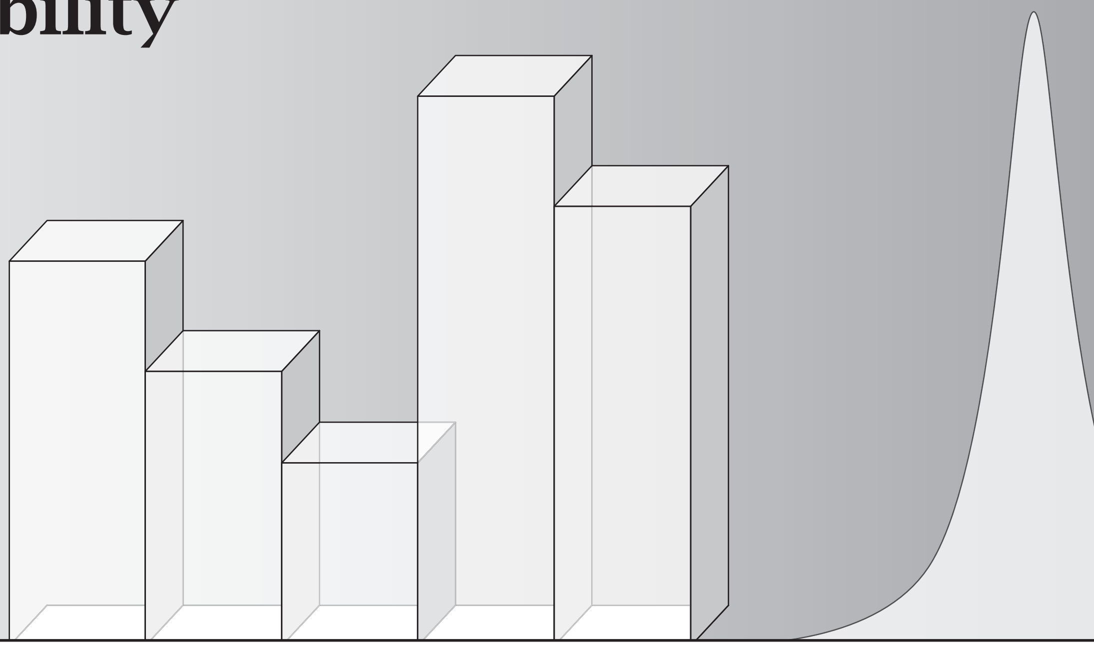
1 Principles of Probability



The Principles of Probability Are the Foundations of Entropy

Fluids flow, boil, freeze, and evaporate. Solids melt and deform. Oil and water don't mix. Metals and semiconductors conduct electricity. Crystals grow. Chemicals react and rearrange, take up heat, and give it off. Rubber stretches and retracts. Proteins catalyze biological reactions. What forces drive these processes? This question is addressed by statistical thermodynamics, a set of tools for modeling molecular forces and behavior, and a language for interpreting experiments.

The challenge in understanding these behaviors is that the properties that can be measured and controlled, such as density, temperature, pressure, heat capacity, molecular radius, or equilibrium constants, do not predict the tendencies and equilibria of systems in a simple and direct way. To predict equilibria, we must step into a different world, where we use the language of *energy*, *entropy*, *enthalpy*, and *free energy*. Measuring the density of liquid water just below its boiling temperature does not hint at the surprise that, just a few degrees higher, above the boiling temperature, the density suddenly drops more than a thousandfold. To predict density changes and other measurable properties, you need to know about the driving forces, the entropies and energies. We begin with entropy.

Entropy is one of the most fundamental concepts in statistical thermodynamics. It describes the tendency of matter toward disorder. Entropy explains how expanding gases drive car engines, liquids mix, rubber bands retract, heat flows from hot objects to cold objects, and protein molecules tangle together in some disease states. The concepts that we introduce in this chapter, *probability*, *multiplicity*, *combinatorics*, *averages*, and *distribution functions*, provide a foundation for describing entropy.

What Is Probability?

Here are two statements of probability. In 1990, the probability that a person in the United States was a scientist or an engineer was 1/250. That is, there were about a million scientists and engineers out of a total of about 250 million people. In 1992, the probability that a child under 13 years old in the United States ate a fast-food hamburger on any given day was 1/30 [1].

Let's generalize. Suppose that the possible outcomes fall into categories A, B, or C. 'Event' and 'outcome' are generic terms. An event might be the flipping of a coin, resulting in heads or tails. Alternatively, it might be one of the possible conformations of a molecule. Suppose that outcome A occurs 20% of the time, B 50% of the time, and C 30% of the time. Then the probability of A is 0.20, the probability of B is 0.50, and the probability of C is 0.30.

The **definition of probability** is as follows: If N is the total number of possible outcomes, and n_A of the outcomes fall into category A, then p_A, the probability of outcome A, is

$$p_A = \left(\frac{n_A}{N}\right). \tag{1.1}$$

Probabilities are quantities in the range from zero to one. If only one outcome is possible, the process is *deterministic*—the outcome has a probability of one. An outcome that never occurs has a probability of zero.

Probabilities can be computed for different combinations of events. Consider one roll of a six-sided die, for example (die, unfortunately, is the singular of dice). The probability that a **4** appears face up is 1/6 because there are $N = 6$ possible outcomes and only $n_4 = 1$ of them is a **4**. But suppose you roll a six-sided die three times. You may ask for the probability that you will observe the sequence of two **3**'s followed by one **4**. Or you may ask for the probability of rolling two **2**'s and one **6** in any order. The rules of probability and combinatorics provide the machinery for calculating such probabilities. Here we define the relationships among events that we need to formulate these rules.

Definitions: Relationships Among Events

Mutually Exclusive. Outcomes $A_1, A_2, \ldots, A_t$ are mutually exclusive if the occurrence of each one of them precludes the occurrence of all the others. If A and B are mutually exclusive, then if A occurs, B does not. If B occurs, A does not. For example, on a single die roll, **1** and **3** are mutually exclusive because only one number can appear face up each time the die is rolled.

Collectively Exhaustive. Outcomes $A_1, A_2, \ldots, A_t$ are collectively exhaustive if they constitute the entire set of possibilities, and no other outcomes are possible. For example, [heads, tails] is a collectively exhaustive set of outcomes for a coin toss, provided that you don't count the occasions when the coin lands on its edge.

