Chapter 3

Concepts of Probability

12 October 2004

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

3.1 Probability in everyday life

Chapter 2 provided an introduction to thermodynamics using macroscopic arguments. Our eventual goal, which we will begin in Chapter 4, is to relate the behavior of various macroscopic quantities to the underlying microscopic behavior of atoms and molecules. To do so, we will need to introduce some ideas from probability theory.

We all use the ideas of probability in everyday life. For example, every morning many of us decide what to wear based on the probability of rain. We cross streets knowing that the probability of being hit by a car is small, but it is much higher than the sun not rising tomorrow. How do we know? Because you probably know people who have been hit by cars, but have not observed the sun not rise. We can even make rough estimates of the probability of being hit by a car. It must be less that one in a thousand, because you have crossed streets thousands of times and hopefully you have not been hit. Of course 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 logic of probabilities ...*

However, our intuition only takes us so far. Consider airplane travel. Is it safe to fly? Suppose that there is a one chance in 100,000 of a plane crashing on a given flight and that there are a 1000 flights a day. Then every 100 days or so there would be a reasonable likelihood of a plane crashing. This estimate is in rough accord with what we read. For a given flight, your chances of crashing are approximately one part in 10^5, and if you fly five times a year for 80 years, it seems that flying is not too much of a risk. However, suppose that instead of living 80 years, you could live 20,000 years. In this case you would take 100,000 flights, and it would be much more risky to fly if you wished to live your full 20,000 years. Although this last statement seems reasonable, can you explain why?

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

Given the long history of games of chance and the interest in estimating probability in a variety of contexts, it is remarkable that the theory of probability took so long to develop. One reason is that the idea of probability is subtle and is capable of many interpretations. An understanding of probability is elusive due in part to the fact that it cannot be measured at one time, but can only be estimated by a series of measurements. Although the rules of probability are defined by simple mathematical rules, an understanding of probability is greatly aided by experience with real data and concrete problems. To test your current understanding of probability, try to solve Problems 3.1–3.6 before reading the rest of this chapter. Then in Problem 3.7 formulate the laws of probability as best as you can based on your solutions to these problems.

Problem 3.1. 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 the marble is orange, (b) the marble is red, (c) the marble is orange or blue.

Problem 3.2. 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. A girl tosses a pair of dice at the same time. Find the probability that (a) both dice show the same number, (b) both dice show a number less than 5, (c) both dice show an even number, (d) the product of the numbers is 12.

Problem 3.4. A boy hits 16 free throws out of 25 attempts. What is the probability that he will make a free throw on his next attempt?

Problem 3.5. Consider an experiment in which a die is tossed 150 times and the number of times each face is observed is counted. The value of A, the number of dots on the face of die, and the number of times that it appeared is shown in Table 3.1. What is the predicted average value of A? What is the average value of A observed in this experiment?

Problem 3.6. 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.

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

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

Problem 3.8. Suppose that you are offered a choice of the following: a certain \$10 or the chance of rolling a die and receiving \$36 if it comes up 5 or 6, but nothing otherwise. Make arguments in favor of each choice.

Problem 3.9. Suppose that you are offered the following choice: (a) a prize of \$100 if a flip of a coin yields heads or (b) a certain prize of \$40. What choice would you make? Explain your reasoning.

Problem 3.10. Suppose that you are offered the following choice: (a) a prize of \$100 is awarded for each head found in ten flips of a coin or (b) a certain prize of \$400. What choice would you make? Explain your reasoning.

Problem 3.11. Suppose that someone gives you a dollar to play the lottery. What sequence of six numbers between 1 and 36 would you choose?

Problem 3.12. Suppose you toss a coin 8 times and obtain heads each time. Estimate the probability that you will obtain heads on your ninth toss.

Problem 3.13. What is the probability that it will rain tomorrow? What is the probability that the Dow Jones industrial average will increase tomorrow?

Problem 3.14. Give several examples of the use of probability in everyday life. Distinguish between various types of probability.

3.2 The rules of probability

We now summarize the basic rules and ideas of probability.[1] Suppose that there is an operation or a process that has several distinct possible *outcomes*. The process might be the flip of a coin or the

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

roll of a six-sided die.[2] We call each flip a *trial*. The list of all the possible *outcomes* or *events* is called the *sample space*. We assume that the events are *mutually exclusive*, that is, the occurrence of one outcome 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.

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

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

and

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

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

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

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

The rules of probability will be summarized further in (3.3) and (3.4). 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.

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

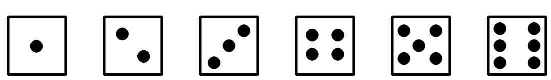

Figure 3.1: The six possible outcomes of the toss of a die.

Suppose that we know that the probability of rolling any face of a die in one throw is equal to 1/6, and we want to find the probability of finding face 3 *or* face 6 in one throw. In this case

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

we wish to know the probability of a trial that is a combination of more elementary operations for which the probabilities are already known. That is, we want to know the probability of the outcome, i or j, where i is distinct from j. According to the rules of probability, the probability of event i *or* j is given by

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

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

Example 3.2. What is the probability of throwing a three or a six with one throw of a die?
Solution. The probability that the face exhibits either 3 or 6 is $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Example 3.3. What is the probability of *not* throwing a six with one throw of die?
Solution. The answer is the probability of either "1 or 2 or 3 or 4 or 5." The addition rule gives that the probability $P(\text{not six})$ is

$$\begin{aligned} P(\text{not six}) &= P(1) + P(2) + P(3) + P(4) + P(5) \\ &= 1 - P(6) = \frac{5}{6}, \end{aligned}$$

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

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

$$P(i \text{ and } j) = P(j)\, P(j). \qquad \text{(multiplication rule)} \tag{3.4}$$

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

To understand the applicability of (3.4) and the meaning of the independence of events, consider the problem of determining the probability that a person chosen at random is a female over six feet tall. Suppose that we know that the probability of a person to be over six feet tall is $P(6^+) = \frac{1}{5}$, and the probability of being female is $P(\text{female}) = \frac{1}{2}$. We might conclude that the probability of being a tall female is $P(\text{female}) \times P(6^+) = \frac{1}{2} \times \frac{1}{5} = \frac{1}{10}$. The same probability would hold for a tall male. However, this reasoning is incorrect, because the probability of being a tall female differs from the probability of being a tall male. The problem is that the two events – being over six feet tall and being female – are not independent. On the other hand, consider the probability that a person chosen at random is female and was born on September 6. We can reasonably assume equal likelihood of birthdays for all days of the year, and it is correct to conclude that this probability is $\frac{1}{2} \times \frac{1}{365}$ (not counting leap years). Being a woman and being born on September 6 are independent events.

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

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

Example 3.5. What is the probability of the same face appearing on two successive throws of a die?
Solution. We know that the probability of any specific combination of outcomes, for example, (1,1), (2,2), ... (6,6) is $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$. Hence, by the addition rule

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

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

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

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

$$\begin{aligned} P(6,6) &= \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \\ P(6,\text{not } 6) &= P(\text{not } 6,6) = \frac{1}{6} \times \frac{5}{6} = \frac{5}{36} \\ P(\text{not } 6,\text{not } 6) &= \frac{5}{6} \times \frac{5}{6} = \frac{25}{36}. \end{aligned}$$

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

$$\begin{aligned} P(\text{at least one } 6) &= P(6,6) + P(6,\text{not } 6) + P(\text{not } 6,6) \\ &= \frac{1}{36} + \frac{5}{36} + \frac{5}{36} = \frac{11}{36}. \end{aligned}$$

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

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

Example 3.7. What is the probability of obtaining at least one six in four throws of a die?
Solution. We know that in one throw of a die, there are two outcomes with $P(6) = \frac{1}{6}$ and $P(\text{not } 6) = \frac{5}{6}$. Hence, in four throws of a die there are sixteen possible outcomes, only one of which has no six. That is, in the fifteen mutually exclusive outcomes, there is at least one six. We can use the multiplication rule (3.3) to find that

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

and hence

$$\begin{aligned} P(\text{at least one six}) &= 1 - P(\text{not } 6, \text{not } 6, \text{not } 6, \text{not } 6) \\ &= 1 - \big(\frac{5}{6}\big)^4 \\ &= \frac{671}{1296} \approx 0.517. \end{aligned}$$

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

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

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

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

Problem 3.15. Find the probability distribution $P(n)$ for throwing a sum n with two dice and plot $P(n)$ as a function of n.

Problem 3.16. What is the probability of obtaining at least one double six in twenty-four throws of a pair of dice?

Problem 3.17. Suppose that three die are thrown at the same time. What is the probability that the sum of the three faces is 10 compared to 9?

Problem 3.18. 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.

Problem 3.19. In two tosses of a single coin, what is the probability that heads will appear at least once? Use the rules of probability to show that the answer is $\frac{3}{4}$. However, d'Alembert, a distinguished French mathematician of the eighteenth century, reasoned that there are only 3 possible outcomes: heads on the first throw, heads on the second throw, and no heads at all. The first two of these three outcomes is favorable. Therefore the probability that heads will appear at least once is $\frac{2}{3}$. What is the fallacy in this reasoning?

3.3 Mean values

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

$$\overline{x} \equiv x_1 P(1) + x_2 P(2) + \ldots + x_n P(n) \tag{3.6a}$$

$$= \sum_{i=1}^{n} x_i P(i), \tag{3.6b}$$

where $P(i)$ is the probability of x_i, and we have assumed that the probability is normalized. If $f(x)$ is a function of x, then the mean value of $f(x)$ is defined by

$$\overline{f(x)} = \sum_{i=1}^{n} f(x_i) P(i). \tag{3.7}$$

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

$$\begin{aligned}\overline{f(x) + g(x)} &= \sum_{i=1}^{n} [f(x_i) + g(x_i)] P(i) \\ &= \sum_{i=1}^{n} f(x_i) P(i) + \sum_{i=1}^{n} g(x_i) P(i),\end{aligned}$$

or

$$\overline{f(x) + g(x)} = \overline{f(x)} + \overline{g(x)}. \tag{3.8}$$

Problem 3.20. Show that if c is a constant, then

$$\overline{cf(x)} = c\overline{f(x)}. \tag{3.9}$$

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

$$\overline{x^m} \equiv \sum_{i=1}^{n} {x_i}^m P(i), \tag{3.10}$$

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

Problem 3.21. Suppose that the variable x takes on the values -2, -1, 0, 1, and 2 with probabilities 1/4, 1/8, 1/8, 1/4, and 1/4, respectively. Calculate the first two moments of x.

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

$$\Delta x \equiv x - \overline{x}, \tag{3.11}$$

and

$$\overline{\Delta x} = \overline{(x - \overline{x})} = \overline{x} - \overline{x} = 0. \tag{3.12}$$

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

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

$$\overline{\Delta x^2} \equiv \overline{\big(x - \overline{x}\big)^2}. \tag{3.13}$$

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

$$\begin{aligned} \overline{\big(x - \overline{x}\big)^2} &= \overline{\big(x^2 - 2x\overline{x} + \overline{x}^2\big)} \\ &= \overline{x^2} - 2\overline{x}\,\overline{x} + \overline{x}^2, \end{aligned} \tag{3.14}$$

or

$$\overline{\big(x - \overline{x}\big)^2} = \overline{x^2} - \overline{x}^2. \tag{3.15}$$

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

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

$$\sigma_x = \sqrt{\overline{\Delta x^2}} = \sqrt{\big(\overline{x^2} - \overline{x}^2\big)}. \tag{3.16}$$

Example 3.9. Find the mean value $\overline{x}$, the variance $\overline{\Delta x^2}$, and the standard deviation σ_x for the value of a single throw of a die.

