ldots p(x_N) \tag{3.195c}$$

$$= \phi(\frac{k}{N^{1/2}})^N. \tag{3.195d}$$

We next take the logarithm of both sides of (3.195d) and expand the right-hand side in powers of k to find

$$\ln \phi_z(k) = \sum_{m=2}^{\infty} \frac{(ik)^m}{m!} N^{1-m/2} C_m. \tag{3.196}$$

The $m = 1$ term does not contribute in (3.196) because we have assumed that $\overline{x} = 0$. More importantly, note that as $N \to \infty$, the higher-order terms are suppressed so that

$$\ln \phi_z(k) \to -\frac{k^2}{2} C_2, \tag{3.197}$$

or

$$\phi_z(k) \to e^{-k^2\sigma^2/2} + \ldots \tag{3.198}$$

Because the inverse Fourier transform of a Gaussian is also a Gaussian, we find that

$$p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2}. \tag{3.199}$$

The leading correction to $\phi(k)$ in (3.199) gives rise to a term of order $N^{-1/2}$, and therefore does not contribute in the limit $N \to \infty$.

The only requirements for the applicability of the central limit theorem are that the various x_i be statistically independent and that the second moment of $p(x)$ exists. Not all probabilities satisfy this latter requirement as demonstrated by the Lorentz distribution (see Problem 3.43). It is not necessary that all the x_i have the same distribution.

Vocabulary

sample space, events, outcome

uncertainty, principle of least bias or maximum uncertainty

probability distribution, probability density

mean value, moments, variance, standard deviation

conditional probability, Bayes' theorem

binomial distribution, Gaussian distribution, Poisson distribution random walk, random additive processes, central limit theorem

Stirling's approximation

Monte Carlo sampling

Rare or extreme events

Notation

probability distribution $P(i)$, mean value $\overline{f(x)}$, variance $\overline{\Delta x^2}$, standard deviation σ

conditional probability $P(A|B)$, probability density $p(x)$

Additional problems

Listing of inline problems.

Problem 3.55. In Figure 3.9 we show a square lattice of 16^2 sites each of which is occupied with probability p. Estimate the probability that a site in the lattice is occupied.

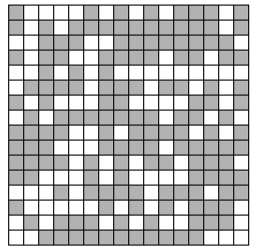

Figure 3.9: Representation of a square lattice of 16×16 sites. The sites are represented by squares and are either occupied (shaded) with probability p or are empty (white) with probability $1 - p$.

Problem 3.56. Three coins (in a fountain)

Three coins are tossed in succession. Assume that the simple events are equiprobable. Find the probabilities of the following:

(a) the first coin is heads;

(b) exactly two heads have occurred;

(c) not more than two heads have occurred.

Problem 3.57. Fallacious reasoning

A student tries to solve Problem 3.13 by using the following reasoning. The probability of a double six is 1/36. Hence the probability of finding at least one double six in 24 throws is 24/36. What is wrong with this reasoning? If you have trouble understanding the error in this reason, try solving the problem of finding the probability of at least one double six in two throws of a pair of dice. What are the possible outcomes? Is each outcome equally probable?

Problem 3.58. d'Alembert's fallacious reasoning

In two tosses of a single coin, what is the probability that heads will appear at least once? Use the rules of probability to show that the answer is $\frac{3}{4}$. However, d'Alembert, a distinguished French mathematician of the eighteenth century, reasoned that there are only 3 possible outcomes: heads on the first throw, heads on the second throw, and no heads at all. The first two of these three outcomes is favorable. Therefore the probability that heads will appear at least once is $\frac{2}{3}$. What is the fallacy in his reasoning? Even eminent mathematicians (and physicists) have been lead astray by the subtle nature of probability.

Problem 3.59. False positives

A diagnostic test for the presence of the AIDS virus has a probability of 0.005 of producing a false positive. If 200 patients are tested at a clinic, what is the probability that at least one false positive occurs?