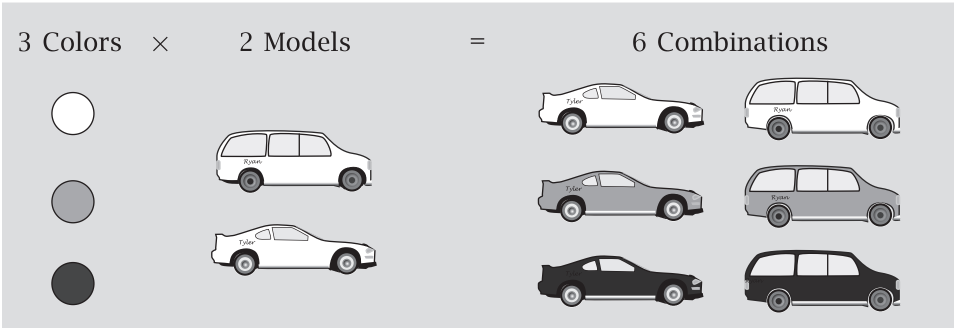


Figure 1.1 If there are three car colors for each of two car models, there are six different combinations of color and model, so the multiplicity is six.

Independent. Events $A_1, A_2, \ldots, A_t$ are independent if the outcome of each one is unrelated to (or not *correlated* with) the outcome of any other. The score on one die roll is independent of the score on the next, unless there is trickery.

Multiplicity. The multiplicity of events is the total number of ways in which different outcomes can possibly occur. If the number of outcomes of type A is n_A, the number of outcomes of type B is n_B, and the number of outcomes of type C is n_C, the total number of possible combinations of outcomes is the multiplicity W:

$$W = n_A n_B n_C. \tag{1.2}$$

Figure 1.1 shows an example of multiplicity.

The Rules of Probability Are Recipes for Drawing Consistent Inferences

The addition and multiplication rules permit you to calculate the probabilities of certain combinations of events.

Addition Rule. If outcomes $A, B, \ldots, E$ are mutually exclusive, and occur with probabilities $p_A = n_A/N$, $p_B = n_B/N, \ldots, p_E = n_E/N$, then the probability of observing either A OR B OR ... OR E (the union of outcomes, expressed as $A \cup B \cup \cdots \cup E$) is the sum of the probabilities:

$$\begin{aligned} p(A \text{ OR } B \text{ OR } \ldots \text{ OR } E) &= \frac{n_A + n_B + \cdots + n_E}{N} \\ &= p_A + p_B + \cdots + p_E. \end{aligned} \tag{1.3}$$

The addition rule holds only if two criteria are met: the outcomes are *mutually exclusive* and we seek the probability of one outcome OR another outcome.

When they are not divided by N, the broader term for the quantities n_i $(i = A, B, \ldots, E)$ is *statistical weights*. If outcomes $A, B, \ldots, E$ are both collectively exhaustive and mutually exclusive, then

$$n_A + n_B + \cdots + n_E = N, \tag{1.4}$$

and dividing both sides of Equation (1.4) by N, the total number of trials, gives

$$p_A + p_B + \cdots + p_E = 1. \tag{1.5}$$

Multiplication Rule. If outcomes $A, B, \ldots, E$ are independent, then the probability of observing A AND B AND... AND E (the intersection of outcomes, expressed as $A \cap B \cap \cdots \cap E$) is the product of the probabilities,

$$\begin{aligned} p(A \text{ AND } B \text{ AND } \ldots \text{ AND } E) &= \left(\frac{n_A}{N}\right)\left(\frac{n_B}{N}\right)\cdots\left(\frac{n_E}{N}\right) \\ &= p_A p_B \cdots p_E. \end{aligned} \tag{1.6}$$

The multiplication rule applies when the outcomes are *independent* and we seek the probability of one outcome AND another outcome AND possibly other outcomes. A more general multiplication rule, described on page 8, applies even when outcomes are not independent.

Here are a few examples using the addition and multiplication rules.

EXAMPLE 1.1 Rolling a die. What is the probability that either a **1** or a **4** appears on a single roll of a die? The probability of a **1** is $1/6$. The probability of a **4** is also $1/6$. The probability of either a **1** OR a **4** is $1/6+1/6 = 1/3$, because the outcomes are mutually exclusive (**1** and **4** can't occur on the same roll) and the question is of the OR type.

EXAMPLE 1.2 Rolling twice. What is the probability of a **1** on the first roll of a die and a **4** on the second? It is $(1/6)(1/6) = 1/36$, because this is an AND question, and the two events are independent. This probability can also be computed in terms of the multiplicity. There are six possible outcomes on each of the two rolls of the die, giving a product of $W = 36$ possible combinations, one of which is **1** on the first roll and **4** on the second.

EXAMPLE 1.3 A sequence of coin flips. What is the probability of getting five heads on five successive flips of an unbiased coin? It is $(1/2)^5 = 1/32$, because the coin flips are independent of each other, this is an AND question, and the probability of heads on each flip is $1/2$. In terms of the multiplicity of outcomes, there are two possible outcomes on each flip, giving a product of $W = 32$ total outcomes, and only one of them is five successive heads.

EXAMPLE 1.4 Another sequence of coin flips. What is the probability of two heads, then one tail, then two more heads on five successive coin flips? It is $p_H^2 p_T p_H^2 = (1/2)^5 = 1/32$. You get the same result as in Example 1.3 because p_H, the probability of heads, and p_T, the probability of tails, are both $1/2$. There are a total of $W = 32$ possible outcomes and only one is the given sequence. The probability $p(n_H, N)$ of observing one particular sequence of N coin flips having exactly n_H heads is

$$p(n_H, N) = p_H^{n_H} p_T^{N-n_H}. \tag{1.7}$$

If $p_H = p_T = 1/2$, then $p(n_H, N) = (1/2)^N$.

EXAMPLE 1.5 Combining events—both, either/or, or neither. If independent events A and B have probabilities p_A and p_B, the probability that *both* events happen is $p_A p_B$. What is the probability that A happens AND B does not? The probability that B does not happen is $(1-p_B)$. If A and B are independent events, then the probability that A happens and B does not is $p_A(1-p_B) = p_A - p_A p_B$. What is the probability that *neither* event happens? It is

$$p(\text{not } A \text{ AND not } B) = (1-p_A)(1-p_B), \tag{1.8}$$

where $p(\text{not } A \text{ AND not } B)$ is the probability that A does not happen AND B does not happen.

EXAMPLE 1.6 Combining events—something happens. What is the probability that *something* happens, that is, A OR B OR both happen? This is an OR question, but the events are independent and not mutually exclusive, so you cannot use either the addition or multiplication rules. You can use a simple trick instead. The trick is to consider the probabilities that events *do not* happen, rather than that events *do* happen. The probability that something happens is $1-p(\text{nothing happens})$:

$$1-p(\text{not } A \text{ AND not } B) = 1-(1-p_A)(1-p_B) = p_A + p_B - p_A p_B. \tag{1.9}$$

Multiple events can occur as ordered sequences in *time*, such as die rolls, or as ordered sequences in *space*, such as the strings of characters in words. Sometimes it is more useful to focus on collections of events rather than the individual events themselves.