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

$$\begin{aligned} \overline{x} &= \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2} = 3.5 \\ \overline{x^2} &= \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{46}{3} \\ \overline{\Delta x^2} &= \overline{x^2} - \overline{x}^2 = \frac{46}{3} - \frac{49}{4} = \frac{37}{12} \approx 3.08 \\ \sigma_x &\approx \sqrt{3.08} = 1.76 \end{aligned}$$

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

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

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

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

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

$$\begin{aligned} m &= p + 2pq + 3pq^2 + 4pq^3 + \cdots \\ &= p(1 + 2q + 3q^2 + \cdots) \\ &= p\frac{d}{dq}(1 + q + q^2 + q^3 + \cdots) \\ &= p\frac{d}{dq}\frac{1}{1-q} = \frac{p}{(1-q)^2} = \frac{1}{p} \end{aligned} \tag{3.17}$$

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

3.4 Kinds of probability

How can we assign the probability? The answer depends on the information available to us. Either we need to make an independent assumption for the probabilities based on mathematical symmetry for example, or we need to estimate the probability empirically. The empirical method is called *sampling* and is equivalent to performing a large number of measurements of a *single* system under identical conditions and counting how often each event occurs. Let M be the number of measurements and M_i represent the number of times that the outcome i occurs. If we observe that the ratio M_i/M tends to a fixed value as M becomes larger and larger, we estimate the probability $P(i)$ as

$$P(i) \approx \frac{M_i}{M}. \qquad (M \text{ is number of measurements}) \tag{3.18}$$

Note that the estimate of the probability in (3.18) is consistent with (3.1) and (3.2).

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

Problem 3.22. Use the Java applet/application at <stp.clarku.edu/simulations/cointoss> to simulate multiple tosses of a single coin. What is the correspondence between this simulation

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

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

of a coin being tossed many times and the actual physical tossing of a coin? If the coin is "fair" (the random number generator is good), what do you think the ratio of the number of heads to the total number of tosses will be? Do you obtain this number after 100 tosses? 10,000 tosses?

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

$$P(i) \approx \frac{M_i}{M}. \qquad (M \text{ is the number of systems in the ensemble}) \tag{3.19}$$

If the system of interest is not changing in time, it is reasonable that a determination 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.

Problem 3.23. Use the Java applet/application at <stp.clarku.edu/simulations/cointoss> to simulate a single toss of many coins at one time. If you toss 100 coins, do you expect to obtain exactly 50 heads? What do you find if you toss 100 coins many times? Do you think that the simulation corresponnds to a fair coin?

Some people believe (erroneously in most cases) that if an event has not occurred for a while, then the probability increases for that event. For example, if you throw ten heads in a row, then it would seem to many that the probability of throwing a tail in the next toss should be greater than 1/2. However, if we do a careful analysis of the way we throw the coin and can physically determine that there is no bias for tossing heads or tails, then we can determine the probability of tossing tails to be 1/2 independent of previous tosses.

The above discussion 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 are using the term probability in its own definition. If we do an experiment to measure the frequencies of outcomes, then we are forced to make two additional assumptions: (1) we have to assume that the measured frequencies will be the same in the future as they were in the past, and (2) we have to take a large number of measurements to insure accuracy, and we have no way of knowing a priori how many measurements are sufficient. Thus, the definition of probability as a frequency really turns out to be a method for estimating probabilities with some hidden assumptions.

However, we can give an intuitive definition that works for all ways of making estimates. This definition, which is sometimes called *subjective probability*, is that *probability is a measure of the degree of belief in the occurrence of an outcome*. This definition contains a very important concept, namely, that probability depends on our prior knowledge, because clearly belief depends on prior knowledge. For example, we might use the knowledge of 100 tails in a row as evidence that the coin or toss is biased, and thus estimate that the probability of throwing another tail is very high. However, if a careful physical analysis shows that there is no bias, then we would stick to our estimate of 1/2. The probability depends on what knowledge we bring to the problem. If we have no knowledge at all except what the possible outcomes are, 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 3.24. (a) Suppose that you know that a couple has two children, at least one of whom is a girl. What is the probability that the other child is a girl? (b) Suppose that instead we know that the oldest child is a girl. What is the probability that the youngest is a girl?

From the above discussion and Problem 3.24, we see that all probability determinations are really examples of *conditional probability*. Let us define $P(A|B)$ as the probability of A occurring given that we know B. Sometimes B is something we have to assume. We now discuss a few results about conditional probability. Clearly,

$$P(A) = P(A|B) + P(A|\overline{B}), \tag{3.20}$$

where $\overline{B}$ means B does not occur. Also, it is clear that

$$P(\text{A and B}) = P(A|B)P(B) = P(B|A)P(A), \tag{3.21}$$

Equation (3.21) means that the probability that A and B occur equals the probability that A occurs given B times the probability that B occurs, which is the same as the probability that B occurs given A times the probability A that occurs. If we are interested in various possible outcomes A_i for the same B, we can rewrite (3.21) as

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}. \tag{3.22}$$

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

$$P(B) = \sum_i P(B|A_i)P(A_i). \tag{3.23}$$

If we substitute (3.23) for $P(B)$ into (3.22), we obtain the important result:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)}. \qquad \text{(Bayes' theorem)} \tag{3.24}$$

Equation 3.24 is known as Bayes' theorem.

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

As an example, consider the following quandary known as the Monty Hall Problem.[3] In this show a contestant is shown three doors. Behind one door is an expensive gift such as a car and behind the other two doors are inexpensive gifts such as a tie. The contestant chooses a door. Suppose she chooses door 1. Then the host opens door 2 containing the tie. The contestant now has a choice – should she stay with her original choice or switch to door 3? What would you do?

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

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

and

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

where A_i denotes car behind door i. We know that all the $P(A_i)$ equal 1/3, because with no information we must 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

$$P(\text{door 2 open after door 1 chosen}|\text{car behind 1}) = 1/2 \tag{3.25a}$$
$$P(\text{door 2 open after door 1 chosen}|\text{car behind 2}) = 0 \tag{3.25b}$$
$$P(\text{door 2 open after door 1 chosen}|\text{car behind 3}) = 1. \tag{3.25c}$$

Using Bayes' theorem we have

$$P(\text{car behind 1}|\text{door 2 open after door 1 chosen}) = \frac{(1/2)(1/3)}{\frac{1}{2}\frac{1}{3} + 0\frac{1}{3} + 1\frac{1}{3}} = 1/3 \tag{3.26a}$$
$$P(\text{car behind 3}|\text{door 2 open after door 1 chosen}) = \frac{(1)(1/3)}{\frac{1}{2}\frac{1}{3} + 0\frac{1}{3} + 1\frac{1}{3}} = 2/3. \tag{3.26b}$$

The results in (3.26) suggest the contestant has a higher probability of winning the car if she switches doors and chooses door 3. The same logic suggests that she should always switch doors independently of which door she originally chose. A search of the internet for Monty Hall will bring you to many sites that discuss the problem in more detail.

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

Example 3.11. 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. We will also assume that if you do not have the disease, the test will come out negative 98% of the time and positive 2% of the time. You take the test and it comes out positive. What is the probability that you have the disease? Is this test useful?

Solution. Let $P(+|D) = 0.98$ represent the probability of testing positive and having the disease. If $\overline{D}$ represents not having the disease and $-$ represents testing negative, then we are given: $P(-|D) = 0.02$, $P(-|\overline{D}) = 0.98$, $P(+|\overline{D}) = 0.02$, $P(D) = 0.0001$, and $P(\overline{D}) = 0.9999$. From Bayes' theorem we have

$$\begin{aligned} P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|\overline{D})P(\overline{D})} \\ &= \frac{(0.98)(0.0001)}{(0.98)(0.0001) + (0.02)(0.9999)} \\ &= 0.0047 = 0.47\%. \end{aligned} \tag{3.27}$$

Problem 3.25. Imagine that you have a sack of 3 balls that can be either red or green. There are four hypotheses for the distribution of colors for the balls: (1) all are red, (2) 2 are red, (3) 1 is red, and (4) all are green. Initially, you have no information about which hypothesis is correct, and thus you assume that they are equally probable. Suppose that you pick one ball out of the sack and it is green. Use Bayes' theorem to determine the new probabilities for each hypothesis.

Problem 3.26. Make a table that determines the necessary accuracy for a test to give the probability of having a disease if tested positive equal to at least 50% for diseases that occur in 1 in 100, 1 in 1000, 1 in 10,000, and 1 in 100,000 people.

3.5 Binomial distribution

Because most physicists spend little time gambling,[4] we will have to develop our intuitive understanding of probability in other ways. Our strategy will be to first consider some physical systems for which we can calculate the probability distribution by analytical methods. Then we will use the computer to generate more data to analyze.

Noninteracting magnetic moments

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

[4]After a Las Vegas hotel hosted a meeting of the American Physical Society in March, 1986, the physicists were asked not to return.

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

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

$$M = \mu(s_1 + s_2 + \ldots + s_N) = \mu \sum_{i=1}^{N} s_i. \tag{3.28}$$

To compute the mean value of M, we need to take the mean values of both sides of (3.28). In the following, we will take $\mu = 1$ for convenience whenever it will not cause confusion. Alternatively, we can interpret M as the net number of up spins. If we use (3.8), we can interchange the sum and the average and write

$$\overline{M} = \overline{\Big(\sum_{i=1}^{N} s_i\Big)} = \sum_{i=1}^{N} \overline{s}_i. \tag{3.29}$$

Because the probability that any spin has the value ± 1 is the same for each spin, the mean value of each spin is the same, that is, $\overline{s}_1 = \overline{s}_2 = \ldots = \overline{s}_N \equiv \overline{s}$. Therefore the sum in (3.29) consists of N equal terms and can be written as

$$\overline{M} = N\overline{s}. \tag{3.30}$$

The meaning of (3.30) is that the mean magnetization is N times as large as the mean magnetic moment of a single spin. Because $\overline{s} = (1 \times p) + (-1 \times q) = p - q$, we have that

$$\overline{M} = N(p - q). \tag{3.31}$$

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

$$\Delta M = M - \overline{M} = \sum_{i=1}^{N} \Delta s_i, \tag{3.32}$$

where

$$\Delta s_i \equiv s_i - \overline{s}. \tag{3.33}$$

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

$$\begin{aligned}(\Delta M)^2 &= (\Delta s_1 + \Delta s_2 + \Delta s_3)(\Delta s_1 + \Delta s_2 + \Delta s_3) \\ &= \big[(\Delta s_1)^2 + (\Delta s_2)^2 + (\Delta s_3)^2\big] + 2\big[\Delta s_1 \Delta s_2 + \Delta s_1 \Delta s_3 + \Delta s_2 \Delta s_3\big].\end{aligned} \tag{3.34}$$

