

Chapter 3

Concepts of Probability

©2009 by Harvey Gould and Jan Tobochnik
13 August 2009

We introduce the basic concepts of probability and apply them to simple physical systems and everyday life. We will discuss the universal nature of the central limit theorem and the Gaussian distribution for the sum of a large number of random variables. Because of the importance of probability in many contexts, our discussion goes beyond what we will need for the applications of statistical mechanics that we will discuss in later chapters.

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 need to introduce some ideas from probability.

We all use 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. You can make a rough estimate of the probability of being hit by a car. It must be less than 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 many real world phenomena.

However, our intuition only takes us so far. Consider airplane travel. Is it safe to fly? Suppose that there is one chance in five million of a plane crashing on a given flight and that there are about 50,000 flights a day. Then every 100 days or so there would be a reasonable likelihood of

a plane crash somewhere in the world. This estimate is in rough accord with what we know. For a given flight, your chances of crashing are approximately one part in 5×10^6 , and if you fly ten 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 50,000 years. In this case you would take 500,000 flights, and it would be much more risky to fly if you wished to live your full 50,000 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 of success. 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 probability 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, solve Problems 3.1–3.6 before reading the rest of this chapter. Then in Problem 3.7 formulate the laws of probability 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 person 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 person hits 16 free throws out of 25 attempts. What is the probability that this person will make a free throw on the next attempt? \square

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.¹ 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? \square

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 . \square

Problem 3.7. Rules of probability

Based on your solutions to Problems 3.1–3.6, state the rules of probability as you understand them at this time. \square

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 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 prize was \$40? \square

¹The 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).

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 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) In one version of the lottery the player chooses six numbers from 1 through 49. The player wins only if there is an exact match with the numbers that are randomly generated. Suppose that someone gives you a dollar to play the lottery. What sequence of six numbers between 1 and 49 would you choose? Are some choices better than others?
- (c) Suppose you toss a coin 6 times and obtain heads each time. Estimate the probability that you will obtain heads on your seventh toss. Now imagine tossing the coin 60 times, and obtaining heads each time. What do you think would happen on the next 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. In each case discuss how you would estimate the probability. □

3.2 The Rules of Probability

We now summarize the basic rules and ideas of probability.² 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. 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 roll of a die yields one of six possible events.

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

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

²In 1933 the Russian mathematician A. N. Kolmogorov formulated a complete set of axioms for the mathematical definition of probability.

and

$$\sum_i P(i) = 1. \quad (3.2)$$

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

Example 3.1. Sample space of a die

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. \diamond

The rules of probability will be summarized 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.

A common interpretation of probability is based on *symmetry*. Suppose that we have a two-sided coin that shows heads and tails. There are two possible mutually exclusive outcomes, and if the coin is fair, 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 of an outcome *a posteriori*, that is, by the observation of the outcome of many throws. As we will see other kinds of information in addition to symmetry arguments can be used to estimate probabilities.

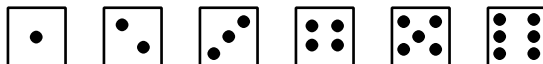


Figure 3.1: The six possible outcomes of the toss of a die.

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

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

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 $1 - P(i)$.

³Is the outcome of a coin toss really random? The outcome of a coin flip is deterministic, but the outcome depends sensitively on the initial conditions, which we don't know precisely. See the references at the end of the chapter.

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 $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$. \diamond

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

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

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

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

Another simple rule concerns 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

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

Events are independent if the occurrence of one event does not affect the probability of 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 $P(6^+) = \frac{1}{10}$, 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 $P(\text{female})P(6^+) = \frac{1}{2} \times \frac{1}{10} = \frac{1}{20}$. This same probability calculation 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. \square

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

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

\diamond

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

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

◇

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

Solution. We know that

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

There are four possible outcomes (6, 6), (6, not 6), (not 6, 6), (not 6, not 6) with the respective probabilities

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

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

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

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

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

$$= \frac{1}{36} + \frac{5}{36} + \frac{5}{36} = \frac{11}{36}. \quad (3.10b)$$

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

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

◇

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 $P(\text{not } 6) = \frac{5}{6}$ as in (3.8). Hence, in four throws of a die there are sixteen possible outcomes, only one of which has no six. We can use the multiplication rule (3.5) to find that

$$P(\text{not } 6, \text{not } 6, \text{not } 6, \text{not } 6) = P(\text{not } 6)^4 = \left(\frac{5}{6}\right)^4, \quad (3.11)$$

and hence

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

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

◇

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 $P(D) = 0.25/6.25 = 0.04$. \diamond

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. Rolling the dice

If a person rolls two dice, what is the probability $P(n)$ of getting the sum n ? Plot $P(n)$ as a function of n . \square

Problem 3.13. An almost even bet

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

Problem 3.14. Rolling three dice

Suppose that three dice are thrown at the same time. What is the ratio of the probabilities that the sum of the three faces is 10 compared to 9? \square

Problem 3.15. Fallacious reasoning

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. \square

3.3 Mean Values

The specification of the *probability distribution* $P(1), P(2), \dots, P(n)$ 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

\bar{x} . The definition of the mean value of x is

$$\bar{x} \equiv x_1P(1) + x_2P(2) + \dots + x_nP(n) \quad (3.13a)$$

$$= \sum_{i=1}^n x_iP(i), \quad (3.13b)$$

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 given by

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

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

$$\overline{f(x) + g(x)} = \sum_{i=1}^n [f(x_i) + g(x_i)]P(i) \quad (3.15a)$$

$$= \sum_{i=1}^n f(x_i)P(i) + \sum_{i=1}^n g(x_i)P(i), \quad (3.15b)$$

or

$$\overline{f(x) + g(x)} = \overline{f(x)} + \overline{g(x)}. \quad (3.15c)$$

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

$$\overline{cf(x)} = c\overline{f(x)}. \quad (3.16)$$

□

We define the m th *moment* of the probability distribution P as

$$\overline{x^m} \equiv \sum_{i=1}^n x_i^m P(i), \quad (3.17)$$

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 the deviation of x from its mean, we have

$$\Delta x \equiv x - \bar{x}, \quad (3.18)$$

and

$$\overline{\Delta x} = \overline{(x - \bar{x})} = \bar{x} - \bar{x} = 0. \quad (3.19)$$

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. Usually, there is more than one outcome and a measure of the width of the probability distribution is given by

$$\overline{\Delta x^2} \equiv \overline{(x - \bar{x})^2}. \quad (3.20)$$

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

$$\overline{(x - \bar{x})^2} = \overline{(x^2 - 2x\bar{x} + \bar{x}^2)} \quad (3.21a)$$

$$= \overline{x^2} - 2\bar{x}\bar{x} + \bar{x}^2, \quad (3.21b)$$

$$= \overline{x^2} - \bar{x}^2, \quad (3.21c)$$

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

The variance is the mean value of $(x - \bar{x})^2$ 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

$$\sigma_x = \sqrt{\overline{\Delta x^2}} = \sqrt{\overline{x^2} - \bar{x}^2}. \quad (3.22)$$

Example 3.9. Find the mean value \bar{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 $i = 1, \dots, 6$, we have that

$$\bar{x} = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2} = 3.5 \quad (3.23a)$$

$$\overline{x^2} = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{46}{3} \quad (3.23b)$$

$$\overline{\Delta x^2} = \overline{x^2} - \bar{x}^2 = \frac{46}{3} - \frac{49}{4} = \frac{37}{12} \approx 3.08 \quad (3.23c)$$

$$\sigma_x \approx \sqrt{3.08} = 1.76. \quad (3.23d)$$

◇

Example 3.10. On the average, how many times must a die be thrown until a 6 appears?

Solution. Although it might be obvious that the answer is six, it is instructive to confirm this answer directly. Let p be the probability of a six on a given throw.

To calculate m , the mean number of throws needed before a six appears, we calculate the probability of obtaining a six for the first time on the i th throw, multiply this probability by i ,

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

Table 3.2: Probability of a six for the first time on throw i , where $p = 1/6$ is the probability of a six on a given throw and $q = 1 - p$.

and then sum over all i . The first few probabilities are listed in Table 3.2. The resulting sum is

$$m = p + 2pq + 3pq^2 + 4pq^3 + \dots \quad (3.24a)$$

$$= p(1 + 2q + 3q^2 + \dots) \quad (3.24b)$$

$$= p \frac{d}{dq} (1 + q + q^2 + q^3 + \dots) \quad (3.24c)$$

$$= p \frac{d}{dq} \frac{1}{1 - q} = \frac{p}{(1 - q)^2} = \frac{1}{p}. \quad (3.24d)$$

Another way to obtain this result is to use the following recursive argument. Because the throws are independent, the mean number of additional throws needed after throwing the die any number of times is still m . If we throw the die once and do not obtain a six, we will need on average m more throws or a total of $m + 1$. If we throw it twice and do not obtain a six on either throw, we will need m more throws or a total of $m + 2$ and so forth. The contribution to the mean due to failing on the first throw and then succeeding sometime later is $q(m + 1)$. The probability of succeeding on the first throw is p and the contribution to the mean is $p(1) = p$. The mean is the sum of these two terms or $m = (1 - p)(m + 1) + p$. The solution for m is $m = 1/p$. \diamond

3.4 The Meaning of Probability

How can we assign probabilities to the various events? If 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 3.85).

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 ensure 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 assessment 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 both children are girls?
- (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.

tosses	heads	fraction of heads
10	4	0.4
50	29	0.58
100	49	0.49
200	101	0.505
500	235	0.470
1,000	518	0.518
10,000	4997	0.4997
100,000	50021	0.50021
500,000	249946	0.49999
1,000,000	500416	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. The robot places these pieces in a hat, shakes the hat, and then chooses 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.

Problem 3.19. Multiple tosses of a single coin

Use program `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? \square

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.

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 directly measurable. If we make further measurements, they can only influence future assessments of the probability.

3.4.1 Information and uncertainty

Consider two experiments that each have two outcomes E_1 and E_2 with probabilities P_1 and P_2 . For example, the experiments could correspond to the toss of a coin. In the first experiment the coin has probabilities $P_1 = P_2 = 1/2$, and in the second experiment (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 than the result of the second experiment.

Next 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 $P_1 = P_2 = P_3 = P_4 = P_5 = P_6 = 1/6$. The fourth experiment is the most uncertain because there are more equally likely outcomes and the second experiment is the least uncertain.

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

$$S(\Omega = 1) = 0. \quad (3.25)$$

We also have that

$$S(\Omega_1) > S(\Omega_2) \text{ if } \Omega_1 > \Omega_2. \quad (3.26)$$

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

We next consider the form of S for 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

$$S(\Omega_1\Omega_2) = S(\Omega_1) + S(\Omega_2). \quad (3.27)$$

It is remarkable that there is an unique functional form that satisfies the three conditions (3.25)–(3.27). We can find this form by writing (3.27) in the form

$$S(xy) = S(x) + S(y), \quad (3.28)$$

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.)

Now we take the partial derivative of $S(xy)$ with respect to x and then with respect to y . We let $z = xy$ and obtain

$$\frac{\partial S(z)}{\partial x} = \frac{\partial z}{\partial x} \frac{dS(z)}{dz} = y \frac{dS(z)}{dz} \quad (3.29a)$$

$$\frac{\partial S(z)}{\partial y} = \frac{\partial z}{\partial y} \frac{dS(z)}{dz} = x \frac{dS(z)}{dz}. \quad (3.29b)$$

From (3.28) we have

$$\frac{\partial S(z)}{\partial x} = \frac{dS(x)}{dx} \quad (3.30a)$$

$$\frac{\partial S(z)}{\partial y} = \frac{dS(y)}{dy}. \quad (3.30b)$$

By comparing the right-hand side of (3.29) and (3.30), we have

$$\frac{dS}{dx} = y \frac{dS}{dz} \quad (3.31a)$$

$$\frac{dS}{dy} = x \frac{dS}{dz}. \quad (3.31b)$$

If we multiply (3.31a) by x and (3.31b) by y , we obtain

$$x \frac{dS(x)}{dx} = y \frac{dS(y)}{dy} = z \frac{dS(z)}{dz}. \quad (3.32)$$

Note that the first term in (3.32) depends only on x and the second term depends only on y . Because x and y are independent variables, the three terms in (3.32) must be equal to a constant. Hence we have the desired condition

$$x \frac{dS(x)}{dx} = y \frac{dS(y)}{dy} = A, \quad (3.33)$$

where A is a constant. The differential equation in (3.33) can be integrated to give

$$S(x) = A \ln x + B. \quad (3.34)$$

The integration constant B must be equal to zero to satisfy the condition (3.25). The constant A is arbitrary so we choose $A = 1$. Hence for equal probabilities we have that

$$S(\Omega) = \ln \Omega. \quad (3.35)$$

What about the case where the probabilities for the various events are unequal? We will not derive the result here but only state the result that the general form of the uncertainty S is

$$S = - \sum_i P_i \ln P_i. \quad (3.36)$$

Note that if all the probabilities are equal, then

$$P_i = \frac{1}{\Omega}. \quad (3.37)$$

for all i . In this case

$$S = - \sum_i \frac{1}{\Omega} \ln \frac{1}{\Omega} = \Omega \frac{1}{\Omega} \ln \Omega = \ln \Omega, \quad (3.38)$$

because there are Ω equal terms in the sum. Hence (3.36) reduces to (3.35) 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 expect 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*. Let's reconsider the toss of a coin. In this case S is given by

$$S = - \sum_i P_i \ln P_i = -(P_1 \ln P_1 + P_2 \ln P_2) \quad (3.39a)$$

$$= -(P_1 \ln P_1 + (1 - P_1) \ln(1 - P_1)), \quad (3.39b)$$

where we have used the fact that $P_1 + P_2 = 1$. We use the principle of maximum uncertainty and set the derivative of S with respect to P_1 equal to zero:⁴

$$\frac{dS}{dP_1} = -[\ln P_1 + 1 - \ln(1 - P_1) - 1] = -\ln \frac{P_1}{1 - P_1} = 0. \quad (3.40)$$

The solution of (3.40) satisfies

$$\frac{P_1}{1 - P_1} = 1, \quad (3.41)$$

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

$$\frac{\partial^2 S}{\partial P_1^2} = -\left[\frac{1}{P_1} + \frac{1}{1 - P_1}\right] = -4, \quad (3.42)$$

which is less than zero, as expected for a maximum.