Elementary and Composite Events

Some problems in probability cannot be solved directly by applying the addition or multiplication rules. Such questions can usually be reformulated in terms of *composite events* to which the rules of probability can be applied. Example 1.7 shows how to do this. Then on page 14 we'll use reformulation to construct probability distribution functions.

EXAMPLE 1.7 Elementary and composite events. What is the probability of a **1** on the first roll of a die OR a **4** on the second roll? If this were an AND question, the probability would be $(1/6)(1/6) = 1/36$, since the two rolls are independent, but the question is of the OR type, so it cannot be answered by direct application of either the addition or multiplication rules. But by redefining the problem in terms of composite events, you can use those rules. An individual coin toss, a single die roll, etc. could be called an elementary event. A composite event is just some set of elementary events, collected together in a convenient way. In this example it's convenient to define each composite event to be a pair of first and second rolls of the die. The advantage is that the complete list of composite events is mutually exclusive. That allows us to frame the problem in terms of an OR question and use the multiplication and addition rules. The composite events are:

[1, 1]*	**[1, 2]***	**[1, 3]***	**[1, 4]***	**[1, 5]***	**[1, 6]***
[2, 1]	**[2, 2]**	**[2, 3]**	**[2, 4]***	**[2, 5]**	**[2, 6]**
[3, 1]	**[3, 2]**	**[3, 3]**	**[3, 4]***	**[3, 5]**	**[3, 6]**
[4, 1]	**[4, 2]**	**[4, 3]**	**[4, 4]***	**[4, 5]**	**[4, 6]**
[5, 1]	**[5, 2]**	**[5, 3]**	**[5, 4]***	**[5, 5]**	**[5, 6]**
[6, 1]	**[6, 2]**	**[6, 3]**	**[6, 4]***	**[6, 5]**	**[6, 6]**

The first and second numbers in the brackets indicate the outcome of the first and second rolls respectively, and * indicates a composite event that satisfies the criterion for 'success' (**1** on the first roll OR **4** on the second roll). There are 36 composite events, of which 11 are successful, so the probability we seek is 11/36.

Since many of the problems of interest in statistical thermodynamics involve huge systems (say, 10^{20} molecules), we need a more systematic way to compute composite probabilities than enumerating them all.

To compute this probability systematically, collect the composite events into three mutually exclusive classes, A, B, and C, about which you can ask an OR question. Class A includes all composite events with a **1** on the first roll AND anything but a **4** on the second. Class B includes all events with anything but a **1** on the first roll AND a **4** on the second. Class C includes the one event in which we get a **1** on the first roll AND a **4** on the second. A, B, and C are mutually exclusive categories. This is an OR question, so add p_A, p_B, and p_C to find the answer:

$$\begin{aligned} p(\mathbf{1}\text{ first OR }\mathbf{4}\text{ second}) &= p_A(\mathbf{1}\text{ first AND not }\mathbf{4}\text{ second}) \\ &\quad + p_B(\text{not }\mathbf{1}\text{ first AND }\mathbf{4}\text{ second}) \\ &\quad + p_C(\mathbf{1}\text{ first AND }\mathbf{4}\text{ second}). \end{aligned} \tag{1.10}$$

The same probability rules that apply to elementary events also apply to composite events. Moreover, p_A, p_B, and p_C are each products of elementary event probabilities because the first and second rolls of the die are independent:

$$\begin{aligned} p_A &= \left(\frac{1}{6}\right)\left(\frac{5}{6}\right), \\ p_B &= \left(\frac{5}{6}\right)\left(\frac{1}{6}\right), \\ p_C &= \left(\frac{1}{6}\right)\left(\frac{1}{6}\right). \end{aligned}$$

Add p_A, p_B, and p_C: $p(\mathbf{1}\text{ first OR }\mathbf{4}\text{ second}) = 5/36+5/36+1/36 = 11/36$. This example shows how elementary events can be grouped together into composite events to take advantage of the addition and multiplication rules. Reformulation is powerful because virtually any question can be framed in terms of combinations of AND and OR operations. With these two rules of probability, you can draw inferences about a wide range of probabilistic events.

EXAMPLE 1.8 A different way to solve it. Often, there are different ways to collect up events for solving probability problems. Let's solve Example 1.7 differently. This time, use $p(\text{success}) = 1 - p(\text{fail})$. Because $p(\text{fail}) = [p(\text{not }\mathbf{1}\text{ first})\ \text{AND}\ p(\text{not }\mathbf{4}\text{ second})] = (5/6)(5/6) = 25/36$, you have $p(\text{success}) = 11/36$.

Two events can have a more complex relationship than we have considered so far. They are not restricted to being either independent or mutually exclusive. More broadly, events can be *correlated*.

Correlated Events Are Described by Conditional Probabilities

Events are correlated if the outcome of one depends on the outcome of the other. For example, if it rains on 36 days a year, the probability of rain is $36/365 \approx 0.1$. But if it rains on 50% of the days when you see dark clouds, then the probability of observing rain (event B) depends upon, or is conditional upon, the appearance of dark clouds (event A). Example 1.9 and Table 1.1 demonstrate the correlation of events when balls are taken out of a barrel.

EXAMPLE 1.9 Balls taken from a barrel with replacement. Suppose a barrel contains one red ball, R, and two green balls, G. The probability of drawing a green ball on the first try is $2/3$, and the probability of drawing a red ball on the first try is $1/3$. What is the probability of drawing a green ball on the second draw? That depends on whether or not you put the first ball back into the barrel before the second draw. If you replace each ball before drawing another, then the probabilities of different draws are uncorrelated with each other. Each draw is an independent event.

However, if you draw a green ball first, and don't put it back in the barrel, then 1 R and 1 G remain after the first draw, and the probability of getting a green ball on the second draw is now $1/2$. The probability of drawing a green ball on the second try is different from the probability of drawing a green ball on the first try. It is *conditional* on the outcome of the first draw.

Table 1.1 All of the probabilities for the three draws without replacement described in Examples 1.9 and 1.10.

1st Draw		2nd Draw		3rd Draw
$p(R_1) = \frac{1}{3}$	$\longrightarrow$	$p(R_2 \mid R_1)p(R_1)$ $0 \cdot \left(\frac{1}{3}\right) = 0$		
		$p(G_2 \mid R_1)p(R_1)$ $1 \cdot \left(\frac{1}{3}\right) = \frac{1}{3}$	$\longrightarrow$	$p(G_3 \mid G_2R_1)p(G_2R_1)$ $1 \cdot \left(\frac{1}{3}\right) = \frac{1}{3}$
$p(G_1) = \frac{2}{3}$	$\longrightarrow$	$p(R_2 \mid G_1)p(G_1)$ $\left(\frac{1}{2}\right) \cdot \left(\frac{2}{3}\right) = \frac{1}{3}$	$\longrightarrow$	$p(G_3 \mid R_2G_1)p(R_2G_1)$ $1 \cdot \left(\frac{1}{3}\right) = \frac{1}{3}$
		$p(G_2 \mid G_1)p(G_1)$ $\left(\frac{1}{2}\right) \cdot \left(\frac{2}{3}\right) = \frac{1}{3}$	$\longrightarrow$	$p(R_3 \mid G_2G_1)p(G_2G_1)$ $1 \cdot \left(\frac{1}{3}\right) = \frac{1}{3}$

Here are some definitions and examples describing the conditional probabilities of correlated events.