We take the mean value of (3.34), interchange the order of the sums and averages, and write

$$\overline{(\Delta M)^2} = \big[\overline{(\Delta s_1)^2} + \overline{(\Delta s_2)^2} + \overline{(\Delta s_3)^2}\big] + 2\big[\overline{\Delta s_1 \Delta s_2} + \overline{\Delta s_1 \Delta s_3} + \overline{\Delta s_2 \Delta s_3}\big]. \tag{3.35}$$

The first term on the right of (3.35) represents the three terms in the sum that are multiplied by themselves. The second term represents all the cross terms arising from different terms in the sum, that is, the products in the second sum refer to different spins. Because different spins are statistically independent (the spins do not interact), we have that

$$\overline{\Delta s_i\,\Delta s_j} = \overline{\Delta s_i}\;\overline{\Delta s_j} = 0, \qquad (i \neq j) \tag{3.36}$$

because $\overline{\Delta s_i} = 0$. That is, each cross term vanishes on the average. Hence (3.36) reduces to a sum of squared terms

$$\overline{(\Delta M)^2} = \big[\overline{(\Delta s_1)^2} + \overline{(\Delta s_2)^2} + \overline{(\Delta s_3)^2}\big]. \tag{3.37}$$

Because each spin is equivalent on the average, each term in (3.37) is equal. Hence, we obtain the desired result

$$\overline{(\Delta M)^2} = 3\overline{(\Delta s)^2}. \tag{3.38}$$

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

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

$$\begin{aligned}\overline{(\Delta s)^2} &= \overline{s^2} - \overline{s}^2 = 1 - (p-q)^2 = 1 - (2p-1)^2 \\ &= 1 - 4p^2 + 4p - 1 = 4p(1-p) = 4pq,\end{aligned} \tag{3.39}$$

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

$$\overline{(\Delta M)^2} = 3(4pq). \tag{3.40}$$

Problem 3.27. Use similar considerations to show that for $N = 3$ that

$$\overline{n} = 3p \tag{3.41}$$

and

$$\overline{(n - \overline{n})^2} = 3pq. \tag{3.42}$$

Explain the difference between (3.31) and (3.41) for $N = 3$, and the difference between (3.40) and (3.42).

Problem 3.28. Use similar considerations for N noninteracting spins and show that

$$\overline{M} = N(p-q), \tag{3.43}$$

and

$$\overline{(\Delta M)^2} = N(4pq). \tag{3.44}$$

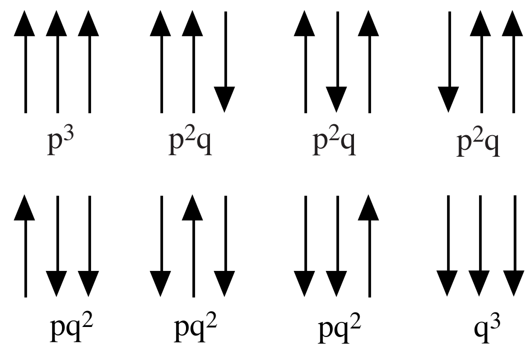

Figure 3.2: An ensemble of $N = 3$ spins. The arrow indicates the direction of the magnetic moment of a spin. The probability of each member of the ensemble is shown.

Because of the simplicity of a system of noninteracting spins, we can calculate the probability distribution itself, and not just the first few moments. As an example, let us consider the statistical properties of a system of $N = 3$ noninteracting spins. Because each spin can be in one of two states, there are $2^{N=3} = 8$ distinct outcomes (see Figure 3.2). What is the probability of each member of the ensemble? Because each spin is independent of the other spins, we can use the multiplication rule (3.4) to calculate the probabilities of each outcome as shown in Figure 3.2. Although each member of the ensemble is distinct, several of the configurations have the same number of up spins. One quantity of interest is the probability $P_N(n)$ that n spins are up out a total of N spins. For example, there are three states with $n = 2$, each with probability p^2q so the probability that two spins are up is equal to $3p^2q$. For $N = 3$ we see from Figure 3.2 that

$$P_3(n = 3) = p^3 \tag{3.45a}$$
$$P_3(n = 2) = 3p^2q \tag{3.45b}$$
$$P_3(n = 1) = 3pq^2 \tag{3.45c}$$
$$P_3(n = 0) = q^3. \tag{3.45d}$$

Example 3.12. Find the first two moments of $P_3(n)$.

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

$$\begin{aligned}\overline{n} &= 0 \times q^3 + 1 \times 3pq^2 + 2 \times 3p^2q + 3 \times p^3 \\ &= 3p\,(q^2 + 2pq + p^2) = 3p\,(q + p)^2 = 3p.\end{aligned} \tag{3.46}$$

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

$$\begin{aligned}\overline{n^2} &= 0 \times q^3 + 1 \times 3pq^2 + 4 \times 3p^2q + 9 \times p^3 \\ &= 3p\,(q^2 + 4pq + 3p^2) = 3p(q + 3p)(q + p) \\ &= 3p\,(q + 3p) = (3p)^2 + 3pq.\end{aligned}$$

Hence

$$\overline{(n - \overline{n})^2} = \overline{n^2} - \overline{n}^2 = 3pq. \tag{3.47}$$

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

Problem 3.29. 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 $q = 1 - p = 1/2$. What is the probability that in three tosses of a coin, there will be two heads?

Problem 3.30. *One-dimensional random walk*. The original statement of the *random walk* problem was posed by Pearson in 1905. If a drunkard begins at a lamp post and takes N steps of equal length in random directions, how far will the drunkard be from the lamp post? We will consider an idealized example of a random walk for which the steps of the walker are restricted to a line (a one-dimensional random walk). Each step is of equal length a, and at each interval of time, the walker either takes a step to the right with probability p or a step to the left with probability $q = 1 - p$. The direction of each step is independent of the preceding one. Let n be the number of steps to the right, and n' the number of steps to the left. The total number of steps $N = n + n'$. What is the form of the probability that the walker has taken n steps to the right after N steps?

Problem 3.31. What is the probability that a random walker in one dimension has taken three steps to the right out of four steps?

From the above examples and problems, we see that the probability distributions of noninteracting magnetic moments, the flip of a coin, and a random walk are identical. These examples have two characteristics in common. First, in each trial there are only *two* outcomes, for example, up or down, heads or tails, and right or left. Second, the result of each trial is independent of all previous trials, for example, the drunken sailor has no memory of his or her previous steps. This type of process is called a *Bernoulli* process.[5]

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

$$P_N(n) = W_N(n, n')\, p^n q^{n'}, \tag{3.48}$$

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 with $(n - 1)$ up spins and n' down spins, or

(b) down with n up spins and n' down spins.

[5] These processes are named after the mathematician Jacob Bernoulli, 1654 – 1705.

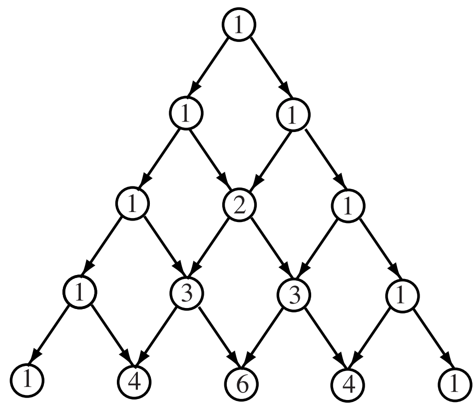

Figure 3.3: The values of the first few coefficients $W_N(n,n')$. Each number is the sum of the two numbers to the left and right above it. This construction is called a Pascal triangle.

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

$$W_N(n,n') = W_{N-1}(n-1,n') + W_{N-1}(n,n'-1). \tag{3.49}$$

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

$$W_2(2,0) = W_1(1,0) + W_1(2,-1) = 1+0 = 1. \tag{3.50a}$$
$$W_2(1,1) = W_1(0,1) + W_1(1,0) = 1+1 = 2. \tag{3.50b}$$
$$W_2(0,2) = W_1(-1,2) + W_1(0,1) = 0+1. \tag{3.50c}$$

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

It is straightforward to show by induction that the expression

$$W_N(n,n') = \frac{N!}{n!\,n'!} = \frac{N!}{n!(N-n)!} \tag{3.51}$$

satisfies the relation (3.49). Note the convention $0!=1$. We can combine (3.48) and (3.51) to find the desired result

$$P_N(n) = \frac{N!}{n!\,(N-n)!}\,p^n q^{N-n}. \qquad \text{(binomial distribution)} \tag{3.52}$$

The form (3.52) is called the *binomial distribution*. Note that for $p=q=1/2$, $P_N(n)$ reduces to

$$P_N(n) = \frac{N!}{n!\,(N-n)!}2^{-N}. \tag{3.53}$$

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

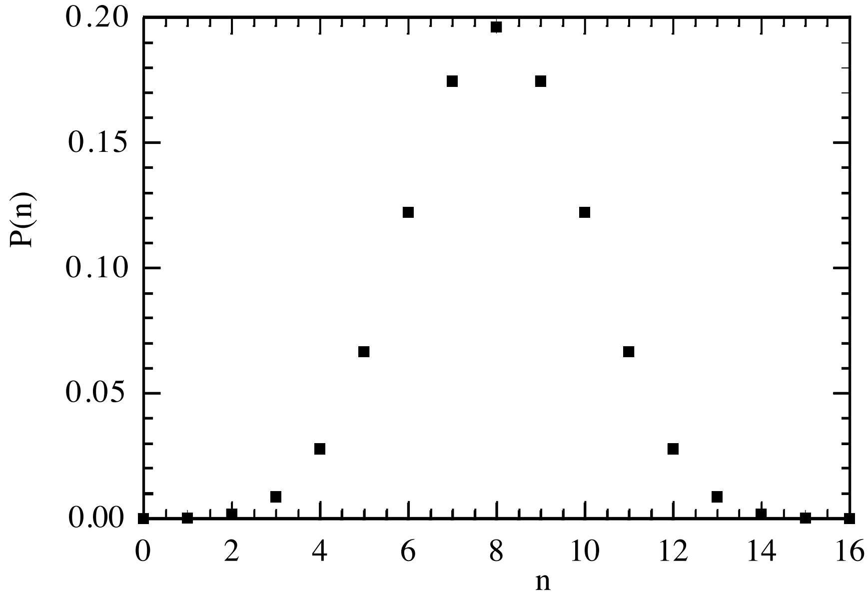

Figure 3.4: The binomial distribution $P_{16}(n)$ for $p = q = 1/2$ and $N = 16$. What is your visual estimate for the width of the distribution?

Problem 3.32. Explain why the binomial distribution applies to all Bernoulli processes.

Problem 3.33. (a) Plot the distribution $P_N(n)$ that n spins are up out of a total of N for $N = 4$ and $N = 16$. Assume $p = q = 1/2$. Visually estimate the width of the distribution for each value of N. What is the qualitative dependence of the width on N? Also compare the relative heights of the maximum of P_N. (b) Calculate the mean values of n and n^2 using your tabulated values of $P_N(n)$. (c) Plot $P_N(n)$ as a function of $n/\overline{n}$ for $N = 4$ and $N = 16$ on the same graph as in (a). Visually estimate the width of the distribution for each value of N. (d) Plot $\ln P_N(n)$ versus $n/\overline{n}$ for $N = 16$. Describe the behavior of $\ln P_N(n)$. Can $\ln P_N(n)$ be fitted to a parabola? (e) 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 (b)? (f) For what value of p and q do you think the width is a maximum for a given N?