Problem 3.20. Uncertainty

- 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?
- What is the uncertainty for $P_1 = 1/5$ and $P_2 = 4/5$? How does the uncertainty in this case compare to that in part (a)?

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

(c) On page 119 we discussed four experiments with various outcomes. Compare the uncertainty S of the third and fourth experiments. \square

Example 3.11. 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, respectively. 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 consistent with the information that $f = 1.9$?

Solution. We have

$$S = -[P_1 \ln P_1 + P_2 \ln P_2 + P_3 \ln P_3]. \quad (3.43)$$

We also know that

$$f = 1P_1 + 2P_2 + 3P_3, \quad (3.44)$$

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 (3.44) as

$$f = P_1 + 2P_2 + 3(1 - P_1 - P_2) = 3 - 2P_1 - P_2. \quad (3.45)$$

We then use (3.45) to eliminate P_2 and P_3 from (3.43) with $P_2 = 3 - f - 2P_1$ and $P_3 = f - 2 + P_1$:

$$S = -[P_1 \ln P_1 + (3 - f - 2P_1) \ln(3 - f - 2P_1) + (f - 2 + P_1) \ln(f - 2 + P_1)]. \quad (3.46)$$

Because S in (3.46) depends on only P_1 , we can differentiate S with respect to P_1 to find its maximum value:

$$\frac{dS}{dP_1} = -[\ln P_1 - 1 - 2[\ln(3 - f - 2P_1) - 1] + [\ln(f - 2 + P_1) - 1]] \quad (3.47a)$$

$$= \ln \frac{P_1(f - 2 + P_1)}{(3 - f - 2P_1)^2} = 0. \quad (3.47b)$$

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). \diamond

Problem 3.21. Fill in the missing steps in Example 3.11 and solve for P_1 , P_2 , and P_3 . \square

In Section 3.11.1 we consider maximizing 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. Conditional probability and Bayes' theorem are very important for the analysis of data including spam filters for email and image restoration, for example. Bayes' theorem gives us a way of understanding how the probability that a hypothesis is true is affected by new evidence.

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

$$P(A) = P(A|B) + P(A|-B), \quad (3.48)$$

where $\neg B$ means that B did not occur. We also know that

$$P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A). \quad (3.49)$$

Equation (3.49) 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 that A occurs. Note that $P(A \text{ and } B)$ is the same as $P(B \text{ and } A)$, but $P(A|B)$ does not have the same meaning as $P(B|A)$.

We can rearrange (3.49) to obtain Bayes' theorem

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}. \quad (\text{Bayes' theorem}) \quad (3.50)$$

We can generalize (3.50) for multiple possible outcomes A_i for the same B. We rewrite (3.50) as

$$\boxed{P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}}. \quad (\text{multiple outcomes}) \quad (3.51)$$

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

$$P(B) = \sum_i P(B|A_i)P(A_i). \quad (3.52)$$

If we substitute (3.52) for $P(B)$ into (3.51), we obtain

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)}. \quad (3.53)$$

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 $P(B|A_i)P(A_i)$ change.

Example 3.12. A chess program has two modes, expert (E) and novice (N). The expert mode beats you 75% of the time and the novice mode beats you 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 $P(N|WW)$, which is difficult to calculate directly. Bayes theorem lets us use the probability $P(WW|N)$, which is easy to calculate, to determine $P(N|WW)$. We use (3.50) to write

$$P(N|WW) = \frac{P(WW|N)P(N)}{P(WW)}. \quad (3.54)$$

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

We next have to calculate $P(WW)$. There are two ways that the program could have won the two games: You chose the novice mode and it won twice, or you chose the expert mode and

it won twice. Because N and E are mutually exclusive, we have $P(WW) = P(N \text{ and } WW) + P(E \text{ and } WW)$. From (3.49) we have

$$P(WW) = P(WW|N)P(N) + P(WW|E)P(E) \quad (3.55a)$$

$$= (1/2 \times 1/2 \times 1/2) + (3/4 \times 3/4 \times 1/2) = \frac{13}{32}. \quad (3.55b)$$

Hence

$$P(N|WW) = \frac{P(WW|N)P(N)}{P(WW)} = \frac{(1/4 \times 1/2)}{\frac{13}{32}} = \frac{4}{13} \approx 0.31. \quad (3.56)$$

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

Example 3.13. 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 $P(\text{red}|A) = 0.4$ and $P(\text{red}|B) = 0.3$.

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

Solution. We can use the total probability law (3.49) to write

$$P(\text{red}) = P(\text{red}|A)P(A) + P(\text{red}|B)P(B) \quad (3.57a)$$

$$= (0.4 \times 0.3) + (0.3 \times 0.7) = 0.33. \quad (3.57b)$$

So Alice will find on average 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 (3.53) and obtain

$$P(A|\text{red}) = \frac{P(\text{red}|A)P(A)}{P(\text{red}|A)P(A) + P(\text{red}|B)P(B)} \quad (3.58a)$$

$$= \frac{0.4 \times 0.3}{(0.4 \times 0.3) + (0.3 \times 0.7)} = \frac{12}{33} = \frac{4}{11} \approx 0.36. \quad (3.58b)$$

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. \diamond

Example 3.14. 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

$$P(5H) = P(5H|\text{fair})P(\text{fair}) + P(5H|\text{not fair})P(\text{not fair}) \quad (3.59a)$$

$$= [(1/2)^5 \times 3/4] + [1 \times 1/4] = 35/128 \approx 0.27. \quad (3.59b)$$

$$P(\text{fair}|5H) = P(5H|\text{fair})P(\text{fair})/P(5H) \quad (3.59c)$$

$$= \frac{[(1/2)^5 \times 3/4]}{35/128} = 3/35 = 0.12. \quad (3.59d)$$

Thus the probability that the coin is fair given the five heads in succession is less than the probability $3/4$ of picking a fair coin randomly out of the bag. \diamond

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 (a)?
- (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 coin 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? \square

Example 3.15. Let's Make A Deal

Consider the quandary known as the Let's Make A Deal or Monty Hall problem.⁵ 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 the contestant 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 – stay with the original choice or switch to door 3? What would you do?

Let us use Bayes' theorem (3.53) to determine the best course of action. We want to calculate

$$P(A_1|B) = P(\text{car behind door 1}|\text{door 2 open after door 1 chosen}), \quad (3.60a)$$

and

⁵This question was posed on the TV game show, "Let's Make A Deal," hosted by Monty Hall. See for example, www.letsmakeadeal.com/problem.htm and www.nytimes.com/2008/04/08/science/08monty.html.

$$P(A_3|B) = P(\text{car behind door 3}|\text{door 2 open after door 1 chosen}), \quad (3.60b)$$

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

$$P(\text{door 2 open after door 1 chosen}|\text{car behind 1}) = \frac{1}{2} \quad (3.61a)$$

$$P(\text{door 2 open after door 1 chosen}|\text{car behind 2}) = 0 \quad (3.61b)$$

$$P(\text{door 2 open after door 1 chosen}|\text{car behind 3}) = 1. \quad (3.61c)$$

From Bayes' theorem we have

$$P(\text{car behind 1}|\text{door 2 open after door 1 chosen}) = \frac{\frac{1}{2} \times \frac{1}{3}}{(\frac{1}{2} \times \frac{1}{3}) + (0 \times \frac{1}{3}) + (1 \times \frac{1}{3})} = \frac{1}{3} \quad (3.62a)$$

$$P(\text{car behind 3}|\text{door 2 open after door 1 chosen}) = \frac{1 \times \frac{1}{3}}{(\frac{1}{2} \times \frac{1}{3}) + (0 \times \frac{1}{3}) + (1 \times \frac{1}{3})} = \frac{2}{3}. \quad (3.62b)$$

The results in (3.62) suggest the contestant has a higher probability of winning the car by switching doors and choosing door 3. The same logic suggests that one should always switch doors independently of which door was originally chosen.

Problem 3.23. Simple considerations

Make a table showing the three possible arrangements of the car and explain in simple terms why switching doubles the chances of winning.

Problem 3.24. 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 Let's Make A Deal 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 (a) the same as found in Example 3.15? Discuss why the probability that the car is behind door 1 depends on what the host knows. \square

Example 3.16. 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 disease, 98% of the times the test will come out positive and 2% negative. Also 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? Answer the question before you read the solution using Bayes theorem.

Solution. Let $P(D)$ represent the probability of having the disease given no other information except the age of the person. In this example $P(D) = 1/10000 = 0.0001$. The probability of not having the disease is $P(N) = 1 - P(D) = 0.9999$. Let $P(+|D) = 0.98$ represent the probability of testing positive given that you have the disease, and $P(+|N) = 0.02$ represent the probability of testing positive given that you do not have the disease. We wish to find the probability $P(D|+)$ that you will test positive for the disease and actually be sick. From Bayes' theorem we have

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|N)P(N)} \quad (3.63a)$$

$$= \frac{(0.98)(0.0001)}{(0.98)(0.0001) + (0.02)(0.9999)} \quad (3.63b)$$

$$= 0.0047 = 0.47\%. \quad (3.63c)$$

We expect that you will find this result difficult to accept. How can it be that the probability of having the disease is so small given the high reliability of the test? This example, and others like it, shows our lack of intuition about many statistical problems. \diamond

Problem 3.25. Testing accuracy

Suppose that a person tests positive for a disease that occurs in 1 in 100, 1 in 1000, 1 in 10,000, or 1 in 100,000 people. Determine in each case how accurate the test must be for the test to give a probability equal to at least 50% of actually having a disease. \square

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 the surgery, 2% die due to complications, and the rest are cured. We have that

$$P(\text{death rate due to cancer}) = P(\text{death}|\text{cancer})P(\text{cancer}) \quad (3.64a)$$

$$= 0.25 \times 0.001 = 0.00025 \quad (3.64b)$$

$$P(\text{death due to test}) = P(\text{die}|\text{surgery})P(\text{surgery}|\text{positive})P(\text{positive}|\text{cancer}) \quad (3.64c)$$

$$= 0.02 \times 0.90 \times 0.02 = 0.00036. \quad (3.64d)$$

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

Problem 3.26. Three balls in a sack

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. \square

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.

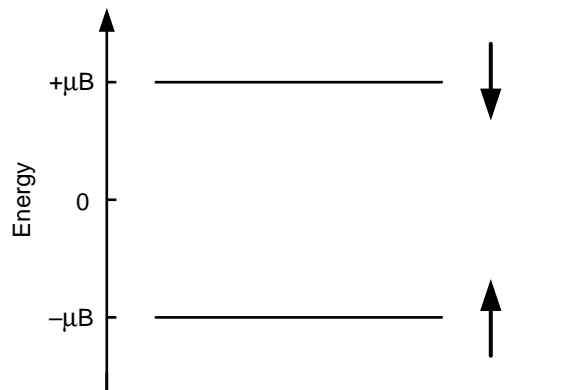


Figure 3.2: The energy of a spin 1/2 magnetic dipole. Note that the state of lowest energy is negative.