Problem 3.60. Number of fish in a pond

A farmer wants to estimate how many fish are in her pond. She takes out 200 fish and tags them and returns them to the pond. After sufficient time to allow the tagged fish to mix with the others, she removes 250 fish at random and finds that 25 of them are tagged. Estimate the number of fish in the pond.

	x_i, y_i		x_i, y_i
1	0.984, 0.246	6	0.637, 0.581
2	0.860, 0.132	7	0.779, 0.218
3	0.316, 0.028	8	0.276, 0.238
4	0.523, 0.542	9	0.081, 0.484
5	0.349, 0.623	10	0.289, 0.032

Table 3.7: A sequence of ten random pairs of numbers.

Problem 3.61. Estimating the area of a pond

A farmer owns a field that is 10 m × 10 m. In the midst of this field is a pond of unknown area. Suppose that the farmer is able to throw 100 stones at random into the field and finds that 40 of the stones make a splash. How can the farmer use this information to estimate the area of the pond?

Problem 3.62. Monte Carlo integration

Consider the ten pairs of numbers, (x_i, y_i), given in Table 3.7. The numbers are all are in the range $0 < x_i, y_i \le 1$. Imagine that these numbers were generated by counting the clicks generated by a Geiger counter of radioactive decays, and hence they can be considered to be a part of a sequence of random numbers. Use this sequence to estimate the magnitude of the integral

$$F = \int_0^1 dx\, \sqrt{(1 - x^2)}. \tag{3.200}$$

If you have been successful in estimating the integral in this way, you have found a simple version of a general method known as *Monte Carlo integration*.[12] An applet for estimating integrals by Monte Carlo integration can be found at <stp.clarku.edu/simulations/estimate>.

Problem 3.63. Bullseyes

A person playing darts hits a bullseye 20% of the time on the average. Why is the probability of b bullseyes in N attempts a binomial distribution? What are the values of p and q? Find the probability that the person hits a bullseye

(a) once in five throws;

(b) twice in ten throws. Why are these probabilities not identical?

Problem 3.64. There are 10 children in a given family. Assuming that a boy is as likely to be born as a girl, find the probability of the family having

(a) 5 boys and 5 girls;

(b) 3 boys and 7 girls.

Problem 3.65. What is the probability that five children produced by the same couple will consist of the following:

[12] Monte Carlo methods were first developed to estimate integrals that could not be performed by other ways.

(a) three sons and two daughters?

(b) alternating sexes?

(c) alternating sexes starting with a son?

(d) all daughters? Assume that the probability of giving birth to a boy and a girl is the same.

Problem 3.66. Probability in baseball

A good hitter in baseball has a batting average of 300 or more, which means that the hitter will be successful 3 times out of 10 tries on the average. Assume that on average a hitter gets three hits for each 10 times at bat and that he has 4 times at bat per game.

(a) What is the probability that he gets zero hits in one game?

(b) What is the probability that he will get two hits or less in a three game series?

(c) What is the probability that he will get five or more hits in a three game series? Baseball fans might want to think about the significance of "slumps" and "streaks" in baseball.

Problem 3.67. Playoff winners

In the World Series in baseball and in the playoffs in the National Basketball Association and the National Hockey Association, the winner is determined by a best of seven series. That is, the first team that wins four games wins the series and is the champion. Do a simple statistical calculation assuming that the two teams are evenly matched and the winner of any game might as well be determined by a coin flip and show that a seven game series should occur 31.25% of the time. What is the probability that the series lasts n games? More information can be found at <www.mste.uiuc.edu/hill/ev/seriesprob.html> and at <www.aip.org/isns/reports/2003/080.html>.

Problem 3.68. Galton board

The Galton board (named after Francis Galton (1822–1911)), is a triangular array of pegs. The rows are numbered $0, 1, \ldots$ from the top row down such that row n has $n+1$ pegs. Suppose that a ball is dropped from above the top peg. Each time the ball hits a peg, it bounces to the right with probability p and to the left with probability $1-p$, independently from peg to peg. Suppose that N balls are dropped successively such that the balls do not encounter one another. How will the balls be distributed at the bottom of the board? Links to applets that simulate the Galton board can be found in the references.

Problem 3.69. The birthday problem

What if somebody offered to bet that at least two people in your physics class had the same birthday? Would you take the bet?

(a) What are the chances that at least two people in your class have the same birthday? Assume that the number of students is 25.

(b) What are the chances that at least one other person in your class has the same birthday as you? Explain why the chances are less in the second case.