Example 3.13. Show that the expression (3.52) for $P_N(n)$ satisfies the normalization condition (3.2).

Solution. The reason that (3.52) is called the binomial distribution is that its form represents a typical term in the expansion of $(p+q)^N$. By the binomial theorem we have

$$(p+q)^N = \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!}\, p^n\, q^{N-n}. \tag{3.54}$$

We use (3.52) and write

$$\sum_{n=0}^{N} P_N(n) = \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n} = (p+q)^N = 1^N = 1, \tag{3.55}$$

where we have used (3.54) and the fact that $p+q=1$.

Calculation of the mean value

We now find an analytical expression for the dependence of $\overline{n}$ on N and p. From the definition (3.6) and (3.52) we have

$$\overline{n} = \sum_{n=0}^{N} n\, P_N(n) = \sum_{n=0}^{N} n\, \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n}. \tag{3.56}$$

We evaluate the sum in (3.56) by using a technique that is useful in a variety of contexts.[6] The technique is based on the fact that

$$p\, \frac{\partial}{\partial p} p^n = n p^n. \tag{3.57}$$

We use (3.57) to rewrite (3.56) as

$$\overline{n} = \sum_{n=0}^{N} n \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n} \tag{3.58}$$

$$= \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!} \big(p\frac{\partial}{\partial p} p^n\big) q^{N-n}. \tag{3.59}$$

We interchange the order of summation and differentiation in (3.59) and write

$$\overline{n} = p\frac{\partial}{\partial p}\Big[\sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n}\Big] \tag{3.60}$$

$$= p\frac{\partial}{\partial p}\,(p+q)^N, \tag{3.61}$$

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

$$\overline{n} = pN(p+q)^{N-1}. \tag{3.62}$$

The result (3.62) is valid for arbitrary p and q, and hence it is applicable for $p+q=1$. Thus our desired result is

$$\overline{n} = pN. \tag{3.63}$$

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

The dependence of $\overline{n}$ on N and p should be intuitively clear. Compare the general result (3.63) to the result (3.46) for $N = 3$. What is the dependence of $\overline{n'}$ on N and p?

Calculation of the relative fluctuations

To determine $\overline{\Delta n^2}$ we need to know $\overline{n^2}$ (see the relation (3.15)). The average value of n^2 can be calculated in a manner similar to that for $\overline{n}$. We write

$$\begin{aligned}
\overline{n^2} &= \sum_{n=0}^{N} n^2 \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n} && (3.64)\\
&= \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!} \big(p \frac{\partial}{\partial p}\big)^2 p^n q^{N-n} \\
&= \big(p \frac{\partial}{\partial p}\big)^2 \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n} = \big(p \frac{\partial}{\partial p}\big)^2 (p+q)^N \\
&= \big(p \frac{\partial}{\partial p}\big) \big[pN(p+q)^{N-1}\big] \\
&= p\big[N(p+q)^{N-1} + pN(N-1)(p+q)^{N-2}\big]. && (3.65)
\end{aligned}$$

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

$$\begin{aligned}
\overline{n^2} &= p\,[N + pN(N-1)] \\
&= p\,[pN^2 + N(1-p)] = (pN)^2 + p\,(1-p)N \\
&= \overline{n}^2 + pqN, && (3.66)
\end{aligned}$$

where we have used (3.63) and let $q = 1 - p$. Hence, from (3.66) we find that the variance of n is given by

$$\sigma_n{}^2 = \overline{(\Delta n)^2} = \overline{n^2} - \overline{n}^2 = pqN. \tag{3.67}$$

Compare the calculated values of σ_n from (3.67) with your estimates in Problem 3.33 and to the exact result (3.47) for $N = 3$.

The relative width of the probability distribution of n is given by (3.63) and (3.67)

$$\frac{\sigma_n}{\overline{n}} = \frac{\sqrt{pqN}}{pN} = \big(\frac{q}{p}\big)^{\frac{1}{2}} N^{-\frac{1}{2}}. \tag{3.68}$$

We see that the relative width goes to zero as $N^{-1/2}$.

Problem 3.34. Frequently we need to evaluate $\ln N!$ for $N >> 1$. What is the largest value of $\ln N!$ that you can calculate exactly using a typical hand calculator? A simple approximation for $\ln N!$, known as *Stirling's approximation*, is

$$\ln N! \approx N \ln N - N \qquad N \to \infty. \tag{3.69}$$

A more accurate approximation is given by

$$\ln N! \approx N \ln N - N + \frac{1}{2}\ln(2\pi N). \tag{3.70}$$

A simple derivation of Stirling's approximation is given in Appendix A. Compare the approximations (3.69) and (3.70) to each other and to the exact value of $\ln N!$ for $N = 5$, 10, 20, and 50. If necessary, compute $\ln N!$ directly using the relation

$$\ln N! = \sum_{m=1}^{N} \ln m. \tag{3.71}$$

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

$$\frac{d}{dx}\ln x! = \ln x. \tag{3.72}$$

Use the more accurate form of Stirling's approximation to find the maximum value of $P_N(n)$ for $p = q = 1/2$. Plot $\ln P_N(n)$ versus n for $N = 100$.

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

Problem 3.36. 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 nearly independent of all other molecules. Although the density will be uniform on the average, there are fluctuations in the density. Divide the volume V into two parts V_1 and V_2, where $V = V_1 + V_2$. (a) What is the probability p that a particular molecule is in each part? (b) What is the probability that N_1 molecules are in V_1 and N_2 molecules are in V_2? (c) What is the average number of molecules in each part? (d) What are the relative fluctuations of the number of particles in each part?

Problem 3.37. Suppose that a random walker takes n steps to the right and n' steps to the left and each step is of equal length a. Denote x as the net displacement of a walker. What is the mean value $\overline{x}$ for a N-step random walk? What is the analogous expression for the dispersion $\overline{(\Delta x)^2}$?