Conditional Probability. The conditional probability $p(B \mid A)$ is the probability of event B, *given that* some other event A has occurred. Event A is the *condition* upon which we evaluate the probability of event B. In Example 1.9, event B is getting a green ball on the second draw, event A is getting a green ball on the first draw, and $p(G_2 \mid G_1)$ is the probability of getting a green ball on the second draw, given a green ball on the first draw.

Joint Probability. The joint probability of events A and B is the probability that both events A AND B occur. The joint probability is expressed by the notation $p(A \text{ AND } B)$, or more concisely, $p(AB)$.

General Multiplication Rule (Bayes' Rule). If outcomes A and B occur with probabilities $p(A)$ and $p(B)$, the joint probability of events A AND B is

$$p(AB) = p(B \mid A)p(A) = p(A \mid B)p(B). \tag{1.11}$$

If events A and B happen to be independent, the pre-condition A has no influence on the probability of B. Then $p(B \mid A) = p(B)$, and Equation (1.11) reduces to $p(AB) = p(B)p(A)$, the multiplication rule for independent events. A probability $p(B)$ that is not conditional is called an *a priori* probability. The conditional quantity $p(B \mid A)$ is called an *a posteriori* probability. The general multiplication rule is general because independence is not required. It defines the probability of the *intersection* of events, $p(AB) = p(A \cap B)$.

General Addition Rule. A general rule can also be formulated for the union of events, $p(A \cup B) = p(A) + p(B) - p(A \cap B)$, when we seek the probability of A OR B for events that are not mutually exclusive. When A and B *are* mutually exclusive, $p(A \cap B) = 0$, and the general addition rule reduces to the simpler addition rule on page 3. When A and B are independent, $p(A \cap B) = p(A)p(B)$, and the general addition rule gives the result in Example 1.6.

Degree of Correlation. The degree of correlation g between events A and B can be expressed as the ratio of the conditional probability of B, given A, to the unconditional probability of B alone. This indicates the degree to which A influences B:

$$g = \frac{p(B \mid A)}{p(B)} = \frac{p(AB)}{p(A)p(B)}. \tag{1.12}$$

The second equality in Equation (1.12) follows from the general multiplication rule, Equation (1.11). If $g = 1$, events A and B are independent and not correlated. If $g > 1$, events A and B are positively correlated. If $g < 1$, events A and B are negatively correlated. If $g = 0$ and A occurs, then B will not. If the *a priori* probability of rain is $p(B) = 0.1$, and if the conditional probability of rain, given that there are dark clouds, A, is $p(B \mid A) = 0.5$, then the degree of correlation of rain with dark clouds is $g = 5$. Correlations are important in statistical thermodynamics. For example, attractions and repulsions among molecules in liquids can cause correlations among their positions and orientations.

EXAMPLE 1.10 Balls taken from that barrel again. As before, start with three balls in a barrel, one red and two green. The probability of getting a red ball on the first draw is $p(R_1) = 1/3$, where the notation R_1 refers to a red ball on the first draw. The probability of getting a green ball on the first draw is $p(G_1) = 2/3$. If balls are not replaced after each draw, the joint probability for getting a red ball first and a green ball second is $p(R_1G_2)$:

$$p(R_1G_2) = p(G_2 \mid R_1)p(R_1) = (1)\left(\frac{1}{3}\right) = \frac{1}{3}. \tag{1.13}$$

So, getting a green ball second is correlated with getting a red ball first:

$$g = \frac{p(R_1G_2)}{p(R_1)p(G_2)} = \frac{\frac{1}{3}}{\left(\frac{1}{3}\right)\left(\frac{1}{3}+\frac{1}{3}\right)} = \frac{3}{2}. \tag{1.14}$$

Rule for Adding Joint Probabilities. The following is a useful way to compute a probability $p(B)$ if you know joint or conditional probabilities:

$$p(B) = p(BA)+p(BA') = p(B \mid A)p(A)+p(B \mid A')p(A'), \tag{1.15}$$

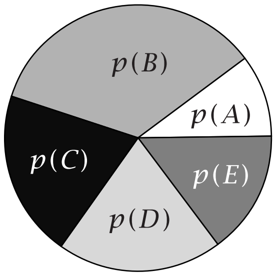
where A' means that event A does not occur. If the event B is rain, and if the event A is that you see clouds and A' is that you see no clouds, then the probability of rain is the sum of joint probabilities of (rain, you see clouds) plus (rain, you see no clouds). By summing over the mutually exclusive conditions, A and A', you are accounting for all the ways that B can happen.

EXAMPLE 1.11 Applying Bayes' rule: Predicting protein properties. *Bayes' rule*, a combination of Equations (1.11) and (1.15), can help you compute hard-to-get probabilities from ones that are easier to get. Here's a toy example. Let's figure out a protein's structure from its amino acid sequence. From modern genomics, it is easy to learn protein sequences. It's harder to learn protein structures. Suppose you discover a new type of protein structure, call it a *helicoil h*. It's rare; you've searched 5000 proteins and found only 20 helicoils, so $p(h) = 0.004$. If you could discover some special amino acid *sequence feature*, call it sf, that predicts the h structure, you could search other genomes to find other helicoil proteins in nature. It's easier to turn this around. Rather than looking through 5000 sequences for patterns, you want to look at the 20 helicoil proteins for patterns. How do you compute $p(\mathrm{sf} \mid h)$? You take the 20 given helicoils and find the fraction of them that have your sequence feature. If your sequence feature (say alternating glycine and lysine amino acids) appears in 19 out of the 20 helicoils, you have $p(\mathrm{sf} \mid h) = 0.95$. You also need $p(\mathrm{sf} \mid \bar{h})$, the fraction of non-helicoil proteins (let's call those $\bar{h}$) that have your sequence feature. Suppose you find $p(\mathrm{sf} \mid \bar{h}) = 0.001$. Combining Equations (1.11) and (1.15) gives Bayes' rule for the probability you want:

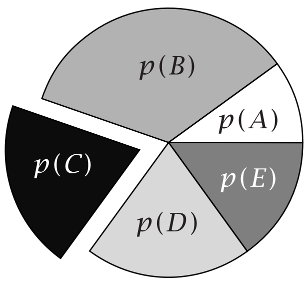
$$\begin{aligned} p(h \mid \mathrm{sf}) &= \frac{p(\mathrm{sf} \mid h)p(h)}{p(\mathrm{sf})} = \frac{p(\mathrm{sf} \mid h)p(h)}{p(\mathrm{sf} \mid h)p(h) + p(\mathrm{sf} \mid \bar{h})p(\bar{h})} \\ &= \frac{(0.95)(0.004)}{(0.95)(0.004) + (0.001)(0.996)} = 0.79. \end{aligned} \tag{1.16}$$

In short, if a protein has the sf sequence, it will have the h structure about 80% of the time.