Problem 3.70. A random walk down Wall Street

Many analysts attempt to select stocks by looking for correlations in the stock market as a whole or for patterns for particular companies. Such an analysis is based on the belief that there are repetitive patterns in stock prices. To understand one reason for the persistence of this belief do the following experiment. Construct a stock chart (a plot of stock price versus time) showing the movements of a hypothetical stock initially selling at $50 per share. On each successive day the closing stock price is determined by the flip of a coin. If the coin toss is a head, the stock closes 1/2 point ($0.50) higher than the preceding close. If the toss is a tail, the price is down by 1/2 point. Construct the stock chart for a long enough time to see "cycles" and other patterns appear. The moral of the charts is that a sequence of numbers produced in this manner is identical to a random walk, yet the sequence frequently appears to be correlated.

Problem 3.71. Suppose that a random walker takes N steps of unit length with probability p of a step to the right. The displacement m of the walker from the origin is given by $m = n - n'$, where n is the number of steps to the right and n' is the number of steps to the left. Show that $\overline{m} = (p - q)N$ and $\sigma_m^2 = \overline{(m - \overline{m})^2} = 4Npq$.

Problem 3.72. The result (3.79) for $\overline{(\Delta M)^2}$ differs by a factor of four from the result for σ_n^2 in (3.101). Why? Compare (3.79) to the result of Problem 3.37.

Problem 3.73. Size of the airways in the mammalian lung

The geometry of branched structures such as blood vessels or airways are important factors in determining the efficiency of physiological processes. The airways of the bronchial tree of mammalian lungs branch at regular intervals with a systematic reduction of their diameter. In the human lung the conducting airway tree ends at about 2^{17} divisions.

In this problem we consider a simplified model of how the average diameter of the airways in the mammalian lung change down the bronchial tree. Assume that the diameter at the $n = 0$ branch (the trachea) is one. Suppose that the airways branch into two nearly equal parts of width p and q. (In this case p and q are not probabilities and hence $p + q$ is not necessarily unity.) After the first division, the average diameter is $L(1) = (p + q)/2$; after two divisions (generations), the average diameter is $L(2) = (p^2 + 2pq + q^2)/4$. Show that the average diameter of the bronchial tube after n generations is given by

$$L(n) = (p + q)^n/2^n = e^{-n/n_0}, \tag{3.201}$$

where $n_0^{-1} = |\ln[(p + q)/2]|$. What is the behavior of $L(n)$ for $p = q = 1/3$? This exponential behavior of $L(n)$ is observed for $n \leq 10$. To explain the behavior of $L(n)$ for larger N, it is necessary to assume that the division is not precisely given by p and q, but is only p and q on the average.

Problem 3.74. Watching a drunkard

A random walker is observed to take a total of N steps, n of which are to the right.

(a) Suppose that a curious observer finds that on ten successive nights the walker takes $N = 20$ steps and that the values of n are given successively by 14, 13, 11, 12, 11, 12, 16, 16, 14, 8. Compute $\overline{n}$, $\overline{n^2}$, and σ_n. Use this information to estimate p. If your reasoning gives different values for p, which estimate is likely to be the most accurate?

(b) Suppose that on another ten successive nights the same walker takes $N = 100$ steps and that the values of n are given by 58, 69, 71, 58, 63, 53, 64, 66, 65, 50. Compute the same quantities as in part (a) and estimate p. How does the ratio of σ_n to $\overline{n}$ compare for the two values of N? Explain your results.

(c) Compute $\overline{m}$ and σ_m, where $m = n - n'$ is the net displacement of the walker. This problem inspired an article by Zia and Schmittmann (see the references).

Problem 3.75. Alternative derivation of the Gaussian distribution

In Section 3.7 we evaluated the derivatives of $P(n)$ to determine the parameters A, B, and $\tilde{n}$ in (3.120). Another way to determine these parameters is to assume that the binomial distribution can be approximated by a Gaussian and require that the first several moments of the Gaussian and binomial distribution be equal. We write

$$P(n) = Ae^{-\frac{1}{2}B(n-\tilde{n})^2}, \tag{3.202}$$

and require that

$$\int_0^N P(n)\, dn = 1. \tag{3.203}$$

Because $P(n)$ depends on the difference $n - \tilde{n}$, it is convenient to change the variable of integration in (3.203) to $x = n - \tilde{n}$. We have