3.5 Bernoulli Processes and the Binomial Distribution

Because most physicists spend little time gambling,⁶ 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 dipoles each having a magnetic moment μ and associated spin in an external magnetic field B . The field B is in the up ($+z$) direction. According to quantum mechanics the component of the magnetic dipole moment along a given axis is limited to certain discrete values. Spin 1/2 implies that a magnetic dipole can either point up (parallel to B) or down (antiparallel to B). We will use the word spin as a shorthand for magnetic dipole. The energy of interaction of a spin with the magnetic field is $E = -\mu B$ if the spin is up and $+\mu B$ if the spin is down (see Figure 3.2). More generally, we can write $E = -s\mu B$, where $s = +1$ (spin up) or $s = -1$ (spin down). As discussed in Section 1.9, page 22, this model is a simplification of a more realistic magnetic system.

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 given. 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

⁶After a Las Vegas hotel hosted a meeting of the American Physical Society (APS) in March, 1986, the APS was asked never to return. The Las Vegas newspaper headline read, "Meeting of physicists in town, lowest casino take ever."

magnetic moment of the system. For a system of N spins the magnetization is given by

$$M = \mu(s_1 + s_2 + \dots + s_N) = \mu \sum_{i=1}^N s_i. \quad (3.65)$$

In the following, we will take $\mu = 1$ for convenience whenever it will not cause confusion.

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 calculate the mean value of M , we need to take the mean values of both sides of (3.65). We interchange the sum and the average (see (3.15c)) and write

$$\overline{M} = \overline{\left(\sum_{i=1}^N s_i \right)} = \sum_{i=1}^N \overline{s_i}. \quad (3.66)$$

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, $\overline{s_1} = \overline{s_2} = \dots = \overline{s_N} \equiv \overline{s}$. Therefore the sum in (3.66) consists of N equal terms and can be written as

$$\overline{M} = N\overline{s}. \quad (3.67)$$

The meaning of (3.67) 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

$$\overline{M} = N(p - q). \quad (3.68)$$

To calculate the variance of M , that is, $\overline{(M - \overline{M})^2}$, we write

$$\Delta M = M - \overline{M} = \sum_{i=1}^N \Delta s_i, \quad (3.69)$$

where

$$\Delta s_i \equiv s_i - \overline{s}. \quad (3.70)$$

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

$$(\Delta M)^2 = (\Delta s_1 + \Delta s_2 + \Delta s_3)(\Delta s_1 + \Delta s_2 + \Delta s_3) \quad (3.71a)$$

$$= [(\Delta s_1)^2 + (\Delta s_2)^2 + (\Delta s_3)^2] + 2[\Delta s_1 \Delta s_2 + \Delta s_1 \Delta s_3 + \Delta s_2 \Delta s_3]. \quad (3.71b)$$

We take the mean value of (3.71b), interchange the order of the sums and averages, and write

$$\overline{(\Delta M)^2} = \overline{[(\Delta s_1)^2 + (\Delta s_2)^2 + (\Delta s_3)^2]} + 2[\overline{\Delta s_1 \Delta s_2} + \overline{\Delta s_1 \Delta s_3} + \overline{\Delta s_2 \Delta s_3}]. \quad (3.72)$$

The first term in brackets on the right of (3.72) 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

$$\overline{\Delta s_i \Delta s_j} = \overline{\Delta s_i} \overline{\Delta s_j} = 0, \quad (i \neq j) \quad (3.73)$$

because $\overline{\Delta s_i} = 0$. That is, each cross term vanishes on the average. Hence (3.73) reduces to a sum of squared terms

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

Because each spin is equivalent on the average, each term in (3.74) is equal. Hence, we obtain the desired result

$$\overline{(\Delta M)^2} = 3\overline{(\Delta s)^2}. \quad (3.75)$$

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

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

$$\overline{(\Delta s)^2} = \overline{s^2} - \overline{s}^2 = 1 - (p - q)^2 = 1 - (2p - 1)^2 \quad (3.76a)$$

$$= 1 - 4p^2 + 4p - 1 = 4p(1 - p) = 4pq, \quad (3.76b)$$

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

$$\overline{(\Delta M)^2} = 3(4pq). \quad (3.77)$$

Problem 3.27. Variance of N spins

In the text we showed that $\overline{(\Delta M)^2} = 3\overline{(\Delta s)^2}$ for $N = 3$ spins (see (3.75) and (3.77)). Use similar considerations for N noninteracting spins to show that

$$\overline{(\Delta M)^2} = N(4pq). \quad (3.78)$$

□

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 3.3). 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 3.3. Although each outcome is distinct, several of the configurations have the same number of up spins. The main quantity of interest is the probability $P_N(n)$ that n spins point up out a total of N spins. For example for $N = 3$ spins, there are three states with $n = 2$, each with probability p^2q so the probability that two spins are up is equal to $3p^2q$. From Figure 3.3 we see that

$$P_3(n = 3) = p^3 \quad (3.79a)$$

$$P_3(n = 2) = 3p^2q \quad (3.79b)$$

$$P_3(n = 1) = 3pq^2 \quad (3.79c)$$

$$P_3(n = 0) = q^3. \quad (3.79d)$$

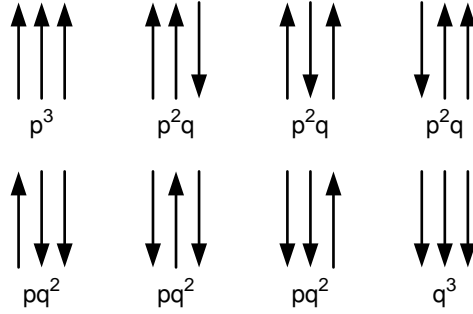


Figure 3.3: An ensemble of $N = 3$ spins. The arrow indicates the direction of a spin. The probability of each member of the ensemble is shown.

Example 3.17. Find the first two moments of $P_3(n)$.

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

$$\bar{n} = (0 \times q^3) + (1 \times 3pq^2) + (2 \times 3p^2q) + (3 \times p^3) \quad (3.80a)$$

$$= 3p(q^2 + 2pq + p^2) = 3p(q + p)^2 = 3p. \quad (3.80b)$$

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

$$\overline{n^2} = (0 \times q^3) + (1 \times 3pq^2) + (4 \times 3p^2q) + (9 \times p^3) \quad (3.81a)$$

$$= 3p(q^2 + 4pq + 3p^2) = 3p(q + 3p)(q + p) \quad (3.81b)$$

$$= 3p(q + 3p) = (3p)^2 + 3pq. \quad (3.81c)$$

Hence

$$\overline{(n - \bar{n})^2} = \overline{n^2} - \bar{n}^2 = 3pq. \quad (3.82)$$

◇

The mean magnetization M or the mean number of up spins minus the mean number of down spins is given by $\bar{M} = [\bar{n} - (3 - \bar{n})] = 2\bar{n} - 3 = 6p - 3$, or $\bar{M} = 3(2p - 1) = 3(p - q)$ in agreement with (3.68).

Problem 3.28. Coin flips

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

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?⁷ We will consider an idealized example of a random

⁷The history of the random walk problem is discussed by Montroll and Shlesinger (see the references).

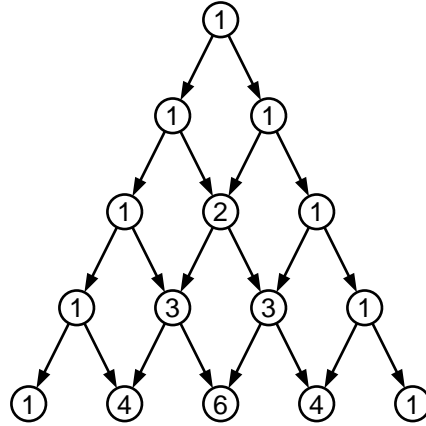


Figure 3.4: The values of the first few coefficients $W_N(n, n')$. Each number is the sum of the two numbers to the left and right above it. This construction is called a Pascal triangle.

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 is $N = n + n'$. What is the probability that a random walker in one dimension has taken three steps to the right out of four steps? \square

From the above examples and problems, we see that the probability distributions of noninteracting spins, the flip of a coin, and a simple one-dimensional 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 noninteracting spins with 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

$$P_N(n) = W_N(n, n') p^n q^{n'}, \quad (3.83)$$

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

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

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

$$W_N(n, n') = W_{N-1}(n - 1, n') + W_{N-1}(n, n' - 1). \quad (3.84)$$

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 (3.84) to construct $W_N(n, n')$ for any desired N . For example,

$$W_2(2, 0) = W_1(1, 0) + W_1(2, -1) = 1 + 0 = 1 \quad (3.85a)$$

$$W_2(1, 1) = W_1(0, 1) + W_1(1, 0) = 1 + 1 = 2. \quad (3.85b)$$

$$W_2(0, 2) = W_1(-1, 2) + W_1(0, 1) = 0 + 1. \quad (3.85c)$$

In Figure 3.4 we show that $W_N(n, n')$ forms a pyramid or (a Pascal) triangle.

It is straightforward to show by induction that the expression

$$W_N(n, n') = \frac{N!}{n!n!} = \frac{N!}{n!(N-n)!} \quad (3.86)$$

satisfies the relation (3.84). Note the convention $0! = 1$. We can combine (3.83) and (3.86) to find the desired result

$$P_N(n) = \frac{N!}{n!(N-n)!} p^n q^{N-n}. \quad (\text{binomial distribution}) \quad (3.87)$$

The form (3.87) is the *binomial distribution*. Note that for $p = q = 1/2$, $P_N(n)$ reduces to

$$P_N(n) = \frac{N!}{n!(N-n)!} 2^{-N}. \quad (3.88)$$

The probability $P_N(n)$ is shown in Figure 3.5 for $N = 64$.

Problem 3.30. Binomial distribution

- (a) Calculate the probability $P_N(n)$ that n spins are up out of a total of N for $N = 4$ and $N = 16$ and put your results in 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 $p = q = 1/2$ for simplicity. Although it is better to first do the calculation of $P_N(n)$ by hand, you can use the program `Binomial`.
- (b) Use program `Binomial` to plot $P_N(n)$ for larger values of N . Assume that $p = q = 1/2$. Determine the value of n corresponding to the maximum of the probability and visually estimate the width for each value of N . What is your measure of the width? One measure is to use the value of n at which $P_N(n)$ is equal to half its value at its maximum. What is the qualitative dependence of the width on N ? Also compare the relative heights of the maximum of P_N for increasing values of N .

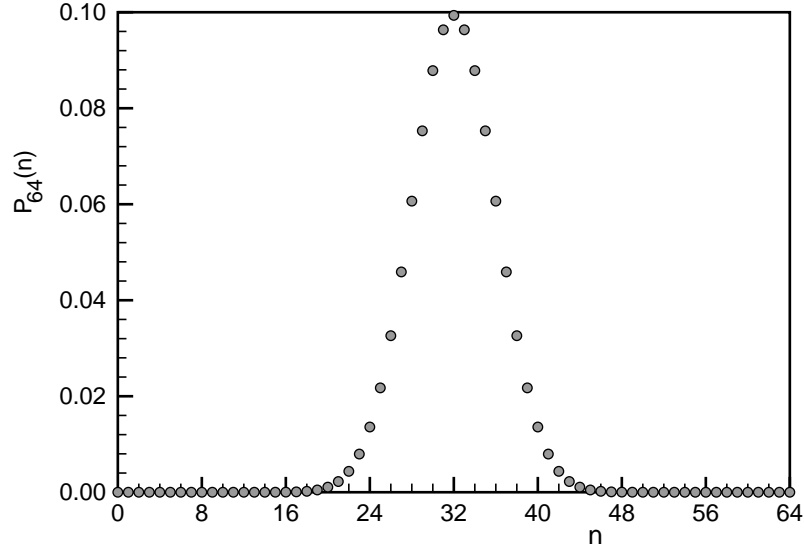


Figure 3.5: The binomial distribution $P_{64}(n)$ for $p = q = 1/2$. What is your visual estimate of the width of the distribution?

- (c) Program `Binomial` also plots $P_N(n)$ versus n/\bar{n} . Does the width of $P_N(n)$ appear to become larger or smaller as N is increased?
- (d) Plot $\ln P_N(n)$ versus n for $N = 16$. (Choose `Log Axes` under the `Views` menu.) Describe the qualitative dependence of $\ln P_N(n)$ on n . Can $\ln P_N(n)$ be fitted to a parabola of the form $A + B(n - \bar{n})^2$, where A and B are fit parameters? \square

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 values of p and q do you think the width is a maximum for a given N ? \square

Example 3.18. Show that the expression (3.87) for $P_N(n)$ satisfies the normalization condition (3.2).