Problem 3.38. *Monte Carlo simulation.* We can gain more insight into the nature of the Bernoulli distribution by doing a Monte Carlo simulation, that is, by using a computer to "flip coins" and average over many measurements.[7] In the context of random walks, we can implement 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
```

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

The function `rnd` generates a random number between zero and one. The quantity x is the net displacement assuming that the steps are of unit length. It is necessary to save the value of x after N steps and average over many walkers. Write a simple program or use the applet/application at <stp.clarku.edu/simulations/OneDimensionalWalk> to compute $P_N(x)$. First choose $N = 4$ and $p = 1/2$ and make sure that the simulation converges to the exact distribution as the number of measurements is increased. Then take $N = 100$ and describe the qualitative x-dependence of $P_N(x)$.

3.6 Continuous probability distributions

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

For continuous variables, the probability of obtaining a particular value is not meaningful. For example, consider a one-dimensional random walker who steps at random to the right or to the left with equal probability, but with step lengths that are chosen at random between zero and a maximum length a. The continuous nature of the length of each step implies that the position x of the walker is a continuous variable. Because there are an infinite number of possible x values in a finite interval of x, the probability of obtaining any particular value of x is zero. Instead, we have to reformulate the question and ask for the probability that the position of the walker is between x and $x + \Delta x$ after N steps. If we do a simulation of such a walker, we would record the number of times, $H(x, \Delta x)$, that a walker is in a bin of width Δx a distance x from the origin, and plot the histogram $H(x, \Delta x)$ as a function of x (see Figure 3.5). If the number of walkers that is sampled is sufficiently large, $H(x, \Delta x)$ is proportional to the estimated probability that a walker is in a bin of width Δx a distance x from the origin after N steps. To obtain the probability, we divide $H(x, \Delta x)$ by the total number of walkers.

In practice, the choice of the bin width is a compromise. If Δx is too big, the features of the histogram would be lost. If Δx is too small, there would be more bins and many of the bins would be empty for a finite sample of walkers. Hence, our estimate of the number of walkers in each bin would be less accurate. Because we expect the number to be proportional to the width of the bin, we can write $H(x, \Delta x)$ as $H(x, \Delta x) = p(x)\,\Delta x$. The quantity $p(x)$ is the *probability density*. In the limit that $\Delta x \to 0$, $H(x, \Delta x)$ becomes a continuous function of x, and we can write the probability that a walker is in the range between a and b as

$$P(a \text{ to } b) = \int_a^b p(x)\,dx. \tag{3.73}$$

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

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

$$\int_{-\infty}^{\infty} p(x)\,dx = 1. \tag{3.74}$$

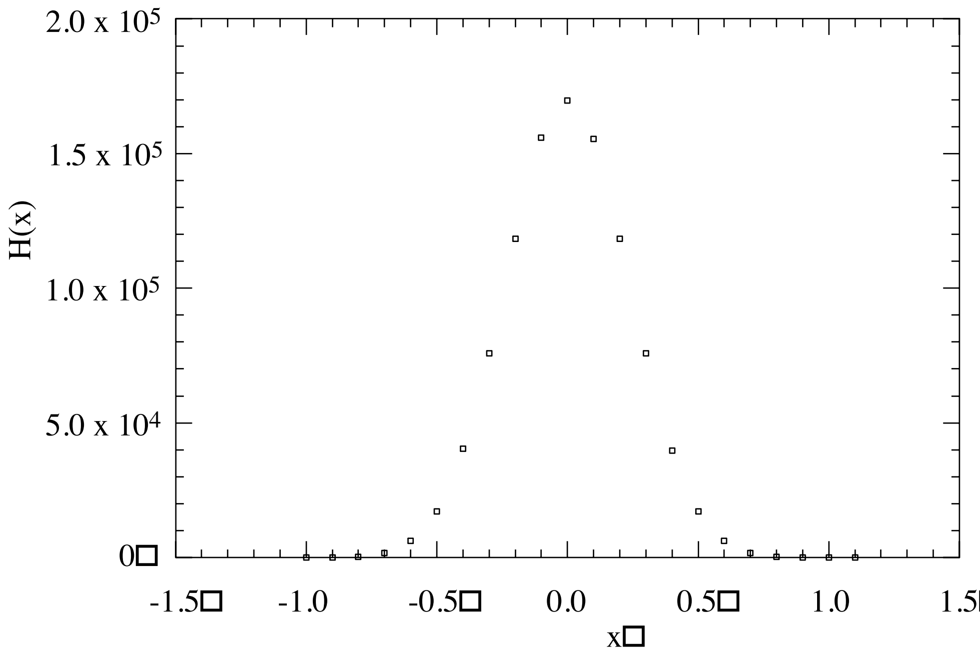

Figure 3.5: Histogram of the number of times that the displacement of a one-dimensional random walker is between x and $x + \Delta x$ after $N = 16$ steps. The data was generated by simulating 10^6 walkers. The length of each step was chosen at random between zero and unity and the bin width is $\Delta x = 0.1$.

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

$$\overline{f(x)} = \int_a^b f(x)\, p(x)\, dx. \tag{3.75}$$

Problem 3.39. The random variable x has the probability density

$$p(x) = \begin{cases} A\, e^{-\lambda x} & \text{if } 0 \leq x \leq \infty \\ 0 & x < 0. \end{cases} \tag{3.76}$$

(a) Determine the normalization constant A in terms of λ. (b) What is the mean value of x? What is the most probable value of x? (c) Choose $\lambda = 1.0$ and determine the probability that a measurement of x yields a value less than 0.3.

Problem 3.40. Consider the probability density function, $p(\mathbf{v}) = (a/\pi)^{3/2}\, e^{-av^2}$, for the velocity $\mathbf{v}$ of a particle. Each of the three velocity components can range from $-\infty$ to $+\infty$ and a is a constant. (a) Show that $p(\mathbf{v})$ is normalized to unity. Use the fact that

$$\int_0^\infty e^{-au^2}\,du = \frac{1}{2}\sqrt{\frac{\pi}{a}}. \tag{3.77}$$

Note that this calculation involves doing three similar integrals that can be evaluated separately. (b) What is the probability that $v_x \geq 0$, $v_y \geq 0$, $v_z \geq 0$ simultaneously?

Problem 3.41. Find the first four moments of the probability density given by

$$p(x) = \begin{cases} (2a)^{-1} & \text{for } |x| \leq a \\ 0 & \text{for } |x| > a \end{cases} \tag{3.78}$$

Plot $p(x)$ and describe its form.

Problem 3.42. Find the first four moments of the Gaussian probability density

$$p(x) = (2\pi)^{-\frac{1}{2}} e^{-x^2/2}. \quad (-\infty < x < \infty) \tag{3.79}$$

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

$$C_4 = \overline{x^4} - 4\overline{x^3}\,\overline{x} - 3\,\overline{x^2}^2 + 12\,\overline{x^2}\,\overline{x}^2 - 6\,\overline{x}^4. \tag{3.80}$$

The quantity C_4 is called the fourth-order cumulant (see (3.131)). What is C_4 for the probability density in (3.79)?

Problem 3.43. Not all probability densities have a finite variance. Sketch the *Lorentz* or *Cauchy distribution* given by

$$p(x) = \frac{1}{\pi}\frac{\gamma}{(x-a)^2 + \gamma^2}. \qquad (-\infty < x < \infty) \tag{3.81}$$

Choose $a = 0$ and $\gamma = 1$ and compare the form of $p(x)$ in (3.81) to the Gaussian distribution given by (3.79). Give a simple argument for the existence of the first moment of the Lorentz distribution. Does the second moment exist?

3.7 The Gaussian distribution as a limit of the binomial distribution

In Problem 3.33 we found that for large N, the binomial distribution has a well-defined maximum at $n = pN$ and can be approximated by a smooth, continuous function even though only integer values of n are physically possible. We now find the form of this function of n.

The first step is to realize that for $N >> 1$, $P_N(n)$ is a rapidly varying function of n near $n = pN$, and for this reason we do not want to approximate $P_N(n)$ directly. However, because the logarithm of $P_N(n)$ is a slowly varying function (see Problem 3.33), we expect that the power series expansion of $\ln P_N(n)$ to converge. Hence, we expand $\ln P_N(n)$ in a Taylor series about the value of $n = \tilde{n}$ at which $\ln P_N(n)$ reaches its maximum value. We will write $p(n)$ instead of $P_N(n)$ because we will treat n as a continuous variable and hence $p(n)$ is a probability density. We find

$$\ln p(n) = \ln p(n = \tilde{n}) + (n - \tilde{n}) \frac{d \ln p(n)}{dn}\Big|_{n=\tilde{n}} + \frac{1}{2}(n - \tilde{n})^2 \frac{d^2 \ln p(n)}{d^2 n}\Big|_{n=\tilde{n}} + \cdots \tag{3.82}$$

Because we have assumed that the expansion (3.82) is about the maximum $n = \tilde{n}$, the first derivative $d \ln p(n)/dn\big|_{n=\tilde{n}}$ must be zero. For the same reason the second derivative $d^2 \ln p(n)/dn^2\big|_{n=\tilde{n}}$ must be negative. We assume that the higher terms in (3.82) can be neglected and adopt the notation

$$\ln A = \ln p(n = \tilde{n}), \tag{3.83}$$

and

$$B = -\frac{d^2 \ln p(n)}{dn^2}\Big|_{n=\tilde{n}}. \tag{3.84}$$

The above approximations and notation allows us to write

$$\ln p(n) \approx \ln A - \frac{1}{2} B (n - \tilde{n})^2, \tag{3.85}$$

or

$$p(n) \approx A\, e^{-\frac{1}{2} B (n - \tilde{n})^2}. \tag{3.86}$$

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

$$\ln p(n) = \ln N! - \ln n! - \ln(N - n)! + n \ln p + (N - n) \ln q. \tag{3.87}$$

It is straightforward to use the relation (3.72) to obtain

$$\frac{d \ln p}{dn} = -\ln n + \ln(N - n) + \ln p - \ln q. \tag{3.88}$$

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

$$\frac{N - \tilde{n}}{\tilde{n}} = \frac{q}{p}, \tag{3.89}$$

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

$$\tilde{n} = pN. \tag{3.90}$$

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

The second derivative can be found from (3.88). We have

$$\frac{d^2 \ln p}{dn^2} = -\frac{1}{n} - \frac{1}{N - n}. \tag{3.91}$$

Hence, the coefficient B defined in (3.84) is given by

$$B = -\frac{d^2 \ln p}{dn^2} = \frac{1}{\tilde{n}} + \frac{1}{N - \tilde{n}} = \frac{1}{Npq}. \tag{3.92}$$

From the relation (3.67) we see that

$$B = \frac{1}{\sigma^2}, \tag{3.93}$$

where σ^2 is the variance of n.

If we use the simple form of Stirling's approximation (3.69) to find the normalization constant A from the relation $\ln A = \ln p(n = \tilde{n})$, we would find that $\ln A = 0$. Instead, we have to use the more accurate form of Stirling's approximation (3.70). The result is

$$A = \frac{1}{(2\pi Npq)^{1/2}} = \frac{1}{(2\pi\sigma^2)^{1/2}}. \tag{3.94}$$

If we substitute our results for $\tilde{n}$, B, and A into (3.86), we find the standard form for the Gaussian distribution

$$p(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(n-\overline{n})^2/2\sigma^2}. \qquad \text{(Gaussian probability density)} \tag{3.95}$$

An alternative derivation of the parameters A, B, and $\tilde{n}$ is given in Problem 3.70.

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 distribution (3.95) for $p = q = 1/2$.

From our derivation we see that (3.95) is valid for large values of N and for values of n near $\overline{n}$. Even for relatively small values of N, the Gaussian approximation is a good approximation for most values of n. A comparison of the Gaussian approximation to the binomial distribution is given in Table 3.4.

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

3.8 The central limit theorem or why is thermodynamics possible?

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

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

$$S = \sum_{i=1}^{N} s_i, \tag{3.96}$$

where

$$s_i = \begin{cases} 1, & \text{if the } i\text{th throw gives 1} \\ 0 & \text{otherwise.} \end{cases} \tag{3.97}$$

If N is large, then S/N approaches 1/6. How does this ratio approach the limit? We can empirically answer this question by repeating the measurement M times. (Each measurement of S consists of N throws of a die.) Because S itself is a random variable, we know that the measured values of S will not be identical. In Figure 3.6 we show the results of $M = 10,000$ measurements of S for $N = 100$ and $N = 800$. We see that the form of the distribution of values of S is approximately Gaussian. In Problem 3.44 we calculate the absolute and relative width of the distributions.

Problem 3.44. Estimate the absolute width and the relative width of the distributions shown in Fig. 3.6 for $N = 100$ and $N = 800$. Does the error of any one measurement of S decreases with increasing N as expected? How would the plot change if M were increased to $M = 10,000$?

In Section 3.12 we show that in the limit of large N, the probability density $p(S)$ is given by

$$p(S) = \frac{1}{\sqrt{2\pi\sigma_S^2}} e^{-(S-\overline{S})^2/2\sigma_S^2}, \tag{3.98}$$

where

$$\overline{S} = N\overline{s} \tag{3.99}$$

$$\sigma_S^2 = N\sigma^2, \tag{3.100}$$

with $\sigma^2 = \overline{s^2} - \overline{s}^2$. The quantity $p(S)\Delta S$ is the probability that the value of $\sum_{i=1}^{N} s_i$ is between S and $S + \Delta S$. Equation (3.98) is equivalent to the central limit theorem. Note that the Gaussian form in (3.98) 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 inclusion of 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 sum of a large number of random variables

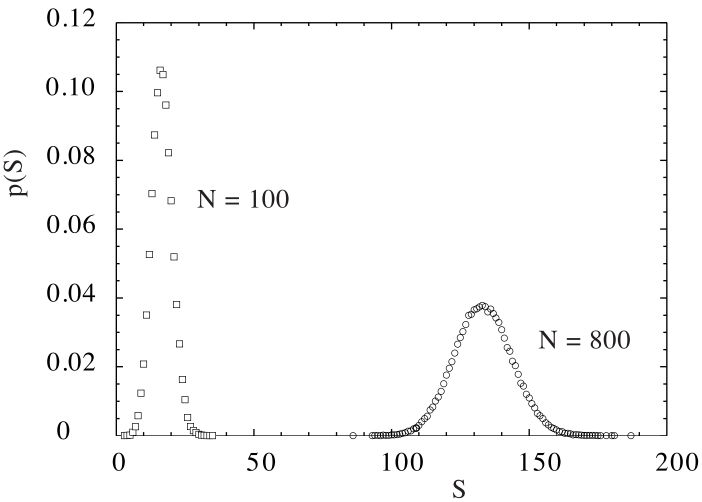

Figure 3.6: The distribution of the measured values of $M = 10,000$ different measurements of the sum S for $N = 100$ and $N = 800$ terms in the sum. The quantity S is the number of times that face 1 appears in N throws of a die. For $N = 100$, the measured values are $\overline{S} = 16.67$, $\overline{S^2} = 291.96$, and $\sigma_S = 3.74$. For $N = 800$, the measured values are $\overline{S} = 133.31$, $\overline{S^2} = 17881.2$, and $\sigma_S = 10.52$. What are the estimated values of the relative width for each case?

will approximate a Gaussian distribution. Moreover, the approximation steadily improves as the number of variables in the sum increases.