$$\int_{-\tilde{n}}^{N-\tilde{n}} P(x)\, dx = 1, \tag{3.204}$$

where

$$P(x) = Ae^{-\frac{1}{2}Bx^2}. \tag{3.205}$$

In the limit of large N, we can extend the upper and lower limits of integration in (3.204) and write

$$\int_{-\infty}^{\infty} P(x)\, dx = 1, \tag{3.206}$$

The first moment of $P(n)$ is given by

$$\overline{n} = \int_0^N nP(n)\, dn = pN. \tag{3.207}$$

Make a change of variables and show that

$$\int_{-\infty}^{\infty} xP(x)\, dx = \overline{n} - \tilde{n}. \tag{3.208}$$

Because the integral in (3.208) is zero, we can conclude that $\tilde{n} = \overline{n}$. We also have that

$$\overline{(n - \overline{n})^2} = \int_0^N (n - \overline{n})^2\, P(n)\, dn = pqN. \tag{3.209}$$

Do the integrals in (3.209) and (3.206) (see (A.17) and (A.21)) and confirm that the values of B and A are given by (3.126) and (3.128), respectively. The generality of the arguments leading to the Gaussian distribution suggests that it occurs frequently in probability when large numbers are involved. Note that the Gaussian distribution is characterized completely by its mean value and its width.

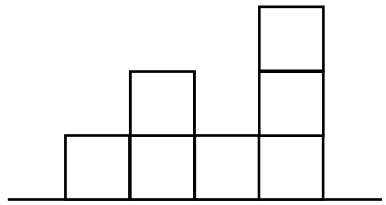

Figure 3.10: Example of a castle wall as explained in Problem 3.76.

Problem 3.76. Consider a two-dimensional 'wall constructed from N squares as shown in Figure 3.10. The base row of the cluster must be continuous, but higher rows can have gaps. Each column must be continuous and self-supporting. Determine the total number W_N of different N-site clusters, that is, the number of possible arrangements of N squares consistent with the above rules. Assume that the squares are identical.

Problem 3.77. First passage time

Suppose that a one-dimensional unbiased random walker starts out at the origin $x = 0$ at $t = 0$. How many steps will it take for the walker to reach a site at $x = 4$? This quantity, known as the *first passage time*, is a random variable because it is different for different possible realizations of the walk. Possible quantities of interest are the probability distribution of the first passage time and the mean first passage time, τ. Write a computer program to estimate $\tau(x)$ and then determine its analytical dependence on x. Why is it more difficult to estimate τ for $x = 8$ than for $x = 4$?

Problem 3.78. Heads you win

Two people take turns tossing a coin. The first person to obtain heads is the winner. Find the probabilities of the following events:

(a) the game terminates at the fourth toss;

(b) the first player wins the game;

(c) the second player wins the game.

*__Problem 3.79.__ Range of validity of the Gaussian distribution

How good is the Gaussian distribution as an approximation to the binomial distribution as a function of N? To determine the validity of the Gaussian distribution, consider the next two terms in the power series expansion of $\ln P(n)$:

$$\frac{1}{3!}(n-\tilde{n})^3 C + \frac{1}{4!)}(n-\tilde{n})^4 D, \tag{3.210}$$

with $C = d^3 \ln P(n)/d^3 n$ and $D = d^4 \ln P(n)/d^4 n$ evaluated at $n = \tilde{n}$.

(a) Show that $C = 0$ if $p = q$. Calculate D for $p = q$ and estimate the order of magnitude of the first nonzero correction. Compare this correction to the magnitude of the first nonzero term in $\ln P(n)$ (see (3.116)) and determine the conditions for which the terms beyond $(n-\tilde{n})^2$ can be neglected.

(b) Define the error as

$$E(n) = 1 - \frac{\text{Binomial}(n)}{\text{Gaussian}(n)} \tag{3.211}$$

Plot $E(n)$ versus n and determine the approximate width of $E(n)$.

(c) Show that if N is sufficiently large and neither p nor q is too small, the Gaussian distribution is a good approximation for n near the maximum of $P(n)$. Because $P(n)$ is very small for large $(n - \overline{n})$, the error in the Gaussian approximation for larger n is negligible.