(a) Who will win?



(b) Given that C won...



(c) Who will place second?

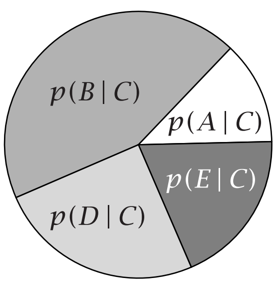


Figure 1.2 (a) *A priori* probabilities of outcomes A to E, such as a horse race. (b) To determine the *a posteriori* probabilities of events A, B, D, and E, given that C has occurred, remove C and keep the relative proportions of the rest the same. (c) *A posteriori* probabilities that horses A, B, D, and E will come in second, given that C won.

Conditional probabilities are useful in a variety of situations, including card games and horse races, as the following example shows.

EXAMPLE 1.12 A gambling equation. Suppose you have a collection of mutually exclusive and collectively exhaustive events $A, B, \ldots, E$, with probabilities $p_A, p_B, \ldots, p_E$. These could be the probabilities that horses $A, B, \ldots, E$ will win a race (based on some theory, model, or prediction scheme), or that card types A to E will appear on a given play in a card game. Let's look at a horse race [2]. If horse A wins, then horses B or C don't win, so these are mutually exclusive.

Suppose you have some information, such as the track records of the horses, that predicts the *a priori* probabilities that each horse will win. Figure 1.2 gives an example. Now, as the race proceeds, the events occur in order, one at a time: one horse wins, then another comes in second, and another comes in third. Our aim is to compute the conditional probability that a particular horse will come in second, given that some other horse has won. The *a priori* probability that horse C will win is $p(C)$. Now *assume* that horse C has won, and you want to know the probability that horse A will be second, $p(A \text{ is second} \mid C \text{ is first})$. From Figure 1.2, you can see that this conditional probability can be determined by eliminating region C, and finding the fraction of the remaining area occupied by region A:

$$p(A \text{ is second} \mid C \text{ is first}) = \frac{p(A)}{p(A)+p(B)+p(D)+p(E)} = \frac{p(A)}{1-p(C)}. \tag{1.17}$$

$1-p(C) = p(A)+p(B)+p(D)+p(E)$ follows from the mutually exclusive addition rule.

The probability that event i is first is $p(i)$. Then the conditional probability that event j is second is $p(j)/[1-p(i)]$. The joint probability that i is first, j is second, and k is third is

$$p(i \text{ is first}, j \text{ is second}, k \text{ is third}) = \frac{p(i)p(j)p(k)}{[1-p(i)][1-p(i)-p(j)]}. \tag{1.18}$$

Equations (1.17) and (1.18) are useful for computing the probability of drawing the queen of hearts in a card game, once you have seen the seven of clubs and the ace of spades. They are also useful for describing the statistical thermodynamics of liquid crystals, and ligand binding to DNA (see pages 575–578).

Combinatorics Describes How to Count Events

Combinatorics, or counting events, is central to statistical thermodynamics. It is the basis for entropy, and the concepts of *order* and *disorder*, which are defined by the numbers of ways in which a system can be configured. Combinatorics is concerned with the *composition* of events rather than the *sequence* of events. For example, compare the following two questions. The first is a question of sequence: What is the probability of the specific sequence of four coin flips, $HTHH$? The second is a question of composition: What is the probability of observing three H's and one T in *any* order? The sequence question is answered by using Equation (1.7): this probability is $1/16$. However, to answer the composition question you must count the number of different possible sequences with the specified composition: $HHHT$, $HHTH$, $HTHH$, and $THHH$. The probability of getting three H's and one T in any order is $4/16 = 1/4$. When you seek the probability of a certain *composition* of events, you count the possible sequences that have the correct composition.

EXAMPLE 1.13 Permutations of ordered sequences. How many permutations, or different sequences, of the letters **w**, **x**, **y**, and **z** are possible? There are 24:

wxyz	**wxzy**	**wyxz**	**wyzx**	**wzxy**	**wzyx**
xwyz	**xwzy**	**xywz**	**xyzw**	**xzwy**	**xzyw**
ywxz	**ywzx**	**yxwz**	**yxzw**	**yzwx**	**yzxw**
zwyx	**zwxy**	**zxwy**	**zxyw**	**zywx**	**zyxw**

How can you compute the number of different sequences without having to list them all? You can use the strategy developed for drawing letters from a barrel without replacement. The first letter of a sequence can be any one of the four. After drawing one, the second letter of the sequence can be any of the remaining three letters. The third letter can be any of the remaining two letters, and the fourth must be the one remaining letter. Use the definition of multiplicity W (Equation (1.2)) to combine the numbers of outcomes n_i, where i represents the position 1, 2, 3, or 4 in the sequence of draws. We have $n_1 = 4$, $n_2 = 3$, $n_3 = 2$, and $n_4 = 1$, so the number of permutations is $W = n_1 n_2 n_3 n_4 = 4 \cdot 3 \cdot 2 \cdot 1 = 24$.

In general, for a sequence of N *distinguishable* objects, the number of different permutations W can be expressed in factorial notation

$$\begin{aligned} W &= N(N-1)(N-2)\cdots 3\cdot 2\cdot 1 = N! \\ &= 4\cdot 3\cdot 2\cdot 1 = 24. \end{aligned} \tag{1.19}$$

The Factorial Notation

The notation $N!$, called N factorial, denotes the product of the integers from one to N:

$$N! = 1\cdot 2\cdot 3\cdots (N-2)(N-1)N.$$

0! is defined to equal 1.

EXAMPLE 1.14 Letters of the alphabet. Consider a barrel containing one each of the 26 letters of the alphabet. What is the probability of drawing the letters out in exactly the order of the alphabet, **A** to **Z**? The probability of drawing the **A** first is $1/26$. If you replace each letter after it is drawn, the probability of drawing the **B** on the second try would be $1/26$, and the probability of drawing the alphabet in order would be $(1/26)^{26}$. But if each letter were *not* replaced in the barrel, the probability of drawing the **B** on the second trial would be $1/25$.

The probability of drawing the **C** on the third trial would be 1/24. Without replacement, the probability of drawing the exact sequence of the alphabet is

$$p(\mathbf{ABC\ldots XYZ}) = \frac{1}{26\cdot 25\cdot 24\cdots 2\cdot 1} = \frac{1}{N!}, \tag{1.20}$$

where $N = 26$ is the number of letters in the alphabet. $N!$ is the number of permutations, or different sequences in which the letters could be drawn. $1/N!$ is the probability of drawing one particular sequence.