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

Note that if we let S represent the displacement of a walker after N steps, and let σ^2 equal the mean square displacement for a single step, then the result (3.98)–(3.100) is equivalent to our results for random walks in the limit of large N. Or we can let S represent the magnetization of a system of noninteracting spins and obtain similar results. That is, a random walk and its equivalents are examples of an *additive* random process.

The central limit theorem shows why the Gaussian distribution is ubiquitous in nature. If a random process is related to a sum of a large number of microscopic processes, the sum will be distributed according to the Gaussian distribution independently of the nature of the distribution of the microscopic processes. The central limit theorem implies that macroscopic bodies have well defined macroscopic properties even though their constituent parts are changing rapidly. For example in a gas or liquid, the particle positions and velocities are continuously changing at a rate much faster than a typical measurement time. For this reason we expect that during a measurement of the pressure of a gas or a liquid, there are many collisions with the wall and hence the pressure has a well defined average. We also expect that the probability that the measured pressure deviates from its average value is proportional to $1/\sqrt{N}$, 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 3.45. Use the central limit theorem to show that the probability that a one-dimensional random walker has a displacement between x and $x + dx$. (There is no need to derive the central limit theorem.)

Problem 3.46. Write a program to test the applicability of the central limit theorem. For simplicity, assume that the variable s_i is uniformly distributed between 0 and 1. First compute the mean and standard deviation of s and compare your numerical results with your analytical calculation. Then sum $N = 10\,000$ values of s_i to obtain one measurement of S. Compute the sum for many measurements, say $M = 1000$. Store in an array $H(S)$ the number of times S is between S and $S + \Delta S$. Plot your results for $H(S)$ and determine how $H(S)$ depends on N. How do you results change if $M = 10000$? Do your results for the form of $H(S)$ depend strongly on the number of measurements M?

3.9 The Poisson distribution and should you fly in airplanes?

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

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

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

$$P(n) = \frac{N!}{n!\,(N-n)!}\,p^n(1-p)^{N-n}. \tag{3.101}$$

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

$$\begin{aligned}
\ln\frac{N!}{(N-n)!} &= \ln N! - \ln(N-n)! \qquad (3.102)\\
&\approx N\ln N - (N-n)\ln(N-n)\\
&\approx N\ln N - (N-n)\ln N\\
&= N\ln N - N\ln N + n\ln N\\
&= n\ln N. \qquad (3.103)
\end{aligned}$$

From (3.103) we obtain

$$\frac{N!}{(N-n)!} \approx e^{n\ln N} = N^n. \tag{3.104}$$

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

$$P(n) \approx \frac{N^n}{n!}\, p^n e^{-pN} = \frac{(Np)^n}{n!} e^{-pN}, \tag{3.105}$$

or

$$P(n) = \frac{\overline{n}^n}{n!} e^{-\overline{n}}, \qquad \text{(Poisson distribution)} \tag{3.106}$$

where

$$\overline{n} = pN. \tag{3.107}$$

The form (3.106) is the Poisson distribution.

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

Problem 3.47. Show that the Poisson distribution is properly normalized, and calculate the mean and variance of n. Because $P(n)$ for $n > N$ is negligibly small, you can sum $P(n)$ from $n = 0$ to $n = \infty$ even though the maximum value of n is N. Plot the Poisson distribution $P(n)$ as a function of n for $p = 0.01$ and $N = 100$.

3.10 *Traffic flow and the rxponential distribution

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

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

$$\int_0^t w(t)\Delta t, \tag{3.108}$$

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

$$1 - \int_0^t w(t)\Delta t. \tag{3.109}$$

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

$$\begin{aligned} w(t)\Delta t &= \text{probability that no event occurs in the interval } [0, t] \\ &\quad \times \text{probability that an event occurs in interval } [t, t + \Delta t] \\ &= \big[1 - \int_0^t w(t)dt\big] \times \lambda\Delta t. \end{aligned} \tag{3.110}$$

If we cancel Δt from each side of (3.110) and differentiate both sides with respect to t, we find

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

so that

$$w(t) = Ae^{-\lambda t}. \tag{3.111}$$

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

$$\int_0^\infty w(t)dt = 1 = A\int_0^\infty e^{-\lambda t}dt = A/\lambda. \tag{3.112}$$

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

$$w(t) = \lambda e^{-\lambda t}. \tag{3.113}$$

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

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

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

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \tag{3.114}$$

We first consider the case $n = 0$. If $n = 0$, the probability that no event occurs in the interval t is (see (3.110))

$$P_{n=0}(t) = 1 - \lambda \int_0^t e^{-\lambda t'} dt' = e^{-\lambda t}. \tag{3.115}$$

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

$$P_{n=1}(t) = \int_0^t \lambda e^{-\lambda t'} e^{-\lambda (t'-t)} dt', \tag{3.116}$$

where we have used (3.115) with $t \to (t' - t)$. Hence,

$$P_{n=1}(t) = \int_0^t \lambda e^{-\lambda t} dt = (\lambda t) e^{-\lambda t}. \tag{3.117}$$

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

$$P_n(t) = \int_0^t \lambda e^{-\lambda t'} P_{n-1}(t - t'). \tag{3.118}$$

Equation (3.118) is a recurrence formula that can be used to derive (3.114) by induction. It is easy to see that (3.114) satisfies (3.118) for $n = 0$ and 1. As is usual when solving recursion formula by induction, we assume that (3.114) is correct for $(n - 1)$. We substitute this result into (3.118) and find

$$P_n(t) = \lambda^n e^{-\lambda t} \int_0^t (t - t')^{n-1} dt'/(n - 1)! = \frac{(\lambda t)^b}{n!} e^{-\lambda t}. \tag{3.119}$$

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

Problem 3.48. 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. Is the distribution of the number of vehicles consistent with the Poisson distribution? If so, what is the value of the parameter λ?

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.

See Section xx for a more detailed discussion of traffic flow and various ways of modeling it using methods of statistical physics.

3.11 *Are all probability distributions Gaussian?

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

We now discuss *random multiplicative processes*. Examples of such processes include the distributions of incomes, rainfall, and fragment sizes in rock crushing processes. 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 size is 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 fragments after N blows of the hammer?

To answer this question, consider a binary sequence in which the numbers x_1 and x_2 appear independently with probabilities p and q respectively. If there are N elements in the product Π, we can ask what is $\overline{\Pi}$, the mean value of Π. To compute $\overline{\Pi}$, we define $P(n)$ as 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:

$$P(n) = \frac{N!}{n!\,(N-n)!} p^n q^{N-n}. \tag{3.120}$$

The mean value of the product is given by

$$\overline{\Pi} = \sum_{n=0}^{N} P(n) {x_1}^n {x_2}^{N-n} \tag{3.121}$$

$$= (px_1 + qx_2)^N. \tag{3.122}$$

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

$$\widetilde{\Pi} = ({x_1}^p {x_2}^q)^N. \tag{3.123}$$

We have found that the average value of the sum of random variables is a good approximation to the most probable value of the sum. Let us see if there is a similar relation for a random multiplicative process. We first consider $x_1 = 2$, $x_2 = 1/2$, and $p = q = 1/2$. Then $\overline{\Pi} = [(1/2) \times 2 + (1/2) \times (1/2)]^N = (5/4)^N = e^{N \ln 5/4}$. In contrast $\widetilde{\Pi} = 2^{1/2} \times (1/2)^{1/2} = 1$.

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

*__Problem 3.49.__ (a) Confirm the above general results for $N = 4$ by showing explicitly all the possible values of the product. (b) Consider the case $x_1 = 2$, $x_2 = 1/2$, $p = 1/3$, and $q = 2/3$, and calculate $\overline{\Pi}$ and $\widetilde{\Pi}$.

*__Problem 3.50.__ (a) Show that $\overline{\Pi^m}$ reduces to $(px_1^m)^N$ as $m \to \infty$. This result implies that for $m >> 1$, the mth moment is determined solely by the most extreme event. (b) Based on the Gaussian approximation for the probability of a random additive process, what is a reasonable guess for the continuum approximation to the probability of a random multiplicative process? Such a distribution is called the *log-normal* distribution. Discuss why or why not you expect the log-normal distribution to be a good approximation for $N >> 1$. (c) More insight can be gained by running the applet at <stp.clarku.edu/simulations/product> which simulates the distribution of values of the product $x_1{}^n x_2{}^{N-n}$. Choose $x_1 = 2$, $x_2 = 1/2$, and $p = q = 1/2$ for which we have already calculated the analytical results for $\overline{\Pi}$ and $\widetilde{\Pi}$. First choose $N = 4$ and estimate $\overline{\Pi}$ and $\widetilde{\Pi}$. Do your estimated values converge more or less uniformly to the exact 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. An excellent discussion is given by Redner (see references).

Vocabulary

sample space, events

probability distribution, probability density

ensemble average, time average

mean value, moments, variance, standard deviation

conditional probability, Bayes' theorem

binomial distribution, Gaussian distribution, Poisson distribution

random walk, random additive processes, central limit theorem

Stirling's approximation

Monte Carlo sampling

Notation

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

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

3.12 Appendix 3A: Derivation of the central limit theorem

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

$$\phi(k) = \overline{e^{ikx}} = \int_{-\infty}^{\infty} dx\, e^{ikx} p(x). \tag{3.124}$$

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

$$p(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-ikx}\phi(k). \tag{3.125}$$

Problem 3.51. Calculate the characteristic function of the Gaussian probability density.

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

$$\phi(k) = \sum_{n=0}^{\infty} \frac{k^n}{n!}\frac{d^n\phi(k)}{dk^n}\Big|_{k=0}, \tag{3.126}$$

$$= \overline{e^{ikx}} = \sum_{n=0}^{\infty} \frac{(ik)^n\overline{x^n}}{n!}. \tag{3.127}$$

By comparing coefficients of k^n in (3.126) and (3.127), we see that

$$\overline{x} = -i\frac{d\phi}{dk}\Big|_{k=0}. \tag{3.128}$$

In Problem 3.52 we show that

$$\overline{x^2} - \overline{x}^2 = -\frac{d^2}{dk^2}\ln\phi(k)\Big|_{k=0} \tag{3.129}$$

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

Problem 3.52. The characteristic function generates the *cumulants* C_m defined by

$$\ln\phi(k) = \sum_{m=1}^{\infty}\frac{(ik)^m}{m!}C_m. \tag{3.130}$$

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

$$C_1 = \overline{x} \tag{3.131a}$$

$$C_2 = \sigma^2 = \overline{x^2} - \overline{x}^2 \tag{3.131b}$$

$$C_3 = \overline{x^3} - 3\,\overline{x^2}\,\overline{x} + 2\,\overline{x}^3 \tag{3.131c}$$

$$C_4 = \overline{x^4} - 4\,\overline{x^3}\,\overline{x} - 3\,\overline{x^2}^2 + 12\,\overline{x^2}\,\overline{x}^2 - 6\,\overline{x}^4. \tag{3.131d}$$

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

$$z = x + y. \tag{3.132}$$

How do the probability densities combine? The answer is

$$p(z) = \int dx\,dy\,p_1(x)p_2(y)\,\delta(z - x - y). \tag{3.133}$$

The integral in (3.133) 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 (3.133) 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

$$\begin{aligned}
\phi_z(k) &= \int dz\,e^{ikz}p(z) && (3.134)\\
&= \int dz \int dx \int dy\,e^{ikz}p_1(x)p_2(y)\,\delta(z - x - y)\\
&= \int dx\,e^{ikx}p_1(x) \int dy\,e^{iky}p_2(y)\\
&= \phi_1(k)\phi_2(k). && (3.135)
\end{aligned}$$

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

$$z = x_1 + x_2 + \ldots + x_N. \tag{3.136}$$

Then

$$\phi_z(k) = \prod_{i=1}^{N} \phi_i(k). \tag{3.137}$$