Problem 3.80. Two-dimensional random walk

Consider a random walk on a two-dimensional square lattice where the walker has an equal probability of taking a step to one of four possible directions, north, south, east, or west. Use the central limit theorem to find the probability that the walker is a distance r to $r + dr$ from the origin, where $r^2 = x^2 + y^2$ and r is the distance from the origin after N steps. There is no need to do an explicit calculation.

Problem 3.81. Continuum model of a random walk

One of the first continuum models of a random walk is due to Rayleigh (1919). In the Rayleigh model the length a of each step is a random variable with probability density $p(a)$ and the direction of each step is random. For simplicity consider a walk in two dimensions and choose $p(a)$ so that each step has unit length. Then at each step the walker takes a step of unit length at a random angle. Use the central limit theorem to find the asymptotic form of $p(r, N)\, dr$, the probability that the walker is in the range r to $r + dr$, where r is the distance from the origin after N steps.

Problem 3.82. Suppose there are three boxes each with two balls. The first box has two green balls, the second box has one green and one red ball, and the third box has two red balls. Suppose you choose a box at random and find one green ball. What is the probability that the other ball is green?

Problem 3.83. Open a telephone directory to an random page and make a list corresponding to the last digit n of the first 100 telephone numbers. Find the probability $P(n)$ that the number n appears. Plot $P(n)$ as a function of n and describe its n-dependence. Do you expect that $P(n)$ should be approximately uniform?

***Problem 3.84.** Model of a porous rock

A simple model of a porous rock can be imagined by placing a series of overlapping spheres at random into a container of fixed volume V. The spheres represent the rock and the space between the spheres represents the pores. If we write the volume of the sphere as v, it can be shown the fraction of the space between the spheres or the *porosity* ϕ is $\phi = \exp(-Nv/V)$, where N is the number of spheres. For simplicity, consider a two-dimensional system, and write a program to place disks of diameter unity into a square box. The disks can overlap. Divide the box into square cells each of which has an edge length equal to the diameter of the disks. Find the probability of having 0, 1, 2, or 3 disks in a cell for $\phi = 0.03$, 0.1, and 0.5.

***Problem 3.85.** Benford's law

Do a search of the Web and find a site that lists the populations of various cities in the world (not necessarily the largest ones) or the cities of your state or region. The quantity of interest is the first digit of each population. Alternatively, scan the first page of your local newspaper and record the first digit of each of the numbers you find. (The first digit of a number such as 0.00123 is 1.) What is the probability $P(n)$ that the *first* digit is n, where $n = 1, \ldots, 9$? Do you think that $P(n)$ will be the same for all n?

It turns out that the form of the probability $P(n)$ is given by

$$P(n) = \log_{10}\big(1 + \frac{1}{n}\big). \tag{3.212}$$

The distribution (3.212) is known as *Benford's law* and is named after Frank Benford, a physicist. It implies that for certain data sets, the first digit is distributed in a predictable pattern with a higher percentage of the numbers beginning with the digit 1. What are the numerical values of $P(n)$ for the different values of n? Is $P(n)$ normalized? Suggest a hypothesis for the nonuniform nature of the Benford distribution.

Accounting data is one of the many types of data that is expected to follow the Benford distribution. It has been found that artificial data sets do not have first digit patterns that follow the Benford distribution. Hence, the more an observed digit pattern deviates from the expected Benford distribution, the more likely the data set is suspect. Tax returns have been checked in this way.

The frequencies of the first digit of 2000 numerical answers to problems given in the back of four physics and mathematics textbooks have been tabulated and found to be distributed in a way consistent with Benford's law. Benford's Law is also expected to hold for answers to homework problems (see James R. Huddle, "A note on Benford's law," *Math. Comput. Educ.* **31**, 66 (1997)). The nature of Benford's law is discussed by T. P. Hill, "The first digit phenomenon," *Am. Sci.* **86**, 358–363 (1998).

***Problem 3.86.** Ask several of your friends to flip a coin 200 times and record the results or pretend to flip a coin and fake the results. Can you tell which of your friends faked the results? Hint: What is the probability that a sequence of six heads in a row will occur? Can you suggest any other statistical tests?