In Examples 1.13 and 1.14, all the letters are distinguishable from each other: **w**, **x**, **y**, and **z** are all different. But what happens if some of the objects are indistinguishable from each other?

EXAMPLE 1.15 Counting sequences of distinguishable and indistinguishable objects. How many different arrangements are there of the letters **A**, **H**, and **A**? That depends on whether or not you can tell the **A**'s apart. Suppose first that one **A** has a subscript 1 and the other has a subscript 2: $\mathbf{A}_1$, **H**, and $\mathbf{A}_2$. Then all the characters are distinguishable, as in Examples 1.13 and 1.14, and there are $W = N! = 3! = 6$ different arrangements of these three distinguishable characters:

$$\mathbf{HA}_1\mathbf{A}_2 \qquad \mathbf{A}_1\mathbf{HA}_2 \qquad \mathbf{A}_1\mathbf{A}_2\mathbf{H} \qquad \mathbf{HA}_2\mathbf{A}_1 \qquad \mathbf{A}_2\mathbf{HA}_1 \qquad \mathbf{A}_2\mathbf{A}_1\mathbf{H}.$$

However, now suppose that the two **A**'s are indistinguishable from each other: they have no subscripts. There are now only $W = 3$ distinguishable sequences of letters: **HAA**, **AHA**, and **AAH**. (Distinguishable is a term that applies either to the letters or to the sequences. We have three distinguishable sequences, each containing two distinguishable letters, **A** and **H**.) The previous expression $W = N!$ overcounts by a factor of two when the two **A**'s are indistinguishable. This is because we have counted each sequence of letters, say **AAH**, twice—$\mathbf{A}_1\mathbf{A}_2\mathbf{H}$ and $\mathbf{A}_2\mathbf{A}_1\mathbf{H}$. Written in a more general way, the number of distinguishable sequences is $W = N!/N_A! = 3!/2! = 3$. The $N!$ in the numerator comes from the number of permutations as if all the characters were distinguishable from each other, and the $N_A!$ in the denominator corrects for overcounting. The overcounting correction 2! is simply the count of all the permutations of the indistinguishable characters, the number of ways in which the **A**'s can be arranged among themselves.

EXAMPLE 1.16 Permutations of mixed sequences. Consider the word **cheese** as $\mathbf{che}_1\mathbf{e}_2\mathbf{se}_3$, in which the **e**'s are distinguished from each other by a subscript. Then $N = 6$ and there are $6! = 720$ distinguishable ways of arranging the characters. By counting in this way, we have reckoned that $\mathbf{che}_1\mathbf{e}_2\mathbf{se}_3$ is different from $\mathbf{che}_2\mathbf{e}_1\mathbf{se}_3$. This correct spelling is counted exactly six times because there are six permutations of the subscripted **e**'s. There are also exactly six permutations of the **e**'s in every other specific sequence. For example:

$$\begin{array}{lll} \mathbf{se}_1\mathbf{e}_2\mathbf{che}_3 & \mathbf{se}_1\mathbf{e}_3\mathbf{che}_2 & \mathbf{se}_2\mathbf{e}_1\mathbf{che}_3 \\ \mathbf{se}_2\mathbf{e}_3\mathbf{che}_1 & \mathbf{se}_3\mathbf{e}_1\mathbf{che}_2 & \mathbf{se}_3\mathbf{e}_2\mathbf{che}_1 \end{array}$$

There are $3! = 6$ permutations of $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ for every sequence of the other characters. So when the **e**'s are indistinguishable, there are $6!/3!$ permutations of the letters in the word cheese. The 3! in the denominator corrects for the indistinguishability of the **e**'s. In general, the denominator needs a factor to account for the indistinguishability of each type of character, so $W = N!/(n_c!n_h!n_e!n_s!) = 6!/(1!1!3!1!) = 120$ is the number of different sequences if the **e**'s are indistinguishable from each other.

EXAMPLE 1.17 Another mixed sequence. For the word **freezer**, you have three indistinguishable **e**'s and two indistinguishable **r**'s. There are $7!/(3!2!)$ permutations of the letters that spell freezer.

In general, for a collection of N objects with t categories, of which n_i objects in each category are *indistinguishable* from one another, but distinguishable from the objects in the other $t-1$ categories, the number of permutations W is

$$W = \frac{N!}{n_1!n_2!\cdots n_t!}. \tag{1.21}$$

When there are only two categories (success/failure, or heads/tails, ...), $t = 2$, so $W(n, N)$, the number of sequences with n successes out of N trials, is

$$W(n, N) = \binom{N}{n} = \frac{N!}{n!(N-n)!}, \tag{1.22}$$

where the shorthand notation $\binom{N}{n}$ for combinations is pronounced 'N choose n.' Use $W = N!$ if you can distinguish every single sequence from every other, or $W = N!/n!$ if only heads are indistinguishable from each other, and tails are distinguishable. Or use Equation (1.22) if tails are indistinguishable from other tails, and heads are indistinguishable from other heads, the case we'll be most interested in subsequently. Example 1.18 applies Equation (1.22) to coin flips and die rolls.

EXAMPLE 1.18 Counting sequences of coin flips and die rolls. You flip a coin $N = 4$ times. How many different sequences have three heads? According to Equation (1.22),

$$W(n_H, N) = \frac{N!}{n_H!n_T!} = \frac{4!}{3!1!} = 4.$$

They are $THHH$, $HTHH$, $HHTH$, and $HHHT$. How many different sequences have two heads?

$$W(2, 4) = \frac{4!}{2!2!} = 6.$$

They are $TTHH$, $HHTT$, $THTH$, $HTHT$, $THHT$, and $HTTH$.

You flip a coin 117 times. How many different sequences have 36 heads?

$$W(36, 117) = \frac{117!}{36!81!} \approx 1.84\times10^{30}.$$

We won't write the sequences out.

You roll a die 15 times. How many different sequences have three **1**'s, one **2**, one **3**, five **4**'s, two **5**'s, and three **6**'s? According to Equation (1.21),

$$W = \frac{15!}{3!1!1!5!2!3!} = 151{,}351{,}200.$$

EXAMPLE 1.19 What is the probability of drawing a royal flush in poker? There are four different ways to draw a royal flush in poker: an ace, king, jack, queen, and ten, all from any one of the four suits. To compute the probability, you need to know how many five-card hands there are in a deck of 52 cards. Use the barrel metaphor: put the 52 cards in the barrel. On the first draw, there are 52 possibilities. On the second draw, there are 51 possibilities, etc. In five draws, there are

$$\frac{52\cdot 51\cdot 50\cdot 49\cdot 48}{5!} = \frac{52!}{5!(52-5)!} = 2{,}598{,}960$$

possible poker hands. The 5! in the denominator corrects for all the possible permutations of each sequence (you don't care whether you draw the king or the ace first, for example). The probability is $1/(2{,}598{,}960)$ of drawing a royal flush in one suit or $4/(2{,}598{,}960) = 1.5\times 10^{-6}$ that you will draw a royal flush in any of the four suits.