Solution. The reason that (3.87) 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

$$(p + q)^N = \sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n q^{N-n}. \quad (3.89)$$

We use (3.87) and write

$$\sum_{n=0}^N P_N(n) = \sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n q^{N-n} = (p + q)^N = 1^N = 1, \quad (3.90)$$

where we have used (3.89) and the fact that $p + q = 1$. \diamond

Mean value. We now find an analytical expression for the dependence of the mean number of up spins \bar{n} on N and p . From the definition (3.13) and (3.87) we have

$$\bar{n} = \sum_{n=0}^N n P_N(n) = \sum_{n=0}^N n \frac{N!}{n!(N-n)!} p^n q^{N-n}. \quad (3.91)$$

We evaluate the sum in (3.91) by using a technique that is useful in a variety of contexts.⁸ The technique is based on the fact that

$$p \frac{d}{dp} p^n = n p^n. \quad (3.92)$$

We use (3.92) to rewrite (3.91) as

$$\bar{n} = \sum_{n=0}^N \frac{N!}{n!(N-n)!} \left(p \frac{\partial}{\partial p} p^n \right) q^{N-n}. \quad (3.93)$$

We have used a partial derivative in (3.93) to remind us that the derivative operator does not act on q , which we have temporarily assumed to be an independent variable. We interchange the order of summation and differentiation in (3.93) and write

$$\bar{n} = p \frac{\partial}{\partial p} \left[\sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n q^{N-n} \right] \quad (3.94a)$$

$$= p \frac{\partial}{\partial p} (p+q)^N, \quad (3.94b)$$

Because the operator acts only on p , we have

$$\bar{n} = pN(p+q)^{N-1}. \quad (3.95)$$

The result (3.95) is valid for arbitrary p and q , and hence it is applicable for $p + q = 1$. Thus our desired result is

$$\bar{n} = pN. \quad (3.96)$$

The nature of the dependence of \bar{n} on N and p should be intuitively clear. Compare the general result (3.96) to the result (3.80b) for $N = 3$. What is the dependence of \bar{n}' on N and p ?

Relative fluctuations. To determine $\overline{\Delta n^2}$ we need to know $\overline{n^2}$ (see the relation (3.21)). The

⁸The integral $\int_0^\infty x^n e^{-ax^2}$ for $a > 0$ is evaluated in Appendix A using a similar technique.

average value of n^2 can be calculated in a manner similar to that for \bar{n} . We write

$$\overline{n^2} = \sum_{n=0}^N n^2 \frac{N!}{n!(N-n)!} p^n q^{N-n} \quad (3.97a)$$

$$= \sum_{n=0}^N \frac{N!}{n!(N-n)!} \left(p \frac{\partial}{\partial p}\right)^2 p^n q^{N-n} \quad (3.97b)$$

$$= \left(p \frac{\partial}{\partial p}\right)^2 \sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n q^{N-n} = \left(p \frac{\partial}{\partial p}\right)^2 (p+q)^N \quad (3.97c)$$

$$= \left(p \frac{\partial}{\partial p}\right) [pN(p+q)^{N-1}] \quad (3.97d)$$

$$= p[N(p+q)^{N-1} + pN(N-1)(p+q)^{N-2}]. \quad (3.97e)$$

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

$$\overline{n^2} = p[N + pN(N-1)] \quad (3.98a)$$

$$= p[pN^2 + N(1-p)] = (pN)^2 + p(1-p)N \quad (3.98b)$$

$$= \bar{n}^2 + pqN, \quad (3.98c)$$

where we have used (3.96) and substituted $q=1-p$. Hence, from (3.98c) we find that the variance of n is given by

$$\sigma_n^2 = \overline{(\Delta n)^2} = \overline{n^2} - \bar{n}^2 = pqN. \quad (3.99)$$

Problem 3.32. Width of the binomial distribution

Compare the calculated values of σ_n from (3.99) with your estimates in Problem 3.30 and to the exact result (3.82) for $N=4$. Explain why σ_n is a measure of the width of $P_N(n)$. \square

The relative width of the probability distribution of n is given by (3.96) and (3.99)

$$\frac{\sigma_n}{\bar{n}} = \frac{\sqrt{pqN}}{pN} = \left(\frac{q}{p}\right)^{1/2} \frac{1}{\sqrt{N}}. \quad (3.100)$$

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

Frequently we need to evaluate $\ln N!$ for $N \gg 1$. An approximation for $\ln N!$ known as *Stirling's approximation* is⁹

$$\boxed{\ln N! \approx N \ln N - N + \frac{1}{2} \ln(2\pi N)}. \quad (\text{Stirling's approximation}) \quad (3.101)$$

In some contexts we can neglect the logarithmic term in (3.101) and use the weaker approximation

$$\ln N! \approx N \ln N - N. \quad (3.102)$$

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

⁹It is more accurate to call it the De Moivre-Stirling approximation because de Moivre had already found that $n! \approx c\sqrt{\pi n}/e^n$ for some constant c . Stirling's contribution was to identify the constant c as $\sqrt{2\pi}$.

Problem 3.33. Range of applicability of Stirling's approximation

- (a) What is the largest value of $\ln N!$ that you can calculate exactly using your calculator?
- (b) Compare the approximations (3.101) and (3.102) to each other and to the exact value of $\ln N!$ for $N = 5, 10, 20,$ and 50 . If necessary, compute $\ln N!$ using the relation

$$\ln N! = \sum_{m=1}^N \ln m. \quad (3.103)$$

Put your results in a table. What is the percentage error of the two approximations for $N = 50$?

- (c) Use Stirling's approximation to show that

$$\frac{d}{dx} \ln x! = \ln x \text{ for } x \gg 1. \quad (3.104)$$

□

Problem 3.34. 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 is uniform on the average, there are fluctuations in the density. Divide the volume V into two parts V_1 and V_2 with $V = V_1 + V_2$.

- (a) What is the probability p that a particular molecule is in the volume V_1 ?
- (b) What is the probability that N_1 molecules are in V_1 and N_2 molecules are in V_2 , where $N = N_1 + N_2$?
- (c) What is the average number of molecules in each part?
- (d) What are the relative fluctuations of the number of molecules in each part? □

Problem 3.35. Random walk

Suppose that a random walker takes n steps to the right and n' steps to the left for a total of N steps. 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 after N steps. What is the mean value \bar{x} for a N -step random walk? What is the N -dependence of the variance $\overline{(\Delta x)^2}$? □

Problem 3.36. Monte Carlo simulation of a one-dimensional random walk

The program `RandomWalk1D` simulates a random walk in one dimension. 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 $q = 1 - p$. 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 after N steps 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.¹⁰

- Is the value of x for one trial of any interest? Why do we have to average over many trials?
- Will we obtain the exact answer for the probability distribution by doing a Monte Carlo simulation?
- Describe the changes of the histogram for larger values of N and $p = 1/2$.
- 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? Estimate 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 ?
- Choose $N = 4$ and $p = 1/2$. How does the histogram change, if at all, as the number of walks increases for fixed N ? \square

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

The first step is to realize that $P_N(n)$ for $N \gg 1$ is a rapidly varying function of n near the maximum of $P_N(n)$ at $\tilde{n} = pN$. 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 Taylor series expansion of $\ln P_N(n)$ will 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 have

$$\ln P_N(n) = \ln P_N(n = \tilde{n}) + (n - \tilde{n}) \left. \frac{d \ln P_N(n)}{dn} \right|_{n=\tilde{n}} + \frac{1}{2} (n - \tilde{n})^2 \left. \frac{d^2 \ln P_N(n)}{d^2 n} \right|_{n=\tilde{n}} + \dots \quad (3.105)$$

Because the expansion (3.105) is about the maximum $n = \tilde{n}$, the first derivative $d \ln P_N(n) / dn \big|_{n=\tilde{n}}$ must be zero and the second derivative $d^2 \ln P_N(n) / dn^2 \big|_{n=\tilde{n}}$ must be negative. We assume that

¹⁰The name “Monte Carlo” was first used in by Nicholas Metropolis and Stanislaw Ulam in “The Monte Carlo method,” *Journal of the American Statistical Association* **44** (247), 335–341 (1949).

the higher terms in (3.105) can be neglected (see Problem 3.76) and adopt the notation

$$\ln A = \ln P_N(n = \tilde{n}), \quad (3.106)$$

and

$$B = -\left. \frac{d^2 \ln P_N(n)}{dn^2} \right|_{n=\tilde{n}}. \quad (3.107)$$

The approximation (3.105) and the notation in (3.106) and (3.107) allows us to write

$$\ln P_N(n) \approx \ln A - \frac{1}{2}B(n - \tilde{n})^2, \quad (3.108)$$

or

$$P_N(n) \approx A e^{-\frac{1}{2}B(n-\tilde{n})^2}. \quad (3.109)$$

We next use Stirling's approximation (3.101) to evaluate the first two derivatives of $\ln P_N(n)$ to find the parameters B and \tilde{n} . We first take the logarithm of both sides of (3.87) and obtain

$$\ln P_N(n) = \ln N! - \ln n! - \ln(N - n)! + n \ln p + (N - n) \ln q. \quad (3.110)$$

It is straightforward to use the approximation (3.104) to obtain

$$\frac{d(\ln P_N(n))}{dn} = -\ln n + \ln(N - n) + \ln p - \ln q. \quad (3.111)$$

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

$$\frac{N - \tilde{n}}{\tilde{n}} = \frac{q}{p}, \quad (3.112)$$

or $(N - \tilde{n})p = \tilde{n}q$. The relation $p + q = 1$ allows us to write

$$\tilde{n} = pN, \quad (3.113)$$

as expected. Note that $\tilde{n} = \bar{n}$, that is, the value of n for which $P_N(n)$ is a maximum is also the mean value of n .

The second derivative can be found from (3.111). We have

$$\frac{d^2(\ln P_N(n))}{dn^2} = -\frac{1}{n} - \frac{1}{N - n}. \quad (3.114)$$

Hence, the coefficient B defined in (3.107) is given by

$$B = -\left. \frac{d^2 \ln P_N(n)}{dn^2} \right|_{n=\tilde{n}} = \frac{1}{\tilde{n}} + \frac{1}{N - \tilde{n}} = \frac{1}{Npq}. \quad (3.115)$$

From the relation (3.99) we see that

$$B = \frac{1}{\sigma^2}, \quad (3.116)$$

where σ^2 is the variance of n . In Problem 3.37 you will be asked to show that the coefficient A in (3.107) can be approximated for large N as

$$A = \frac{1}{(2\pi Npq)^{1/2}} = \frac{1}{(2\pi\sigma^2)^{1/2}}. \quad (3.117)$$

We thus find the form of the Gaussian probability distribution

$$P_N(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(n-\bar{n})^2/2\sigma^2}. \quad (\text{Gaussian probability distribution}) \quad (3.118)$$

An alternative derivation of the parameters A , \bar{n} , and B is given in Problem 3.72.

Problem 3.37. Calculation of the normalization constant

Derive the form of A in (3.117) using Stirling's approximation (3.101). Note that the weaker form of Stirling's approximation in (3.102) yields the incorrect result that $\ln A = 0$. \square

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 (3.118) for $p = q = 1/2$.

From our derivation we see that (3.118) is valid for large values of N and for values of n near \bar{n} . The Gaussian approximation is a good approximation even for relatively small values of N for most values of n . A comparison of the Gaussian approximation to the binomial distribution is given in Table 3.4. A discussion of the accuracy of the Gaussian approximation to the binomial distribution is given in Problem 3.76.

The most important feature of the Gaussian probability distribution is that its relative width, σ_n/\bar{n} , decreases as $N^{-1/2}$. The binomial distribution also shares this feature. The alternate derivation of the Gaussian probability distribution in Problem 3.72 shows why the binomial and Gaussian distributions have the same mean and variance.

3.6 Continuous Probability Distributions

In many cases of physical interest the random variables have continuous values. Examples of continuous variables are the positions of the holes left by darts thrown at a dart board, the position and velocity of a particle described by classical mechanics, and the angle of a compass needle.

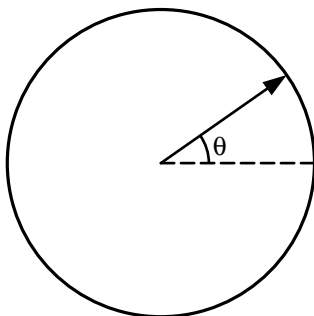


Figure 3.6: The angle θ is an example of a continuous random variable.

As an example, consider a spinner, the equivalent of a wheel of fortune,¹¹ with an arrow that spins around and stops at some angle at random (see Figure 3.6). 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. Thus, 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 angular range $d\theta$ about θ . For example, the probability that θ in Fig. 3.6 is between 0 and π is $1/2$ and the probability that θ is between 0 and $\pi/2$ is $1/4$.

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 perform 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 the bin width $\Delta x = 0.5$ is shown in Figure 3.7. The histogram $H(x)$ is proportional to the estimated probability that a walker lies 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 N_w .

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.

Because we expect the number of walkers in a particular bin to be proportional to the width of the bin, we may write $p(x)\Delta x = H(x)/N_w$. The quantity $p(x)$ is called the *probability density*. In the limit $\Delta x \rightarrow 0$, $H(x)$ becomes a continuous function of x , and we can write the probability that the displacement x of the walker is between a and b as (see Figure 3.8).

$$P(a < x < b) = \int_a^b p(x) dx. \quad (3.119)$$

¹¹The Wheel of Fortune is an American television game. The name of the show comes from the large spinning wheel that determines the dollar amounts and prizes won by the contestants.