***Problem 3.87.** Zipf's law

Analyze a text and do a ranking of the word frequencies. The word with rank r is the rth word when the words of the text are listed with decreasing frequency. Make a log-log plot of word frequency f versus word rank r. The relation between word rank and word frequency was first stated by George Kingsley Zipf (1902–1950). This relation states that to a very good approximation for a given text

$$f \sim \frac{1}{r\ln(1.78R)}, \tag{3.213}$$

where R is the number of different words. Note the inverse power law behavior. The relation (3.213) is known as *Zipf's law*. The top 20 words in an analysis of a 1.6 MB collection of 423 short Time magazine articles (245,412 term occurrences) are given in Table 3.8.

***Problem 3.88.** Time of response to emails

If you receive an email, how long does it take for you to respond to it? If you keep a record of your received and sent mail, you can analyze the distribution of your response times – the number of days (or hours) between receiving an email from someone and then replying.

1	the	15861	11	his	1839
2	of	7239	12	is	1810
3	to	6331	13	he	1700
4	a	5878	14	as	1581
5	and	5614	15	on	1551
6	in	5294	16	by	1467
7	that	2507	17	at	1333
8	for	2228	18	it	1290
9	was	2149	19	from	1228
10	with	1839	20	but	1138

Table 3.8: Ranking of the top 20 words.

It turns out that the time it takes people to reply to emails can be described by a power law; that is, $P(\tau) \sim \tau^{-a}$ with $a \approx 1$. Oliveira and Barabási have shown that the response times of Einstein and Darwin to letters can also be described by a power law, but with an exponent $a \approx 3/2$ (see J. G. Oliveira and A.-L. Barabási, "Darwin and Einstein correspondence patterns," Nature **437**, 1251 (2005). Their results suggest that there is an universal pattern for human behavior in response to correspondence. What is the implication of a power law response?

***Problem 3.89.** A doctor has two drugs, A and B, which she can prescribe to patients with a certain illness. The drugs have been rated in terms of their effectiveness on a scale of 1 to 6, with 1 being the least effective and 6 being the most effective. Drug A is uniformly effective with a value of 3. The effectiveness of drug B is variable and 54% of the time it scores a value of 1, and 46% of the time it scores a value of 5. The doctor wishes to provide her patients with the best possible care and asks her statistician friend which drug has the highest probability of being the most effective. The statistician says, "It is clear that drug A is the most effective drug 54% of the time. Thus drug A is your best bet."

Later a new drug C becomes available. Studies show that on the scale of 1 to 6, 22% of the time this drug scores a 6, 22% of the time it scores a 4, and 56% of the time it scores a 2. The doctor, again wishing to provide her patients with the best possible care, goes back to her statistician friend and asks him which drug has the highest probability of being the most effective. The statistician says, "Because there is this new drug C on the market, your best bet is now drug B, and drug A is your worst bet." Show that the statistician is right.

Problem 3.90. Three cards are in a hat. One card is white on both sides, the second is white on one side and red on the other, and the third is red on both sides. The dealer shuffles the cards, takes one out and places it flat on the table. The side showing is red. The dealer now says, "Obviously this card is not the white-white card. It must be either the red-white card or the red-red card. I will bet even money that the other side is red." Is this bet fair?

Problem 3.91. Will an asteroid impact the Earth?

Estimate the probability that an asteroid will impact the Earth and cause major damage. Does it make sense for society to take steps now to guard itself against such an occurrence?

***Problem 3.92.** Response to rare events

The likelihood of the breakdown of the levees near New Orleans was well known before their occurrence on August 30, 2005. Discuss the various reasons why the decision was made not to strengthen the levees. Relevant issues include the ability of people to think about the probability of rare events and the large amount of money needed to strengthen the levees to withstand such an event.

***Problem 3.93.** Science and society issues

Does capital punishment deter murder? Are vegetarians more likely to have daughters? Does it make sense to talk about a "hot hand" in basketball? Are the digits of π random? See <chance.dartmouth.edu/chancewiki/> and <www.dartmouth.edu/~chance/> and read about other interesting issues involving probability and statistics.

Suggestions for further reading

Vinay Ambegaokar, *Reasoning About Luck*, Cambridge University Press (1996). A book developed for a course intended for non-science majors. An excellent introduction to statistical reasoning and its uses in physics.

Peter L. Bernstein, *Against the Gods: The Remarkable Story of Risk*, John Wiley & Sons (1996). The author is a successful investor and an excellent writer. The book includes an excellent summary of the history of probability.