That is, *for independent variables the characteristic function of the sum is the product of the individual characteristic functions.* If we take the logarithm of both sides of (3.137), we obtain

$$\ln \phi_z(k) = \sum_{i=1}^{N} \ln \phi_i(k). \tag{3.138}$$

Each side of (3.138) can be expanded as a power series and compared order by order in powers of ik. The result is that when random variables are added, their associated cumulants also add. That is, the nth order cumulants satisfy the relation:

$$C_n^z = C_n^1 + C_n^2 + \ldots + C_n^N. \tag{3.139}$$

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

If we denote the mean and standard deviation of the weight of an adult male as $\overline{x}$ and σ respectively, then from (3.131a) and (3.139) we find that the mean weight of N adult males is given by $N\overline{x}$. Similarly from (3.131b) we see that the standard deviation of the weight of N adult males is given by $\sigma_N^2 = N\sigma^2$, or $\sigma_N = \sqrt{N}\sigma$. Hence, we find the now familiar result that the sum of N random variables scales as N while the standard deviation scales as $\sqrt{N}$.

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

$$p(S) = \frac{1}{\sqrt{2\pi N {\sigma_x}^2}}\, e^{-S^2/2N{\sigma_x}^2} \tag{3.140}$$

From (3.131b) 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:

$$\overline{\Big(\frac{x_1 + x_2 + \ldots x_N}{N}\Big)^2} = \frac{N\sigma_x^2}{N^2} = \frac{\sigma_x^2}{N}. \tag{3.141}$$

From (3.141) we see that it is useful to define a scaled sum:

$$z = \frac{1}{\sqrt{N}}(x_1 + x_2 + \ldots + x_N), \tag{3.142}$$

and to write the central limit theorem in the form

$$p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2}. \tag{3.143}$$

To obtain the result (3.143), we write the characteristic function of z as

$$\begin{aligned}\phi_z(k) &= \int dx\, e^{ikz} \int dx_1 \int dx_2 \cdots \int dx_N\, \delta\big(z - \big[\frac{x_1 + x_2 + \ldots + x_N}{N^{1/2}}\big]\big) \\ &\quad \times\, p(x_1)\, p(x_2) \ldots p(x_N) \\ &= \int dx_1 \int dx_2 \ldots \int dx_N\, e^{ik(x_1+x_2+\ldots x_N)/N^{1/2}} p(x_1)\, p(x_2) \ldots p(x_N) \\ &= \phi\big(\frac{k}{N^{1/2}}\big)^N. \end{aligned} \tag{3.144}$$

We next take the logarithm of both sides of (3.144) and expand the right-hand side in powers of k to find

$$\ln \phi_z(k) = \sum_{m=2}^{\infty} \frac{(ik)^m}{m!} N^{1-m/2} C_m. \tag{3.145}$$

The $m = 1$ term does not contribute in (3.145) because we have assumed that $\overline{x} = 0$. More importantly, note that as $N \to \infty$, the higher-order terms are suppressed so that

$$\ln \phi_z(k) \to -\frac{k^2}{2} C_2, \tag{3.146}$$

or

$$\phi_z(k) \to e^{-k^2\sigma^2/2} + \ldots \tag{3.147}$$

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

$$p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2}. \tag{3.148}$$

The leading correction to $\phi(k)$ in (3.148) gives rise to a term of order $N^{-1/2}$, and therefore does not contribute in the limit $N \to \infty$.

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

Additional problems

Problems 3.1, 3.2, 3.3, 3.4, and 3.5, page 83;
Problems 3.6, 3.7, and 3.8, 3.9, 3.10, 3.11, 3.12, 3.13, and 3.14, page 84.
Problems 3.15, 3.16, 3.17, 3.18, and 3.19, page 88.
Problems 3.20 and 3.21, page 89.
Problem 3.22, page 91.
Problem 3.23, page 92.
Problem 3.24, page 93.
Problems 3.25 and 3.26, page 95.
Problems 3.27 and 3.28, page 97.
Problems 3.29, 3.30, and 3.31, page 99.
Problems 3.32 and 3.33, page 101.
Problem 3.34, page 103
Problems 3.35, 3.36, 3.37, and 3.38, page 104
Problem 3.39, page 106
Problems 3.40, 3.41, 3.42, and 3.43, page 107.
Problem 3.44, page 110
Problem 3.45, page 112.
Problem 3.46, page 112.
Problem 3.47, page 113.
Problem 3.48, page 115.
Problems 3.51 and 3.52, page 118.

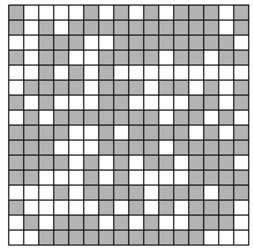

Figure 3.7: Representation of a square lattice of 16×16 sites. The sites are represented by squares and are either occupied (shaded) or empty (white).

Problem 3.53. In Figure 3.7 we show a square lattice of 16^2 sites each of which is either occupied or empty. Estimate the probability that a site in the lattice is occupied.

Problem 3.54. Three coins are tossed in succession. Assume that the simple events are equiprobable. Find the probabilities of the following: (a) the first coin is a heads, (b) exactly two heads have occurred, (c) not more than two heads have occurred.

Problem 3.55. A student tries to solve Problem 3.16 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?

Problem 3.56. A farmer wants to estimate how many fish are in her pond. She takes out 200 fish and tags them and returns them to the pond. After sufficient time to allow the tagged fish to mix with the others, she removes 250 fish at random and finds that 25 of them are tagged. Estimate the number of fish in the pond.

Problem 3.57. A farmer owns a field that is $10\,\mathrm{m} \times 10\,\mathrm{m}$. In the midst of this field is a pond of unknown area. Suppose that the farmer is able to throw 100 stones at random into the field and finds that 40 of the stones make a splash. How can the farmer use this information to estimate the area of the pond?

Problem 3.58. Consider the ten pairs of numbers, (x_i, y_i), given in Table 3.6. The numbers are all are 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

$$F = \int_0^1 dx\, \sqrt{(1 - x^2)}. \tag{3.149}$$

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*.[8] An applet for estimating integrals by Monte Carlo integration can be found at <stp.clarku.edu/simulations/estimation>.

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

	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.

Problem 3.59. A person playing darts hits a bullseye 20% of the time on the average. Why is the probability of b bullseyes in N attempts a binomial distribution? What are the values of p and q? Find the probability that the person hits a bullseye (a) once in five throws; (b) twice in ten throws. Why are these probabilities not identical?

Problem 3.60. 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 3.61. What is the probability that five children produced by the same couple will consist of (a) three sons and two daughters? (b) alternating sexes? (c) alternating sexes starting with a son? (d) all daughters? Assume that the probability of giving birth to a boy and a girl is the same.

Problem 3.62. A good hitter in baseball has a batting average of 300 or more, which means that the hitter will be successful 3 times out of 10 tries on the average. Assume that on average a hitter gets three hits for each 10 times at bat and that he has 4 times at bat per game. (a) What is the probability that he gets zero hits in one game? (b) What is the probability that he will get two hits or less in a three game series? (c) What is the probability that he will get five or more hits in a three game series? Baseball fans might want to think about the significance of "slumps" and "streaks" in baseball.

Problem 3.63. In the World Series in baseball and in the playoffs in the National Basketball Association and the National Hockey Association, the winner is determined by a best of seven series. That is, the first team that wins four games wins the series and is the champion. Do a simple statistical calculation assuming that the two teams are evenly matched and the winner of any game might as well be determined by a coin flip and show that a seven game series should occur 31.25% of the time. What is the probability that the series lasts n games? More information can be found at www.mste.uiuc.edu/hill/ev/seriesprob.html and at www.insidescience.org/reports/2003/080.html.

Problem 3.64. The Galton board (named after Francis Galton (1822–1911)), is a triangular array of pegs. The rows are numbered $0, 1, \ldots$ from the top row down such that row n has $n+1$ pegs. Suppose that a ball is dropped from above the top peg. Each time the ball hits a peg, it bounces to the right with probability p and to the left with probability $1-p$, independently from peg to peg. Suppose that N balls are dropped successively such that the balls do not encounter one another. How will the balls be distributed at the bottom of the board? Links to Java applets that simulate the Galton board can be found in the references.

Problem 3.65. (a) What are the chances that at least two people in your class have the same birthday? Assume that the number of students is 25. (b) What are the chances that at least one other person in your class has the same birthday as you? Explain why the chances are less in the second case.

Problem 3.66. Many analysts attempt to select stocks by looking for correlations in the stock market as a whole or for patterns for particular companies. Such an analysis is based on the belief that there are repetitive patterns in stock prices. To understand one reason for the persistence of this belief do the following experiment. Construct a stock chart (a plot of stock price versus time) showing the movements of a hypothetical stock initially selling at \$50 per share. On each successive day the closing stock price is determined by the flip of a coin. If the coin toss is a head, the stock closes 1/2 point (\$0.50) higher than the preceding close. If the toss is a tail, the price is down by 1/2 point. Construct the stock chart for a long enough time to see "cycles" and other patterns appear. The moral of the charts is that a sequence of numbers produced in this manner is identical to a random walk, yet the sequence frequently appears to be correlated.

Problem 3.67. Suppose that a random walker takes N steps of unit length with probability p of a step to the right. The displacement m of the walker from the origin is given by $m = n - n'$, where n is the number of steps to the right and n' is the number of steps to the left. Show that $\overline{m} = (p - q)N$ and $\sigma_m^2 = \overline{(m - \overline{m})^2} = 4Npq$.

Problem 3.68. The result (3.40) for $\overline{(\Delta M)^2}$ differs by a factor of four from the result for σ_n^2 in (3.67). Why? Compare (3.40) to the result of Problem 3.37.

Problem 3.69. A random walker is observed to take a total of N steps, n of which are to the right. (a) Suppose that a curious observer finds that on ten successive nights the walker takes $N = 20$ steps and that the values of n are given successively by 14, 13, 11, 12, 11, 12, 16, 16, 14, 8. Compute $\overline{n}$, $\overline{n^2}$, and σ_n. Use this information to estimate p. If your reasoning gives different values for p, which estimate is likely to be the most accurate? (b) Suppose that on another ten successive nights the same walker takes $N = 100$ steps and that the values of n are given by 58, 69, 71, 58, 63, 53, 64, 66, 65, 50. Compute the same quantities as in part (a) and estimate p. How does the ratio of σ_n to $\overline{n}$ compare for the two values of N? Explain your results. (c) Compute $\overline{m}$ and σ_m, where $m = n - n'$ is the net displacement of the walker. This problem inspired an article by Zia and Schmittmann (see the references).

Problem 3.70. In Section 3.7 we evaluated the derivatives of $P(n)$ to determine the parameters A, B, and $\tilde{n}$ in (3.86). Another way to determine these parameters is to assume that the binomial distribution can be approximated by a Gaussian and require that the first several moments of the Gaussian and binomial distribution be equal. We write

$$P(n) = Ae^{-\frac{1}{2}B(n-\tilde{n})^2}, \tag{3.150}$$

and require that

$$\int_0^N P(n)\, dn = 1. \tag{3.151}$$

Because $P(n)$ depends on the difference $n - \tilde{n}$, it is convenient to change the variable of integration in (3.151) to $x = n - \tilde{n}$. We have

$$\int_{-\tilde{n}}^{N-\tilde{n}} P(x)\, dx = 1, \tag{3.152}$$

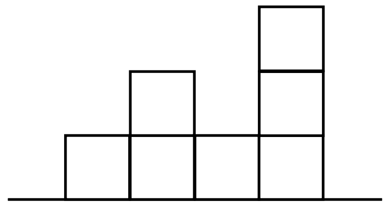

Figure 3.8: Example of a castle wall as explained in Problem 3.71.

where

$$P(x) = Ae^{-\frac{1}{2}Bx^2}. \tag{3.153}$$

In the limit of large N, we can extend the upper and lower limits of integration in (3.152) and write

$$\int_{-\infty}^{\infty} P(x)\, dx = 1, \tag{3.154}$$

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

$$\overline{n} = \int_0^N nP(n)\, dn = pN. \tag{3.155}$$

Make a change of variables and show that

$$\int_{-\infty}^{\infty} xP(x)\, dx = \overline{n} - \tilde{n}. \tag{3.156}$$

Because the integral in (3.156) is zero, we can conclude that $\tilde{n} = \overline{n}$. We also have that