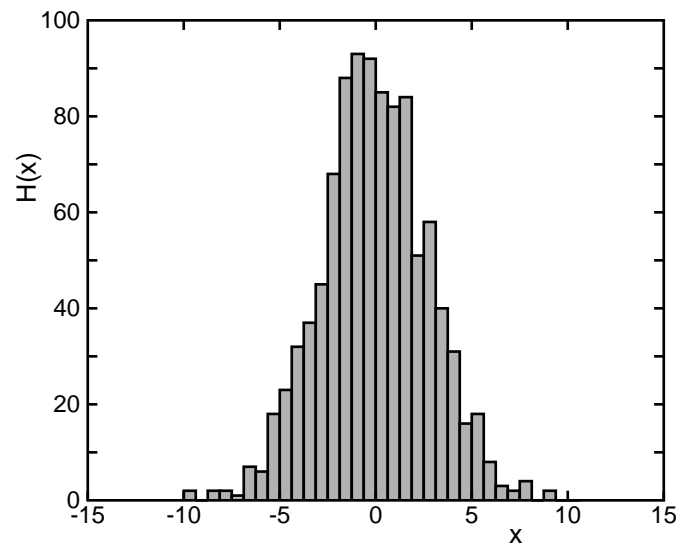


Figure 3.7: Histogram $H(x)$ of the number of times that the displacement of a one-dimensional random walker lies between x and $x + \Delta x$ after $N = 16$ steps (see Problem 3.38). The length of each step is chosen with uniform random probability to be between zero and one. The bin width is $\Delta x = 0.5$. The data was generated with 1000 trials, a relatively small number. The results of this set of trials are the estimates $\bar{x} = -0.045$ and $\overline{x^2} = 4.95$.

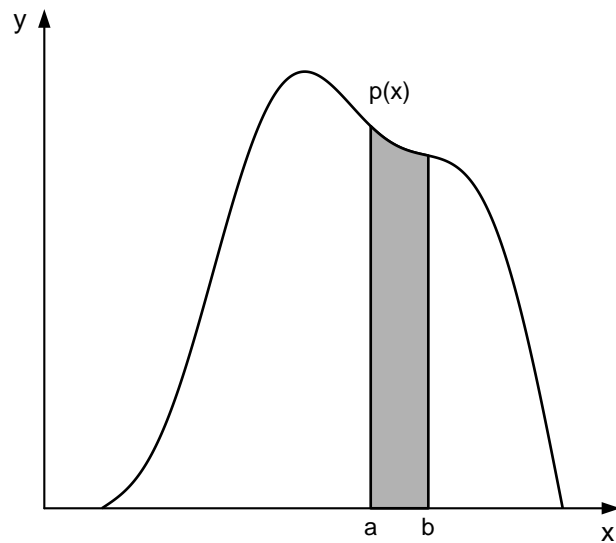


Figure 3.8: The probability that x is between a and b is equal to the shaded area.

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

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

$$\int_{-\infty}^{\infty} p(x) dx = 1. \quad (3.120)$$

The mean value of the function $f(x)$ is given by

$$\bar{f} = \int_{-\infty}^{\infty} f(x) p(x) dx. \quad (3.121)$$

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

The program `RandomWalk1DContinuous` simulates a random walk in one dimension with a variable step length.

- The step length is generated at random with a uniform probability between 0 and 1. Calculate the mean displacement and its variance for one step.
- Compare your analytical results from part (a) to the results of the simulation for $N = 1$.
- How does the variance of the displacement found in the simulation for $N = 16$ depend on the variance of the displacement for $N = 1$ that you calculated in part (a)?
- Explore how the histogram changes with the bin width. What is a reasonable choice of the bin width for $N = 100$? \square

Problem 3.39. Exponential probability density

The random variable x has the probability density

$$p(x) = \begin{cases} A e^{-\lambda x} & (0 \leq x < \infty) \\ 0 & (x < 0). \end{cases} \quad (3.122)$$

The exponential probability density plays an important role in statistical mechanics (see (4.79), page 201).

- Determine the normalization constant A in terms of λ .
- What is the mean value of x ? What is the most probable value of x ?
- What is the mean value of x^2 ?
- Determine the probability for $\lambda = 1$ that a measurement of x yields a value between 1 and 2.
- Determine the probability $\lambda = 1$ that a measurement of x yields a value less than 0.3. \square

Problem 3.40. Probability density for velocity

Consider the probability density function $p(v_x) = (a/\pi)^{3/2} e^{-av_x^2}$ for the velocity of a particle in the x -direction. The probability densities for v_y and v_z have the same form. Each of the three velocity components can range from $-\infty$ to $+\infty$ and a is a constant. This form of the probability density for the velocity will be derived in Section 6.2.2 for a classical system of particles at temperature T .

- (a) Show that $p(\mathbf{v})$ is normalized to one. Use the fact that (see (A.15))

$$\int_0^\infty e^{-au^2} du = \frac{1}{2} \sqrt{\frac{\pi}{a}}. \quad (3.123)$$

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

- (b) 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 $v_z + dv_z$?
- (c) What is the probability that $v_x \geq 0$, $v_y \geq 0$, $v_z \geq 0$ simultaneously? \square

Problem 3.41. Gaussian probability density

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

$$p(x) = (2\pi)^{-1/2} e^{-x^2/2}. \quad (-\infty < x < \infty) \quad (3.124)$$

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

$$C_4 = \overline{x^4} - 4\overline{x^3}\overline{x} - 3\overline{x^2}^2 + 12\overline{x^2}\overline{x}^2 - 6\overline{x}^4. \quad (3.125)$$

\square

Problem 3.42. Uniform probability distribution

Consider the probability density given by

$$p(x) = \begin{cases} (2a)^{-1} & (|x| \leq a) \\ 0 & (|x| > a) \end{cases} \quad (3.126)$$

- (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 (3.125) for the probability density in (3.126). Compare your result to the corresponding result for C_4 for the Gaussian distribution. \square

Problem 3.43. Other probability distributions

Not all probability densities have a finite variance as you will find in the following.

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

$$p(x) = \frac{1}{\pi} \frac{\gamma}{(x-a)^2 + \gamma^2}. \quad (-\infty < x < \infty) \quad (3.127)$$

Choose $a = 0$ and $\gamma = 1$ and compare the form of $p(x)$ in (3.127) to the Gaussian distribution given by (3.124).

- (b) Calculate the first moment of the Lorentz distribution assuming that $a = 0$ and $\gamma = 1$.
 (c) Does the second moment exist? □

3.7 The Central Limit Theorem (or Why Thermodynamics Is Possible)

We have discussed how to estimate probabilities empirically by sampling, that is, by making repeated measurements of the outcome of independent random 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 example.

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

$$S = \sum_{i=1}^N s_i, \quad (3.128)$$

where

$$s_i = \begin{cases} 1, & \text{if the } i\text{th throw gives a 5} \\ 0 & \text{otherwise.} \end{cases} \quad (3.129)$$

The ratio S/N approaches $1/6$ for large N . 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 3.9 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 3.44 we calculate the absolute and relative width of the distributions.

Problem 3.44. Analysis of Figure 3.9

- (a) Estimate the absolute width and the relative width $\Delta S/\bar{S}$ of the distributions shown in Figure 3.9 for $N = 100$ and $N = 800$.
- (b) Does the error of any one measurement of S decrease with increasing N as expected?
- (c) How would the plot change if the number of measurements M were increased to $M = 100,000$? \square

In Section 3.11.2 we show that in the limit $N \rightarrow \infty$, the probability density $p(S)$ is given by

$$p(S) = \frac{1}{\sqrt{2\pi\sigma_S^2}} e^{-(S-\bar{S})^2/2\sigma_S^2} \quad (\text{central limit theorem}) \quad (3.130)$$

where

$$\bar{S} = N\bar{s} \quad (3.131)$$

$$\sigma_S^2 = N\sigma^2, \quad (3.132)$$

with $\sigma^2 = \overline{s^2} - \bar{s}^2$. The quantity $p(S)\Delta S$ is the probability that the value of the sum $\sum_{i=1}^N s_i$ is between S and $S + \Delta S$. Equation (3.130) is equivalent to the central limit theorem. Note that the Gaussian form in (3.130) 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 distribution of the value of the sum of a large number of random variables is approximately a Gaussian. The approximation improves as the number of variables in the sum increases. For the throw of a die we have $\bar{s} = \frac{1}{6}$, $\overline{s^2} = \frac{1}{6}$, and $\sigma^2 = \overline{s^2} - \bar{s}^2 = \frac{1}{6} - \frac{1}{36} = \frac{5}{36}$. For N throws of a die, we have $\bar{S} = N/6$ and $\sigma_S^2 = 5N/36$. We see that in this example the most probable relative error in any one measurement of S decreases as $\sigma_S/\bar{S} = \sqrt{5/N}$.

If we let S represent the displacement of a walker after N steps and let σ^2 equal the mean square displacement of a single step, then the central limit theorem implies that the probability density of the displacement is a Gaussian which is equivalent to the results that we found for random walks in the limit of large N . Or we can let S represent the magnetization of a system of spins and obtain similar results. The displacement of a random walk after N steps and the magnetization of a system of spins are examples of a *random additive process*. Hence, the probability distribution for random walks, spins, and multiple coin tosses is given by (3.130), and our task reduces to finding expressions for \bar{s} and σ^2 for the particular process of interest.

Problem 3.45. Central limit theorem

Use program `CentralLimitTheorem` to test the applicability of the central limit theorem.

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

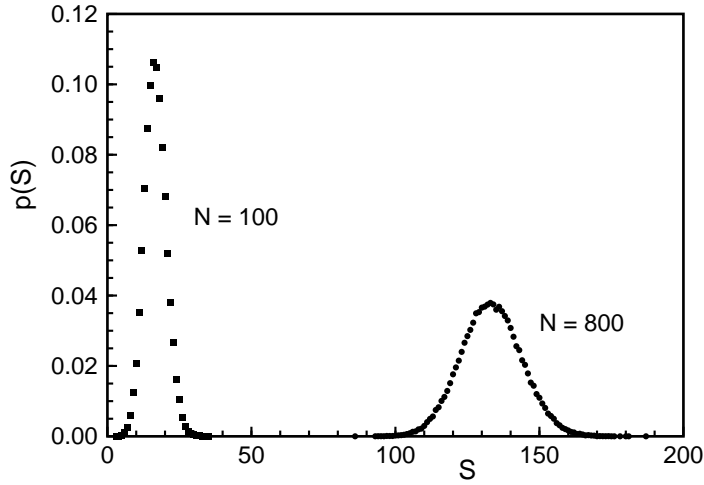


Figure 3.9: The distribution of the measured values of $M = 10,000$ different measurements of the sum S for $N = 100$ and $N = 800$ terms in the sum. The quantity S is the number of times that face 1 appears in N throws of a die. For $N = 100$, the measured values are $\bar{S} = 16.67$, $\overline{S^2} = 291.96$, and $\sigma_S = 3.74$. For $N = 800$, the measured values are $\bar{S} = 133.31$, $\overline{S^2} = 17881.2$, and $\sigma_S = 10.52$. What is the estimated value of the relative width for each case?

- Use the default value of $N = 12$, 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 ?
- 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?
- 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).
- 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)?
- 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

$$\tilde{\sigma}_S^2 = \frac{1}{N-1} \sum_{i=1}^N (S_i - \bar{S})^2. \quad (3.133)$$

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 $\sigma_S^2 = \overline{S^2} - \bar{S}^2$.

- (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 $p(S)$? \square

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.¹²

The central limit theorem implies that macroscopic bodies have well defined macroscopic properties even though their constituent parts are changing rapidly. For example, the particle positions and velocities in a gas or liquid 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, which is a sum of the pressure due to the individual particles, 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 find 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.) \square

3.8 *The Poisson Distribution or Should You Fly?

We now return to the question of whether or not it is safe to fly. If the probability of a plane crashing is $p = 5 \times 10^{-6}$, 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 $N = 1000$, $P_N \approx 0.995$, and for $N = 5 \times 10^5$, $P_N \approx 0.82$. Thus, our intuition is verified that if we took 1000 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.

One way to derive the Poisson distribution is to begin with the binomial distribution:

$$P(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}. \quad (3.134)$$

¹²We will state the central limit theorem more carefully in Section 3.11.2 and note that the theorem holds only if the second moment of the probability distribution of the individual terms in the sum is finite.