David S. Betts and Roy E. Turner, *Introductory Statistical Mechanics*, Addison-Wesley (1992). Section 3.4 is based in part on Chapter 3 of this text.

Jean-Phillippe Bouchaud and Marc Potters, *Theory of Financial Risks*, Cambridge University Press (2000). This book by two physicists is an example of the application of concepts in probability and statistical mechanics to finance. Although the treatment is at the graduate level and assumes some background in finance, it is recommended for students who might be interested in the overlap of physics, finance, and economics. Also see J. Doyne Farmer, Martin Shubik, and Eric Smith, "Is economics the next physical science?," Phys. Today **58** (9), 37–42 (2005). A related book on the rare events is by Nassim Nicholas Taleb, *The Black Swan: The Impact of the Highly Improbable*, Random House (2007).

The <www.dartmouth.edu/~chance/> encourages its users to apply statistics to everyday events.

Giulio D'Agostini, "Teaching statistics in the physics curriculum: Unifying and clarifying role of subjective probability," Am. J. Phys. **67**, 1260–1268 (1999). The author, whose main research interest is in particle physics, discusses subjective probability and Bayes' theorem. Section 3.4 is based in part on this article.

See <www.math.uah.edu/stat/objects/> for a simulation of the Galton board.

Gene F. Mazenko, *Equilibrium Statistical Mechanics*, John Wiley & Sons (2000). Sections 1.7 and 1.8 of this graduate level text discuss the functional form of the missing information.

Elliott W. Montroll and Michael F. Shlesinger, "On the wonderful world of random walks," in *Studies in Statistical Mechanics*, Vol. XI: Nonequilibrium Phenomena II, J. L. Lebowitz and E. W. Montroll, editors, North-Holland (1984).

Elliott W. Montroll and Wade W. Badger, *Introduction to Quantitative Aspects of Social Phenomena*, Gordon and Breach (1974). The applications of probability that are discussed include traffic flow, income distributions, floods, and the stock market.

Richard Perline, "Zipf's law, the central limit theorem, and the random division of the unit interval," Phys. Rev. E **54**, 220–223 (1996).

S. Redner, "Random multiplicative processes: An elementary tutorial," Am. J. Phys. **58**, 267–273 (1990).

Charles Ruhla, *The Physics of Chance*, Oxford University Press (1992).

B. Schmittmann and R. K. P. Zia, "'Weather' records: Musings on cold days after a long hot Indian summer," Am. J. Phys. **67**, 1269–1276 (1999). A relatively simple introduction to the statistics of extreme values. Suppose that somebody breaks the record for the 100 meter dash. How long do records typically survive before they are broken?

Kyle Siegrist at the University of Alabama in Huntsville has developed many applets to illustrate concepts in probability and statistics. See <www.math.uah.edu/stat/> and follow the link to Bernoulli processes.

G. Troll and P. beim Graben, "Zipf's law is not a consequence of the central limit theorem," Phys. Rev. E **57**, 1347–1355 (1998).

Charles A. Whitney, *Random Processes in Physical Systems: An Introduction to Probability-Based Computer Simulations*, John Wiley & Sons (1990).

A good discussion by Eliezer Yudkowsky of the intuitive basis of Bayesian reasoning can be found at <yudkowsky.net/bayes/bayes.html>.

R. K. P. Zia and B. Schmittmann, "Watching a drunkard for 10 nights: A study of distributions of variances," Am. J. Phys. **71**, 859–865 (2003). See Problem 3.74.

The outcome of tossing a coin is not really random. See Ivars Peterson, "Heads or tails?," Science News Online, www.sciencenews.org/articles/20040228/mathtrek.asp and Erica Klarreich, "Toss out the toss-up: Bias in heads-or-tails, Science News **165** (9), 131 (2004), <http://www.sciencenews.org/articles/20040228/fob2.asp>. Some of the original publications include Joseph Ford, "How random is a coin toss?," Phys. Today **36** (4), 40–47 (1983); Joseph B. Keller, "The probability of heads," Am. Math. Monthly **93**, 191–197 (1986); and Vladimir Z. Vulovic and Richard E. Prange, "Randomness of a true coin toss," Phys. Rev. A **33**, 576-582 (1986).