$$\overline{(n-\overline{n})^2} = \int_0^N (n-\overline{n})^2\, P(n)\, dn = pqN. \tag{3.157}$$

Do the integrals in (3.157) and (3.154) (see (A.17) and (A.21)) and confirm that the values of B and A are given by (3.92) and (3.94), respectively. The generality of the arguments leading to the Gaussian distribution suggests that it occurs frequently in probability when large numbers are involved. Note that the Gaussian distribution is characterized completely by its mean value and its width.

Problem 3.71. Consider a two-dimensional "castle wall" constructed from N squares as shown in Figure 3.8. The base row of the cluster must be continuous, but higher rows can have gaps. Each column must be continuous and self-supporting. Determine the total number W_N of different N-site clusters, that is, the number of possible arrangements of N squares consistent with the above rules. Assume that the squares are identical.

Problem 3.72. Suppose that a one-dimensional unbiased random walker starts out at the origin $x = 0$ at $t = 0$. How many steps will it take for the walker to reach a site at $x = 4$? This quantity, known as the *first passage time*, is a random variable because it is different for different possible realizations of the walk. Possible quantities of interest are the probability distribution of the first passage time and the mean first passage time, τ. Write a computer program to estimate $\tau(x)$ and then determine its analytical dependence on x. Why is it more difficult to estimate τ for $x = 8$ than for $x = 4$?

Problem 3.73. 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 3.74.__ To determine the validity of the Gaussian distribution as an approximation to the binomial distribution, consider the next two terms in the power series expansion of $\ln P(n)$:

$$\frac{1}{3!}(n-\tilde{n})^3 C + \frac{1}{4!)}(n-\tilde{n})^4 D, \tag{3.158}$$

with $C = d^3 \ln P(n)/d^3 n$ and $D = d^4 \ln P(n)/d^4 n$ evaluated at $n = \tilde{n}$. Because $C = 0$ if $p = q$, we need to consider the next term in the expansion. Calculate D for $p = q$ and estimate the order of magnitude of the first nonzero correction. Compare this correction to the magnitude of the first nonzero term in $\ln P(n)$ (see (3.82)) and determine the conditions for which the terms beyond $(n-\tilde{n})^2$ can be neglected. That is, show that if N is sufficiently large and neither p nor q is too small, the Gaussian distribution is a good approximation for n near the maximum of $P(n)$. Because $P(n)$ is very small for large $(n-\overline{n})$, any error in the Gaussian approximation for larger n is negligible.

Problem 3.75. Consider a random walk on a two-dimensional square lattice where the walker has an equal probability of taking a step to one of four possible directions, north, south, east, or west. Use the central limit theorem to find the probability that the walker is a distance r to $r + dr$ from the origin, where $r^2 = x^2 + y^2$ and r is the distance from the origin after N steps. There is no need to do an explicit calculation.

Problem 3.76. One of the first continuum models of a random walk is due to Rayleigh (1919). In the Rayleigh model the length a of each step is a random variable with probability density $p(a)$ and the direction of each step is random. For simplicity consider a walk in two dimensions and choose $p(a)$ so that each step has unit length. Then at each step the walker takes a step of unit length at a random angle. Use the central limit theorem to find the asymptotic form of $p(r, N)\,dr$, the probability that the walker is in the range r to $r + dr$, where r is the distance from the origin after N steps.

Problem 3.77. Suppose there are three boxes each with two balls. The first box has two green balls, the second box has one green and one red ball, and the third box has two red balls. Suppose you choose a box at random and find one green ball. What is the probability that the other ball is green?

Problem 3.78. Open a telephone directory to an random page and make a list corresponding to the last digit n of the first 100 telephone numbers. Find the probability $P(n)$ that the number n appears. Plot $P(n)$ as a function of n and describe its n-dependence. Do you expect that $P(n)$ should be approximately uniform?

*__Problem 3.79.__ A simple model of a porous rock can be imagined by placing a series of overlapping spheres at random into a container of fixed volume V. The spheres represent the rock and the space between the spheres represents the pores. If we write the volume of the sphere as v, it can be shown the fraction of the space between the spheres or the *porosity* ϕ is $\phi = \exp(-Nv/V)$, where N is the number of spheres. For simplicity, consider a two-dimensional system, and write a program to place disks of diameter unity into a square box. The disks can overlap. Divide the box into square cells each of which has an edge length equal to the diameter of the disks. Find the probability of having 0, 1, 2, or 3 disks in a cell for $\phi = 0.03$, 0.1, and 0.5.

*__Problem 3.80.__ Do a search of the Web and find a site that lists the populations of various cities in the world (not necessarily the largest ones) or the cities of your state or region. The quantity of interest is the first digit of each population. Alternatively, scan the first page of your local newspaper and record the first digit of each of the numbers you find. (The first digit of a number such as 0.00123 is 1.) What is the probability $P(n)$ that the *first* digit is n, where $n = 1, \ldots, 9$? Do you think that $P(n)$ will be the same for all n?

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

$$P(n) = \log_{10}(1 + \frac{1}{n}). \tag{3.159}$$

The distribution (3.159) is known as *Benford's Law* and is named after Frank Benford, a physicist. It implies that for certain data sets, the first digit is distributed in a predictable pattern with a higher percentage of the numbers beginning with the digit 1. What are the numerical values of $P(n)$ for the different values of n? Is $P(n)$ normalized? Suggest a hypothesis for the nonuniform nature of the Benford distribution.

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

The frequencies of the first digit of 2000 numerical answers to problems given in the back of four physics and mathematics textbooks have been tabulated and found to be distributed in a way consistent with Benford's law. Benford's Law is also expected to hold for answers to homework problems (see James R. Huddle, "A note on Benford's law," *Math. and Comput. Educ.* **31**, 66 (1997)).

Understanding the reasons for Benford's law and its applications would make an excellent project. The nature of Benford's law is discussed by T. P. Hill, "The first digit phenomenon," *Amer. Sci.* **86**, 358–363 (1998).

*__Problem 3.81.__ Ask several of your friends to flip a coin 200 times and record the results or pretend to flip a coin and fake the results. Can you tell which of your friends faked the results? Hint: What is the probability that a sequence of six heads in a row will occur? Can you suggest any other statistical tests?

*__Problem 3.82.__ Analyze a text and do a ranking of the word frequencies. The word with rank r is the rth word when the words of the text are listed with decreasing frequency. Make a log-log plot of word frequency f versus word rank r. The relation between word rank and word frequency

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.

was first stated by George Kingsley Zipf (1902–1950). This relation states that to a very good approximation for a given text

$$f \sim \frac{1}{r \ln(1.78R)}, \tag{3.160}$$

where R is the number of different words. Note the inverse power law behavior. The relation (3.160) 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. [xx any volunteers to write a Perl or Python program to analyze text? xx]

Problem 3.83. Three cards are in a hat. One card is white on both sides, the second is white on one side and red on the other, and the third is red on both sides. The dealer shuffles the cards, takes one out and places it flat on the table. The side showing is red. The dealer now says, "Obviously this card is not the white-white card. It must be either the red-white card or the red-red card. I will bet even money that the other side is red." Is this bet fair?

Problem 3.84. Estimate the probability that an asteroid will impact the earth and cause major damage. Does it make sense for society to take steps now to guard itself against such an occurrence?

***Problem 3.85.** Does it make sense to talk about a "hot hand" in basketball? Does capital punishment deter murder? Are vegetarians more likely to have daughters? Visit the Chance database and read about other interesting issues involving probability and statistics.

***Problem 3.86.** A doctor has two drugs, A and B, which she can prescribe to patients with a certain illness. The drugs have been rated in terms of their effectiveness on a scale of 1 to 6, with 1 being the least effective and 6 being the most effective. Drug A is uniformly effective with a value of 3. The effectiveness of drug B is variable and 54% of the time it scores a value of 1, and 46% of the time it scores a value of 5. The doctor wishes to provide her patients with the best possible care and asks her statistician friend which drug has the highest probability of being the most effective. The statistician says, "It is clear that drug A is the most effective drug 54% of the time. Thus drug A is your best bet."

Later a new drug C becomes available. Studies show that on the scale of 1 to 6, 22% of the time this drug scores a 6, 22% of the time it scores a 4, and 56% of the time it scores a 2.

The doctor, again wishing to provide her patients with the best possible care, goes back to her statistician friend and asks him which drug has the highest probability of being the most effective. The statistician says, "Because there is this new drug C on the market, your best bet is now drug B, and drug A is your worst bet." Show that the statistician is right.

Suggestions for Further Reading

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

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

The Chance database encourages its users to apply statistics to everyday events.

Giulio D'Agostini, "Teaching statistics in the physics curriculum: Unifying and clarifying role of subjective probability," Am. J. Phys. **67**, 1260–1268 (1999). The author, whose main research interest is in particle physics, discusses subjective probability and Bayes' theorem.

See <www.math.uah.edu/psol/applets/GaltonBoardExperiment.html> for an excellent simulation of the Galton board.

Elliott W. Montroll and Michael F. Shlesinger, "On the wonderful world of random walks," in *Studies in Statistical Mechanics*, Vol. XI: Nonequilibrium Phenomena II, J. L. Lebowitz and E. W. Montroll, eds., North-Holland (1984).

Elliott W. Montroll and Wade W. Badger, *Introduction to Quantitative Aspects of Social Phenomena*, Gordon and Breach (1974). The applications of probability that are discussed include traffic flow, income distributions, floods, and the stock market.

Richard Perline, "Zipf's law, the central limit theorem, and the random division of the unit interval," Phys. Rev. E **54**, 220–223 (1996).

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

Charles Ruhla, *The Physics of Chance*, Oxford University Press (1992).

B. Schmittmann and R. K. P. Zia, "'Weather' records: Musings on cold days after a long hot Indian summer," Am. J. Phys. **67**, 1269 (1999). A relatively simple introduction to the statistics of extreme values. Suppose that some breaks the record for the 100 meter dash. How long do records typically survive before they are broken?

Kyle Siegrist at the University of Alabama in Huntsville has developed many Java applets to illustrate concepts in probability and statistics. See <www.math.uah.edu/stat/> and follow the link to Bernoulli processes.

G. Troll and P. beim Graben, "Zipf's law is not a consequence of the central limit theorem," Phys. Rev. E **57**, 1347–1355 (1998).

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

R. K. P. Zia and B. Schmittmann, "Watching a drunkard for 10 nights: A study of distributions of variances," Am. J. Phys. **71**, 859 (2003). See Problem 3.69.