We will suppress the N dependence of P . As we have done previously, we will approximate $\ln P(n)$ rather than $P(n)$ directly. We first show that the term $n!/(N-n)!$ can be approximated by N^n in the limit $N \gg n$. We use Stirling's approximation (3.101) (neglecting the logarithmic correction) to write

$$\ln \frac{N!}{(N-n)!} = \ln N! - \ln(N-n)! \quad (3.135a)$$

$$\approx N \ln N - \mathcal{X} - (N-n) \ln(N-n) + \mathcal{X} - n \quad (3.135b)$$

$$\approx \cancel{N \ln N} - \cancel{N \ln N} + n \ln N \quad (3.135c)$$

$$= n \ln N. \quad (3.135d)$$

We next write $\ln(1-p)^{(N-n)} = (N-n) \ln(1-p) \approx -(N-n)p \approx -Np$, and hence $q^{N-n} \approx e^{-pN}$. We then combine these approximations to obtain

$$P(n) \approx \frac{1}{n!} N^n p^n e^{-pN} = \frac{(Np)^n}{n!} e^{-pN}, \quad (3.136)$$

or

$$\boxed{P(n) = \frac{\bar{n}^n}{n!} e^{-\bar{n}}} \quad (\text{Poisson distribution}) \quad (3.137)$$

where $\bar{n} = pN$. The form (3.137) is the Poisson distribution (see Figure 3.10).

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 $\bar{N} = pN$ equals $10^{-5} \times 400 = 0.004$ for $N = 400$ flights and $\bar{N} = 1$ for $N = 10^5$ flights. Thus, the survival probability is $P(0) = e^{-\bar{N}} \approx 0.996$ 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 3.47. Poisson distribution

- Show that the Poisson distribution is properly normalized, and calculate the mean and variance of n (see (A.5)). 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 .
- Plot the Poisson distribution $P(n)$ as a function of n for $p = 0.01$ and $N = 100$. □

3.9 *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 which occur at times t_1, t_2, \dots . Examples of such sequences are the times that a Geiger counter registers a decay of a radioactive nucleus and the times of an accident at a busy intersection. Suppose that we determine the sequence over a very long time T that is much greater than any of the intervals $\tau_i = t_i - t_{i-1}$. We also suppose that the mean number of events is λ per unit time so that in the interval τ , the mean number of events is $\lambda\tau$. We also assume that the events occur at random and are independent of each other.

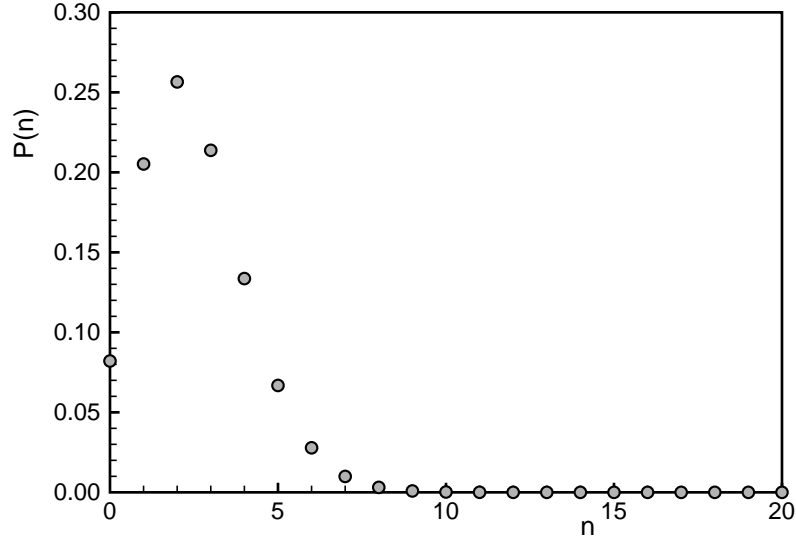


Figure 3.10: Plot of the Poisson probability distribution for $p = 0.0025$ and $N = 1000$.

We wish to find the probability $w(\tau)d\tau$ that the interval between events is between τ and $\tau + d\tau$. If an event occurred at $t = 0$, the probability that at least one other event occurs within the interval $[0, \tau]$ is

$$\int_0^\tau w(\tau') d\tau'. \quad (3.138)$$

The probability that no event occurs in this interval is

$$1 - \int_0^\tau w(\tau') d\tau'. \quad (3.139)$$

Another way of thinking of $w(\tau)$ is that it is the probability that no event occurs in the interval $[0, \tau]$ and then an event occurs within $[\tau, \tau + \Delta\tau]$. Thus,

$$\begin{aligned} w(\tau)\Delta\tau &= \text{probability that no event occurs in the interval } [0, \tau] \\ &\quad \times \text{probability that an event definitely occurs in the interval } [\tau, \tau + \Delta\tau] \\ &= \left[1 - \int_0^\tau w(\tau') d\tau'\right] \lambda \Delta\tau. \end{aligned} \quad (3.140)$$

If we cancel $\Delta\tau$ from each side of (3.140) and differentiate both sides with respect to τ , we find

$$\frac{dw}{d\tau} = -\lambda w, \quad (3.141)$$

so that

$$w(\tau) = Ae^{-\lambda\tau}. \quad (3.142)$$

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

$$\int_0^\infty w(\tau') d\tau' = 1 = A \int_0^\infty e^{-\lambda\tau'} d\tau' = A/\lambda. \quad (3.143)$$

Hence, $w(\tau)$ is the exponential function

$$w(\tau) = \lambda e^{-\lambda\tau}. \quad (3.144)$$

These results for the exponential distribution lead naturally to the Poisson distribution. Let us divide the interval $T \gg 1$ into n smaller intervals $\tau = T/n$. What is the probability that 0, 1, 2, 3, ... events occur in the interval τ ? We will show that the probability that n events occur in the time interval τ is given by the Poisson distribution:

$$P_n(\tau) = \frac{(\lambda\tau)^n}{n!} e^{-\lambda\tau}, \quad (3.145)$$

where we have set $\bar{n} = \lambda\tau$ in (3.137). We first consider the case $n = 0$. If $n = 0$, the probability that no event occurs in the interval τ is (see (3.140))

$$P_{n=0}(\tau) = 1 - \int_0^\tau w(\tau') d\tau' = 1 - \lambda \int_0^\tau e^{-\lambda\tau'} d\tau' = e^{-\lambda\tau}. \quad (3.146)$$

For $n = 1$ there is exactly one event in time interval τ . This event must occur at some time τ' . If it occurs at τ' , then no other event can occur in the interval $[\tau', \tau]$ (otherwise n would not equal 1). Thus, we have

$$P_{n=1}(\tau) = \int_0^\tau w(\tau') P_{n=0}(\tau - \tau') d\tau' \quad (3.147a)$$

$$= \int_0^\tau \lambda e^{-\lambda\tau'} e^{-\lambda(\tau - \tau')} d\tau', \quad (3.147b)$$

where we have used (3.146) with $\tau \rightarrow (\tau - \tau')$. Hence,

$$P_{n=1}(\tau) = \int_0^\tau \lambda e^{-\lambda\tau} d\tau' = (\lambda\tau) e^{-\lambda\tau}. \quad (3.148)$$

If n events are to occur in the interval $[0, \tau]$, the first must occur at some time τ' and exactly $(n - 1)$ must occur in the time $(\tau - \tau')$. Hence,

$$P_n(\tau) = \int_0^\tau \lambda e^{-\lambda\tau'} P_{n-1}(\tau - \tau') d\tau'. \quad (3.149)$$

Equation (3.149) is a recurrence formula that can be used to derive (3.145) by induction. It is easy to see that (3.145) satisfies (3.149) for $n = 1$. As is usual when solving recursion formulas by induction, we assume that (3.145) is correct for $(n - 1)$. We substitute this result into (3.149) and find

$$P_n(\tau) = \lambda^n e^{-\lambda\tau} \int_0^\tau (\tau - \tau')^{n-1} d\tau' / (n - 1)! = \frac{(\lambda\tau)^n}{n!} e^{-\lambda\tau}. \quad (3.150)$$

An application of the Poisson distribution is given in Problem 3.48.

***Problem 3.48.** 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.

N	frequency
0	1
1	7
2	14
3	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.

- (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. \square

3.10 *Are All Probability Distributions Gaussian?

We have discussed *random additive processes* and found that the probability distribution of their sum is a Gaussian for a sufficiently large number of terms. An example of such a process is a one-dimensional random walk 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.¹³ 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 consider the value of the product of the binary sequence of N elements in which the numbers x_1 and x_2 appear independently with probabilities p and q respectively. We write

$$\Pi = x_1 x_1 x_2 x_1 x_2 \dots \quad (3.151)$$

¹³The following discussion is based on an article by Sidney Redner (see the references).

We ask what is $\bar{\Pi}$, the mean value of Π ? To calculate $\bar{\Pi}$ we define $P_N(n)$ 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:

$$P_N(n) = \frac{N!}{n!(N-n)!} p^n q^{N-n}. \quad (3.152)$$

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

$$\bar{\Pi} = \sum_{n=0}^N P_N(n) [x_1^n x_2^{N-n}] = (px_1 + qx_2)^N. \quad (3.153)$$

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

$$\tilde{\Pi} = (x_1^p x_2^q)^N. \quad (3.154)$$

To obtain a better feeling for these results we consider some special cases. For $x_1 = 2$, $x_2 = 1/2$, and $p = q = 1/2$ we have $\bar{\Pi} = (1/4)[x_2^2 + 2x_1x_2 + x_1^2] = (1/4)[4 + 2 + 1/4] = 25/16$ for $N = 2$; for general N we have $\bar{\Pi} = (5/4)^N$. In contrast, the most probable value for $N = 2$ is given by $\tilde{\Pi} = [2^{1/2} \times (1/2)^{1/2}]^2 = 1$; 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 $\bar{\Pi} = 1$ for all N and $\tilde{\Pi} = [2^{1/3} \times (1/2)^{2/3}]^2 = 2^{-2/3}$ for $N = 2$ and $2^{-N/3}$ for any N . We see that $\tilde{\Pi} \neq \bar{\Pi}$ for a random multiplicative process. In contrast, the most probable event is a good approximation to the mean value of the sum of a random additive process (and is identical for $p = q$).

The reason for the large discrepancy between $\bar{\Pi}$ and $\tilde{\Pi}$ 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 $\bar{\Pi}$ and a dominant contribution to the higher moments $\bar{\Pi}^m$.

***Problem 3.49.** A simple multiplicative process

- Confirm the general result in (3.153) for $N = 4$ by showing explicitly all the possible values of the product.
- Consider the case $x_1 = 2$, $x_2 = 1/2$, $p = 1/4$, and $q = 3/4$, and calculate $\bar{\Pi}$ and $\tilde{\Pi}$.
- Show that the mean value of the m th moment $\bar{\Pi}^m = \sum_{n=0}^N P(n) [x_1^n x_2^{N-n}]^m$ reduces to $(px_1^m)^N$ as $m \rightarrow \infty$ (for $x_1 > x_2$). (Hint: Consider the ratio of each term in the sum to the term with x_1^{Nm} .) This result implies that the m th moment is determined solely by the most extreme event for $m \gg 1$.
- Explain why a *log-normal* distribution for which $p(\Pi) \sim e^{-(\ln \Pi - \overline{\ln \Pi})^2 / 2\sigma^2}$ is a reasonable guess for the continuum approximation to the probability of a random multiplicative process for $N \gg 1$. Here $\Pi = x_1^n x_2^{N-n}$. \square

***Problem 3.50.** Simulation of a multiplicative process

- (a) Run the program `MultiplicativeProcess` to simulate 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 $\bar{\Pi}$ and $\tilde{\Pi}$. 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 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

$$\ln T^* = \frac{N}{2pq} \left(p - \frac{pq(x_1/x_2)^m}{q + p(x_1/x_2)^m} \right)^2, \quad (3.155)$$

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 (3.155) compare with the results you observe in the simulation? \square

3.11 *Supplementary Notes

3.11.1 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 $f(x) = x(1 - x^2)$. This approach works only if f can be reduced to a function of one variable. We first consider this case as a way of introducing the general method of undetermined multipliers.

Our goal is to maximize $f(x, y) = xy^2$ 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

$$df - \lambda dg = 0, \quad (3.156)$$

where $df = 0$ at the maximum of f , $dg = 0$ because g expresses the constraint, and λ will be chosen so that (3.156) is satisfied. If we substitute $df = y^2 dx + 2xy dy$ and $dg = 2x dx + 2y dy$ in (3.156), we obtain

$$(y^2 - 2\lambda x) dx + 2(xy - \lambda y) dy = 0. \quad (3.157)$$

We choose $\lambda = y^2/2x$ so that the first term is zero at the maximum. 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 $\lambda = 2$.