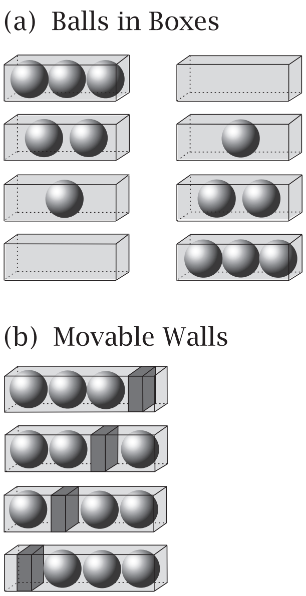


Figure 1.3 How many ways can you put $n = 3$ balls into $M = 2$ boxes (Example 1.20)? (a) There are four ways to partition $n = 3$ balls into $M = 2$ boxes when each box can hold any number of balls. (b) Look at this as four ways to partition three balls and one movable wall.

Here's an example of a type of counting problem in statistical thermodynamics.

EXAMPLE 1.20 Bose–Einstein statistics. How many ways can n indistinguishable particles be put into M boxes, with any number of particles per box? This type of counting is needed to predict the properties of particles called *bosons*, such as photons and ^{4}He atoms. Bose–Einstein statistics counts the ways that n particles can be distributed in M different energy levels, when several particles can occupy the same quantum mechanical energy levels. For now, our interest is not in the physics, but just in the counting problem. Figure 1.3 shows that one way to count the number of arrangements is to think of the system as a linear array of n particles interspersed with $M-1$ movable walls that partition the system into M boxes (spaces between walls). There are $M+n-1$ objects, counting walls plus particles. If the objects were all distinguishable, there would be $(M+n-1)!$ arrangements. However, because the n particles are indistinguishable from each other and each of the $M-1$ walls is indistinguishable from the other walls, and because the walls are distinguishable from the particles, the number of arrangements is

$$W(n,M) = \frac{(M+n-1)!}{(M-1)!n!}. \tag{1.23}$$

Collections of Probabilities Are Described by Distribution Functions

The probabilities of events can be described by *probability distribution functions*. For t mutually exclusive outcomes, $i = 1, 2, 3, \ldots, t$, the distribution

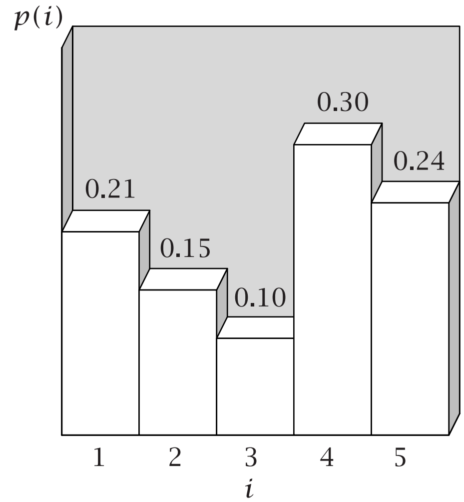


Figure 1.4 A probability distribution function. The possible outcomes are indexed on the horizontal axis. The probability of each outcome is shown on the vertical axis. In this example, outcome **4** is the most probable and outcome **3** is the least probable.

function is $p(i)$, the set of probabilities of all the outcomes. Figure 1.4 shows a probability distribution function for a system with $t = 5$ outcomes.

A property of probability distribution functions is that the sum of the probabilities equals 1. Because the outcomes are mutually exclusive and collectively exhaustive, Equations (1.3) and (1.5) apply and

$$\sum_{i=1}^{t} p(i) = 1. \tag{1.24}$$

For some types of events, the order of the outcomes $i = 1, 2, 3, \ldots, t$ has meaning. For others, it does not. For statistical thermodynamics, the order usually has meaning, and i represents the value of some physical quantity. On the other hand, the index i may be just a label. The index $i = 1, 2, 3$ can represent the colors of socks, [red, green, blue], or [green, red, blue], where the order is irrelevant. Probability distributions can describe either case.

Summations

The sigma notation means to sum terms. For example,

$$\sum_{i=1}^{6} ip_i = p_1 + 2p_2 + 3p_3 + 4p_4 + 5p_5 + 6p_6 \tag{1.25}$$

means 'sum the quantity ip_i from $i = 1$ up to $i = 6$.' Sometimes the index i above and/or below the sigma is omitted in concise shorthand expressions.

Continuous Probability Distribution Functions

In some situations, the outcomes of an event are best represented by a continuous variable x rather than by a discrete variable. Think of a bell curve. Or, for example, a particle might have some probability $p(x)\,dx$ of being between position $x = 1.62$ cm and $x + dx = 1.63$ cm or $p(\theta)\,d\theta$ of having an orientation angle between $\theta = 25.6°$ and $\theta + d\theta = 25.8°$. If x is continuous, $p(x)$ is called a *probability density*, because $p(x)\,dx$ is the probability of finding a value between x and $x+dx$. If x ranges from $x=a$ to $x=b$, Equation (1.24) becomes

$$\int_a^b p(x)\,dx = 1. \tag{1.26}$$

Some distribution functions aren't *normalized*: the statistical weights do not sum to 1. Then you first need to normalize them. For a continuous distribution function $g(x)$, where x ranges from a to b, you can normalize to form a proper probability distribution function. Find the normalization constant g_0 by integrating over x:

$$g_0 = \int_a^b g(x)\,dx. \tag{1.27}$$

The normalized probability density is

$$p(x) = \frac{g(x)}{g_0} = \frac{g(x)}{\displaystyle\int_a^b g(x)\,dx}. \tag{1.28}$$

The Binomial and Multinomial Distribution Functions

Some probability distribution functions occur frequently in nature, and have simple mathematical expressions. Two of the most useful functions are the binomial and multinomial distributions. These will be the basis for our development of the concept of entropy in Chapter 2. The binomial distribution describes processes in which each independent elementary event has two mutually exclusive outcomes, such as heads/tails, yes/no, up/down, or occupied/vacant. Independent trials with two such possible outcomes are called *Bernoulli trials*. Let's label the two possible outcomes $\bullet$ and $\circ$. Let the probability of $\bullet$ be p. Then the probability of $\circ$ is $1-p$. We choose composite events that are pairs of Bernoulli trials. The probability of $\bullet$ followed by $\circ$ is $P_{\bullet\circ} = p(1-p)$. The probabilities of the four possible composite events are

$$\begin{aligned} P_{\bullet\bullet} &= p^2, & P_{\bullet\circ} &= p(1-p), \\ P_{\circ\bullet} &= (1-p)p, & P_{\circ\circ} &= (1-p)^2. \end{aligned} \tag{1.29}$$

This set of composite events is mutually exclusive and collectively exhaustive.