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

$$df = \sum_{i=1}^N \frac{\partial f}{\partial x_i} dx_i = 0, \quad (3.158)$$

and the constraints can be expressed as

$$dg_j = \sum_{i=1}^N \frac{\partial g_j}{\partial x_i} dx_i = 0. \quad (3.159)$$

As in our example, we can combine (3.158) and (3.159) and write $df - \sum_{j=1}^M \lambda_j dg_j = 0$, or

$$\sum_{i=1}^N \left[\frac{\partial f}{\partial x_i} - \sum_{j=1}^M \lambda_j \frac{\partial g_j}{\partial x_i} \right] dx_i = 0. \quad (3.160)$$

We are free to choose all M values of λ_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

$$\frac{\partial f}{\partial x_i} - \sum_{j=1}^M \lambda_j \frac{\partial g_j}{\partial x_i} = 0. \quad (3.161)$$

In Example 3.11 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 – 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 is \bar{n} . We wish to determine the values of P_1, P_2, \dots, P_6 that maximize the uncertainty S subject to the constraints

$$\sum_{j=1}^6 P_j = 1, \quad (3.162)$$

$$\sum_{j=1}^6 jP_j = \bar{n}. \quad (3.163)$$

For a perfect die $\bar{n} = 3.5$. We take

$$f = S = - \sum_{j=1}^6 P_j \ln P_j \quad (3.164a)$$

$$g_1 = \sum_{j=1}^6 jP_j - 1, \quad (3.164b)$$

and

$$g_2 = \sum_{j=1}^6 jP_j - \bar{n}. \quad (3.164c)$$

We have $\partial f/\partial P_j = -(1 + \ln P_j)$, $\partial g_1/\partial P_j = 1$, and $\partial g_2/\partial P_j = j$, and write (3.161) for $j = 1$ and $j = 2$ as

$$-(1 + \ln P_1) - \alpha - \beta = 0 \quad (3.165a)$$

$$-(1 + \ln P_2) - \alpha - 2\beta = 0, \quad (3.165b)$$

where we have taken α and β (instead of λ_1 and λ_2 as the undetermined Lagrange multipliers. The solution of (3.165) for α and β is

$$\alpha = \ln P_2 - 2 \ln P_1 - 1 \quad (3.166a)$$

$$\beta = \ln P_1 - \ln P_2. \quad (3.166b)$$

We solve (3.166b) for $\ln P_2 = \ln P_1 - \beta$ and use (3.166a) to find $\ln P_1 = -1 - \alpha - \beta$. We then use this result to write $\ln P_2$ as $\ln P_2 = -1 - \alpha - 2\beta$. We can independently vary dP_3, \dots, dP_6 because the two constraints are satisfied by the values of P_1 and P_2 . Hence, we have from (3.161) and (3.164) that

$$\ln P_j = -1 - \alpha - j\beta, \quad (3.167)$$

or

$$P_j = e^{-1-\alpha} e^{-\beta j}. \quad (3.168)$$

We eliminate the constant α by the normalization condition (3.162) and write:

$$P_j = \frac{e^{-\beta j}}{\sum_j e^{-\beta j}}. \quad (3.169)$$

The constant β is determined by the constraint (3.45):

$$\bar{n} = \frac{e^{-\beta} + 2e^{-2\beta} + 3e^{-3\beta} + 4e^{-4\beta} + 5e^{-5\beta} + 6e^{-6\beta}}{e^{-\beta} + e^{-2\beta} + e^{-3\beta} + e^{-4\beta} + e^{-5\beta} + e^{-6\beta}}. \quad (3.170)$$

Usually, (3.170) must be solved numerically.

The exponential form (3.169) will become very familiar to us (see (4.79), page 201) and is known as the Boltzmann distribution. In the context of thermal systems the Boltzmann distribution maximizes the uncertainty given the constraints that the probability distribution is normalized and the mean energy is known.

Problem 3.51. Numerical solution of (3.170)

Show that the solution to (3.170) is $\beta = 0$ for $\bar{n} = 7/2$, $\beta = +\infty$ for $\bar{n} = 1$, $\beta = -\infty$ for $\bar{n} = 6$, and $\beta = -0.1746$ for $\bar{n} = 4$. \square

3.11.2 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 $p(x)$:

$$\phi(k) = \overline{e^{ikx}} = \int_{-\infty}^{\infty} e^{ikx} p(x) dx. \quad (3.171)$$

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:

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikx} \phi(k) dk. \quad (3.172)$$

Problem 3.52. Characteristic function of a Gaussian

Calculate the characteristic function of the Gaussian probability density. \square

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

$$\phi(k) = \sum_{n=0}^{\infty} \frac{k^n}{n!} \left. \frac{d^n \phi(k)}{dk^n} \right|_{k=0}, \quad (3.173)$$

$$= \overline{e^{ikx}} = \sum_{n=0}^{\infty} \frac{(ik)^n \overline{x^n}}{n!}. \quad (3.174)$$

By comparing coefficients of k^n in (3.173) and (3.174), we see that

$$\overline{x} = -i \left. \frac{d\phi}{dk} \right|_{k=0}. \quad (3.175)$$

In Problem 3.53 we show that

$$\overline{x^2} - \overline{x}^2 = - \left. \frac{d^2}{dk^2} \ln \phi(k) \right|_{k=0} \quad (3.176)$$

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

Problem 3.53. The first few cumulants

The characteristic function generates the *cumulants* C_m defined by

$$\ln \phi(k) = \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} C_n. \quad (3.177)$$

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

$$C_1 = \overline{x} \quad (3.178a)$$

$$C_2 = \sigma^2 = \overline{x^2} - \overline{x}^2 \quad (3.178b)$$

$$C_3 = \overline{x^3} - 3\overline{x^2}\overline{x} + 2\overline{x}^3 \quad (3.178c)$$

$$C_4 = \overline{x^4} - 4\overline{x^3}\overline{x} - 3\overline{x^2}^2 + 12\overline{x^2}\overline{x}^2 - 6\overline{x}^4. \quad (3.178d)$$

The first few cumulants were calculated for several probability distributions in Problems 3.41(b) and 3.42(c). What is the value of C_4 for the Gaussian distribution? \square

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

$$z = x + y. \quad (3.179)$$

How do the probability densities combine? The answer is given by

$$p(z) = \int p_1(x)p_2(y) \delta(z - x - y) dx dy. \quad (3.180)$$

The integral in (3.180) 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 (3.180) 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

$$\phi_z(k) = \int e^{ikz} p(z) dz \quad (3.181a)$$

$$= \iiint e^{ikz} p_1(x)p_2(y) \delta(z - x - y) dx dy dz \quad (3.181b)$$

$$= \int e^{ikx} p_1(x) dx \int e^{iky} p_2(y) dy \quad (3.181c)$$

$$= \phi_1(k)\phi_2(k). \quad (3.181d)$$

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

$$S = x_1 + x_2 + \dots + x_N. \quad (3.182)$$

Then

$$\phi_S(k) = \prod_{i=1}^N \phi_i(k). \quad (3.183)$$

That is, *the characteristic function of the sum of several independent variables is the product of the individual characteristic functions*. If we take the logarithm of both sides of (3.183), we obtain

$$\ln \phi_S(k) = \sum_{i=1}^N \ln \phi_i(k). \quad (3.184)$$

Each side of (3.184) 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 n th order cumulants satisfy the relation:

$$C_n^S = C_n^1 + C_n^2 + \dots + C_n^N. \quad (3.185)$$

We conclude that if the random variables x_i are independent (uncorrelated), their cumulants and in particular, their variances, add. We saw a special case of this result for the variance in (3.75).

If we denote the mean and standard deviation of the weight of an adult male as \bar{w} and σ respectively, then from (3.178a) and (3.185) we find that the mean weight of N adult males is given by $N\bar{w}$. Similarly from (3.178b) we see that the standard deviation of the weight of N adult males is given by $\sigma_N^2 = N\sigma_w^2$, or $\sigma_N = \sqrt{N}\sigma_w$. 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, \dots, 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 $\bar{x} = 0$, 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

$$p(S) = \frac{1}{\sqrt{2\pi N\sigma_x^2}} e^{-S^2/2N\sigma_x^2} \quad (3.186)$$

From (3.178b) 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 :

$$\overline{\left(\frac{x_1 + x_2 + \dots + x_N}{N}\right)^2} = \frac{N\sigma_x^2}{N^2} = \frac{\sigma_x^2}{N}. \quad (3.187)$$

From (3.187) we see that it is useful to define a scaled sum:

$$z = \frac{1}{\sqrt{N}}(x_1 + x_2 + \dots + x_N), \quad (3.188)$$

and to write the central limit theorem in the form

$$p(z) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-z^2/2\sigma_x^2}. \quad (3.189)$$

To obtain the result (3.189), we write the characteristic function of z as

$$\phi_z(k) = \iiint \dots \int e^{ikz} \delta\left(z - \left[\frac{x_1 + x_2 + \dots + x_N}{N^{1/2}}\right]\right) \quad (3.190a)$$

$$\times p(x_1) p(x_2) \dots p(x_N) dz dx_1 dx_2 \dots dx_N \quad (3.190b)$$

$$= \iint \dots \int e^{ik(x_1 + x_2 + \dots + x_N)/N^{1/2}} p(x_1) p(x_2) \dots p(x_N) dx_1 dx_2 \dots dx_N \quad (3.190c)$$

$$= \phi\left(\frac{k}{N^{1/2}}\right)^N. \quad (3.190d)$$

We next take the logarithm of both sides of (3.190d) and expand the right-hand side in powers of k to find

$$\ln \phi_z(k) = \sum_{m=2}^{\infty} \frac{(ik)^m}{m!} N^{1-m/2} C_m. \quad (3.191)$$

The $m = 1$ term does not contribute in (3.191) because we have assumed that $\bar{x} = 0$. More importantly, note that as $N \rightarrow \infty$, the higher-order terms are suppressed so that

$$\ln \phi_z(k) \rightarrow -\frac{k^2}{2} C_2, \quad (3.192)$$

or

$$\phi_z(k) \rightarrow e^{-k^2 \sigma^2 / 2} + \dots \quad (3.193)$$

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

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

The leading correction to $\phi(k)$ in (3.194) gives rise to a term of order $N^{-1/2}$, and therefore does not contribute in the limit $N \rightarrow \infty$.

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. It is not necessary that all the x_i have the same distribution. Not all probabilities have a finite second moment as demonstrated by the Lorentz distribution (see Problem 3.43), but the requirements for the central limit theorem are weak and the central limit theorem is widely applicable.

Vocabulary

sample space, events, outcome
 uncertainty, principle of least bias or maximum uncertainty
 probability distribution $P(i)$ or P_i , probability density $p(x)$
 mean value $\overline{f(x)}$, moments, variance $\overline{\Delta x^2}$, standard deviation σ
 conditional probability $P(A|B)$, 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, random multiplicative processes
 cumulants, characteristic function

Additional problems

Problem 3.54. Probability that a site is occupied

In Figure 3.11 we show a square lattice of 16^2 sites each of which is occupied with probability p . Estimate the probability p that a site in the lattice is occupied and explain your reasoning. \square

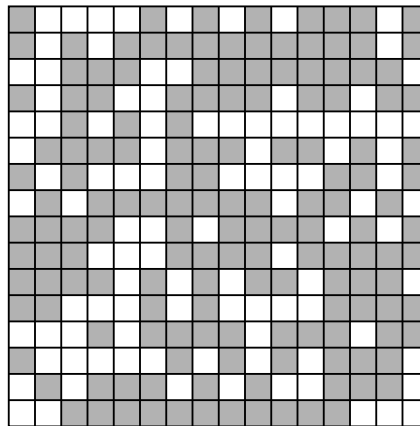


Figure 3.11: Representation of a square lattice of 16×16 sites. The sites are represented by squares. Each site is either occupied (shaded) independently of its neighbors with probability p or is empty (white) with probability $1 - p$. These configurations are discussed in the context of percolation in Section 9.3.

Problem 3.55. Three coins (in a fountain)

Three coins are tossed in succession. Assume that landing heads or tails is 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 3.56. A student's 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 reasoning, 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 3.57. d'Alembert's fallacious reasoning

- (a) What is the probability that heads will appear at least once in two tosses of a single coin? Use the rules of probability to show that the answer is $\frac{3}{4}$.
- (b) 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 3.58. Number of fish in a pond

A farmer wants to estimate how many fish are in a pond. The farmer 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, the farmer removes 250 fish at random and finds that 25 of them are tagged. Estimate the number of fish in the pond. \square

Problem 3.59. Estimating the area of a pond

A farmer owns a field that is $10\text{ m} \times 10\text{ m}$. 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? \square

Problem 3.60. Monte Carlo integration

Consider the ten pairs of numbers, (x_i, y_i) , given in Table 3.6. The numbers are all in the range $0 < x_i, y_i \leq 1$. 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

$$F = \int_0^1 dx \sqrt{1 - x^2}. \quad (3.195)$$

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*.¹⁴

- Show analytically that the integral in (3.195) is equal to $\pi/4$.
- Use the program `MonteCarloEstimation` to estimate the integral (3.195) by Monte Carlo integration. Determine the error (the magnitude of the deviation from the exact answer) for trials of n pairs of points equal to $n = 10^4$, 10^6 , and 10^8 . Does the error decrease with increasing n on the average?
- Estimate the integral using $n = 1000$. Repeat for a total of 10 trials using a different random number seed each time. The easiest way to do so is to press the Reset button and then press the Calculate button. The default is for the program to choose a new seed each time based on the time. Is the magnitude of the variation of your values of the same order as the error between the average value and the exact value? For a large number of trials, the error is estimated from the standard error of the mean, which approximately equals the standard deviation divided by the square root of the number of trials. \square

Problem 3.61. Bullseye

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

¹⁴Monte Carlo methods were first developed to estimate integrals that could not be performed by analytically or by the usual numerical methods.

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 3.6: A sequence of ten random pairs of numbers.

- (a) once in five throws;
- (b) twice in ten throws. Why are these probabilities not identical? □

Problem 3.62. Family values

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 3.63. Fathers and sons (and daughters)

What is the probability that five children produced by the same couple will consist of the following (assume that the probabilities of giving birth to a boy and a girl are the same):

- (a) three sons and two daughters?
- (b) alternating sexes?
- (c) alternating sexes starting with a son?
- (d) all daughters? □

Problem 3.64. Probability in baseball

A good hitter in major league baseball has a batting average of 300, which means that the hitter will be successful 3 times out of 10 tries on the average. Assume that the batter has 4 times at bat per game.

- (a) What is the probability that he gets no 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 3.65. Playoff winners

- (a) 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 games. 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 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.
- (b) Most teams have better records at home. Assume the two teams are evenly matched and each has a 60% chance of winning at home and a 40% change of winning away. In principle, both teams should have an equal chance of winning a seven game series. Determine which pattern of home games is closer to giving each team a 50% chance of winning. Consider the two common patterns: (1) 2 home, 3 away, 2 home; and (2) 2 home, 2 away, 1 home, 1 away, 1 home. \square

Problem 3.66. Galton board

The Galton board (named after Francis Galton (1822–1911)), is a triangular array of pegs. The rows are numbered $0, 1, \dots$ 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. \square

Problem 3.67. The birthday problem

What if somebody offered to bet you 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 this case than in part (a). \square

Problem 3.68. 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. A sequence of numbers produced in this manner is identical to a random walk, yet the sequence frequently appears to be correlated. The lesson of the charts is that our eyes look for patterns even when none exists. \square

Problem 3.69. Displacement and number of steps to the right

- (a) 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 $\sigma_m^2 = \overline{(m - \overline{m})^2} = 4Npq$.
- (b) The result (3.78) for $\overline{(\Delta M)^2}$ differs by a factor of four from the result for σ_n^2 in (3.99). Why? \square

Problem 3.70. 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. Calculate \overline{n} , $\overline{n^2}$, and σ_n . You can use this information to make two estimates of p , the probability of a step to the right. If you obtain different estimates 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. Calculate the same quantities as in part (a) and use this information to estimate p . How does the ratio of σ_n to \overline{n} compare for the two values of N ? Explain your results.
- (c) Calculate \overline{m} and σ_m , where $m = n - n'$ is the net displacement of the walker for parts (a) and (b).

This problem inspired an article by Zia and Schmittmann (see the references). \square

Problem 3.71. Consider the binomial distribution $P_N(n)$ for $N = 16$ and $p = q = 1/2$.

- (a) What is the value of $P_N(n)$ at $n = \overline{n} - \sigma_n$?
- (b) What is the value of the product $P_N(n = \overline{n})(2\sigma_n)$? \square

Problem 3.72. Alternative derivation of the Gaussian distribution

On page 138 we evaluated the binomial probability $P_N(n)$ using Stirling's approximation to determine the parameters A , B , and \tilde{n} in (3.109). Another way to determine these parameters is to approximate the binomial distribution by a Gaussian and require that the zeroth, first, and second moments of the Gaussian and binomial distribution be equal. We write

$$P(n) = Ae^{-B(n-\tilde{n})^2/2}, \quad (3.196)$$

where A , B , and \tilde{n} are the parameters to be determined. We first require that

$$\int_0^N P(n) dn = 1. \quad (3.197)$$

Because $P(n)$ depends on the difference $n - \tilde{n}$, it is convenient to change the variable of integration in (3.197) to $x = n - \tilde{n}$ and write

$$\int_{-Np}^{N(1-p)} P(x) dx = 1, \quad (3.198)$$

where

$$P(x) = Ae^{-Bx^2/2}. \quad (3.199)$$

Because we are interested in the limit $N \rightarrow \infty$, we can extend the limits in (3.198) to $\pm\infty$:

$$\int_{-\infty}^{\infty} P(x) dx = 1. \quad (3.200)$$

(a) The first moment of the Gaussian distribution is

$$\bar{n} = \int_{-\infty}^{\infty} nP(n) dn, \quad (3.201)$$

where $P(n)$ is given by (3.196). Make a change of variables and show that

$$\bar{n} = \int_{-\infty}^{\infty} (x + \tilde{n})P(x) dx = \tilde{n}. \quad (3.202)$$

(b) The first moment of the binomial distribution is given by pN according to (3.96). Show that if we require the first moments of the binomial and Gaussian distributions to be equal, we obtain $\tilde{n} = pN$.

(c) The variance of the binomial distribution is given in (3.99) and is equal to $\overline{(n - \bar{n})^2} = Npq$. The corresponding variance of the Gaussian distribution is given by

$$\overline{(n - \bar{n})^2} = \int_{-\infty}^{\infty} (n - \bar{n})^2 P(n) dn. \quad (3.203)$$

Make the necessary change of variables in (3.203) and do the integrals in (3.200) and (3.203) (see (A.23) and (A.17)) to confirm that the values of B and A are given by (3.115) and (3.117), respectively.

(d) Explain why the third moments of the binomial and Gaussian distribution are not equal. \square

Problem 3.73. A simple two-dimensional wall

Consider a two-dimensional “wall” constructed from N squares as shown in Figure 3.12. The base row of the wall must be continuous, but higher rows can have gaps. Each column must be continuous and self-supporting with no overhangs. Determine the total number W_N of different N -site clusters, that is, the number of possible arrangements of N squares consistent with these rules. Assume that the squares are identical. \square

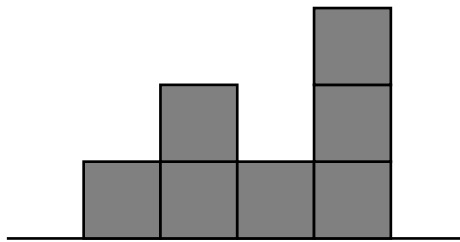


Figure 3.12: Example of a wall as explained in Problem 3.73.

Problem 3.74. 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 3.75. First passage time

Suppose that a one-dimensional unbiased random walker starts out at the origin $x = 0$ at $t = 0$ and takes unit length steps at regular intervals. As usual the probability of a step to the right is p .

- (a) How many steps will it take for the walker to first reach $x = +1$? This quantity, known as the *first passage time*, is a random variable because it is different for different realizations of the walk. Let P_n be the probability that x first equals $+1$ after n steps. What is P_n for $n = 1, 3, 5$, and 7 ?
- (b) Write a program to simulate a random walker in one dimension and determine the number of steps needed to first reach $x = 1$. What is the probability that the walker will eventually reach $x = +1$ assuming that $p = 1/2$? What is the mean number of steps needed to reach $x = +1$? □

***Problem 3.76.** 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 after (3.114) in the power series expansion of $\ln P(n)$:

$$\frac{1}{3!}(n - \tilde{n})^3 C + \frac{1}{4!}(n - \tilde{n})^4 D, \quad (3.204)$$

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

- (a) Show that $|C| < 1/N^2 p^2 q^2$. What does C equal if $p = q$?

(b) Show that $|D| < 4/N^3 p^3 q^3$.

(c) Show that the results for $|C|$ and $|D|$ imply that the neglect of terms beyond second order in $(n - \tilde{n})$ is justified if $|n - \tilde{n}| \ll Npq$. Explain why stopping at second order is justified if $Npq \gg 1$. \square

Problem 3.77. A Lévy flight

A Lévy flight, named after the mathematician Paul Pierre Lévy, is a random walk in which the length ℓ of each step is distributed according to a probability distribution of the form $p(\ell) \propto \ell^{-\mu}$, where $1 < \mu < 3$. Is the form of the probability distribution of the displacement of the walker after N steps a Gaussian? \square

Problem 3.78. Balls and boxes

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? \square

Problem 3.79. Telephone numbers

Open a telephone directory to a random page or look at the phone numbers in your cell phone and make a list corresponding to the last digit of the first 100 telephone numbers you see. Find the probability $P(n)$ that the number n appears in your list. Plot $P(n)$ as a function of n and describe its n -dependence. Do you expect that $P(n)$ is approximately uniform? \square

***Problem 3.80.** Benford's law or looking for number one

Suppose that you constructed a list of the populations of the largest cities in the world, or a list of the house numbers of everybody you know. Other naturally occurring lists include river lengths, mountain heights, radioactive decay half lives, the size of the files on your computer, and the first digit of each of the numbers that you find in a newspaper. (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 $n = 1, \dots, 9$? 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

$$P(n) = \log_{10} \left(1 + \frac{1}{n} \right). \quad (3.205)$$

The distribution (3.205) is known as *Benford's law* and is named after Frank Benford, a physicist, who independently discovered it in 1938, although it was discovered previously by the astronomer Simon Newcomb in 1881. The distribution (3.205) 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?

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

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 3.7: Ranking of the top 20 words.

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). \square

***Problem 3.81.** Faking it

Ask several of your friends to flip a coin 100 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? \square

***Problem 3.82.** Zipf's law

Suppose that we analyze a text and count the number of times a given word appears. The word with rank r is the r th 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 for a given text

$$f \sim \frac{1}{r \ln(1.78R)}, \quad (3.206)$$

where R is the number of different words. Note the inverse power law behavior of the frequency on the rank. The relation (3.206) 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 3.7. \square

***Problem 3.83.** Time of response to emails

When 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 hours between receiving an email from someone and replying to it.

It turns out that the time it takes people to reply to emails can be described by a power law; that is, the probability $p(\tau)d\tau$ that the response is between τ and $\tau + d\tau$ 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? \square

Problem 3.84. Pick any card

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? \square

***Problem 3.85.** Societal response to rare events

- (a) 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?
- (b) The likelihood of the breakdown of the levees near New Orleans was well known before its 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. \square

***Problem 3.86.** 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/ to read about interesting issues involving probability and statistics. \square

Suggestions for further reading

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

Ralph Baierlein, *Atoms and Information Theory*, W. H. Freeman (1971) derives the Boltzmann distribution using arguments similar to those used to obtain (3.169).

Arieh Ben-Naim, *Entropy Demystified: The Second Law Reduced to Plain Common Sense*, World Scientific (2007).

Marta C. González, César A. Hidalgo, and Albert-László Barabási, “Understanding individual human mobility patterns,” *Nature* **453**, 779–782 (2008). The authors studied the trajectories of 100,000 cell phone users over a six-month period and found that human trajectories cannot be simply modeled by a Lévy flight or as an ordinary random walk. Similar studies have been done on animal trajectories. The website barabasilab.com/ has many examples of the application of probability to diverse systems of interest in statistical 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, the first several chapters are a good read for students who are 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 importance of rare events is by Nassim Nicholas Taleb, *The Black Swan: The Impact of the Highly Improbable*, Random House (2007).

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.

F. N. David, *Games, Gods and Gambling: A History of Probability and Statistical Ideas*, Dover Publications (1998).

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.

Leonard Mlodinow, *The Drunkard’s Walk: How Randomness Rules Our Lives*, Vintage Press (2009). A popular book on how the mathematical laws of randomness affect our lives.

Elliott W. Montroll and Michael F. Shlesinger, “On the wonderful world of random walks,” in *Studies in Statistical Mechanics*, Vol. XI: Nonequilibrium Phenomena II, edited by J. L. Lebowitz and E. W. Montroll North-Holland (1984). An excellent article on the history of random walks.

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).

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), 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).

S. Redner, “Random multiplicative processes: An elementary tutorial,” *Am. J. Phys.* **58**, 267–273 (1990).

Jason Rosenhouse, *The Monty Hall Problem: The Remarkable Story of Math’s Most Contentious Brain Teaser*, Oxford University Press (2009).

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 m dash. How long do such 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.

J. Torres, S. Fernández, A. Gamero, and A Sola, “How do numbers begin? (The first digit law),” *Eur. J. Phys.* **28**, L17–L25 (2007).

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).

Hans Christian von Baeyer, *Information: The New Language of Science*, Harvard University Press (2004). This book raises many profound issues. It is not an easy read even though it is well written.

Charles A. Whitney, *Random Processes in Physical Systems: An Introduction to Probability-Based Computer Simulations*, John Wiley & Sons (1990).

Michael M. Woolfson, *Everyday Probability and Statistics*, Imperial College Press (2008). An interesting book for lay people.

A 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 3.70.