The same probability rules apply to the composite events that apply to elementary events. For example, Equation (1.24) for the normalization of discrete distributions requires that the probabilities must sum to 1:

$$\begin{aligned} P_{\bullet\bullet}+P_{\bullet\circ}+P_{\circ\bullet}+P_{\circ\circ} &= p^2+2p(1-p)+(1-p)^2 \\ &= [p+(1-p)]^2 = 1. \end{aligned} \tag{1.30}$$

In Example 1.7, we defined composite events as pairs of elementary events. More generally, a composite event is a sequence of N repetitions of independent elementary events. The probability of a *specific sequence* of n $\bullet$'s and $N-n$ $\circ$'s is given by Equation (1.7). What is the probability that a series of N trials has

n ●'s and $N-n$ ○'s *in any order*? Equation (1.22) gives the total number of sequences that have n ●'s and $N-n$ ○'s. The product of Equations (1.7) and (1.22) gives the probability of n ●'s and $N-n$ ○'s irrespective of their sequence. This is the **binomial distribution**:

$$P(n,N) = p^n(1-p)^{N-n}\frac{N!}{n!(N-n)!}. \tag{1.31}$$

Because the set of all possible sequences of N trials is mutually exclusive and collectively exhaustive, the composite probabilities sum to one, $\sum_{n=0}^{N} P(n,N) = 1$, as illustrated below.

A simple visualization of the combinatoric terms in the binomial distribution is *Pascal's triangle*. Make a triangle in which the lines are numbered $N = 0, 1, 2, \ldots$. Compute $N!/[n!(N-n)!]$ at each position:

$$\begin{array}{lccccccccccc}
N=0 & & & & & & 1 & & & & & \\
N=1 & & & & & 1 & & 1 & & & & \\
N=2 & & & & 1 & & 2 & & 1 & & & \\
N=3 & & & 1 & & 3 & & 3 & & 1 & & \\
N=4 & & 1 & & 4 & & 6 & & 4 & & 1 & \\
N=5 & 1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}$$

Each term in Pascal's triangle is the sum of the two terms to the left and right from the line above it. Pascal's triangle gives the coefficients in the expansion of $(x+y)^N$. For example, for $N = 4$, using $x = p$ and $y = 1-p$, Equation (1.31) is

$$\begin{aligned}[p+(1-p)]^4 &= p^4+4p^3(1-p)+6p^2(1-p)^2 \\ &\quad + 4p(1-p)^3+(1-p)^4.\end{aligned} \tag{1.32}$$

This sums to one, $\sum_{n=0}^{N} P(n,N) = 1$, because $[p+(1-p)]^N = 1$.

EXAMPLE 1.21 Distribution of coin flips. Figure 1.5 shows a distribution function, the probability $p(n_H,N)$ of observing n_H heads in $N=4$ coin flips, given by Equation (1.31) with $p=0.5$. This shows that in four coin flips, the most probable number of heads is two. It is least probable that all four will be heads or all four will be tails.

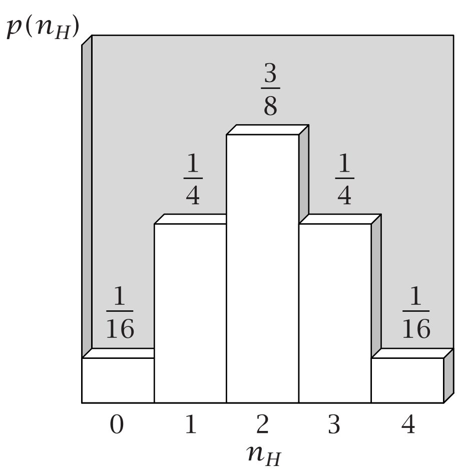


Figure 1.5 The probability distribution for the numbers of heads in four coin flips in Example 1.21.

The multinomial probability distribution is a generalization of the binomial probability distribution. A binomial distribution describes two-outcome events such as coin flips. A multinomial probability distribution applies to t-outcome events where n_i is the number of times that outcome $i = 1, 2, 3, \ldots, t$ appears. For example, $t = 6$ for die rolls. For the multinomial distribution, the number of distinguishable outcomes is given by Equation (1.21): $W = N!/(n_1!n_2!n_3! \cdots n_t!)$. The **multinomial probability distribution** is

$$P(n_1, n_2, \ldots, n_t, N) = p_1^{n_1} p_2^{n_2} p_3^{n_3} \cdots p_t^{n_t} \left(\frac{N!}{n_1!n_2! \cdots n_t!} \right), \tag{1.33}$$

where each factor $n_i!$ accounts for the indistinguishability of objects in category i. The n_i are constrained by the condition $\sum_{i=1}^{t} n_i = N$.

Distribution Functions Have Average Values and Standard Deviations

Averages

A probability distribution function contains all the information that can be known about a probabilistic system. A full distribution function, however, is rarely accessible from experiments. Generally, experiments can measure only certain *moments* of the distribution. The ***n*th moment** of a probability distribution function $p(x)$ is

$$\langle x^n \rangle = \int_a^b x^n p(x)\, dx = \frac{\int_a^b x^n g(x)\, dx}{\int_a^b g(x)\, dx}, \tag{1.34}$$

where the second expression is appropriate for a non-normalized distribution function $g(x)$. Angle brackets $\langle\ \rangle$ are used to indicate the moments, also called the expectation values or averages, of a distribution function. For a *probability* distribution the zeroth moment always equals one, because the sum of the probabilities equals one. The first moment of a distribution function ($n = 1$ in Equation (1.34)) is called the mean, average, or expected value. For discrete functions,

$$\langle i \rangle = \sum_{i=1}^{t} i p(i), \tag{1.35}$$

and for continuous functions,

$$\langle x \rangle = \int_a^b x p(x)\, dx. \tag{1.36}$$

For distributions over t discrete values, the mean of a function $f(i)$ is

$$\langle f(i) \rangle = \sum_{i=1}^{t} f(i) p(i). \tag{1.37}$$

For distributions over continuous values, the mean of a function $f(x)$ is

$$\langle f(x)\rangle = \int_a^b f(x)p(x)\,dx = \frac{\int_a^b f(x)g(x)\,dx}{\int_a^b g(x)\,dx}. \tag{1.38}$$

Equations (1.35)–(1.38) quantify the familiar notion of average, as Example 1.22 shows.

EXAMPLE 1.22 Taking an average. The average of the set of numbers $[\mathbf{3},\mathbf{3},\mathbf{2},\mathbf{2},\mathbf{2},\mathbf{1},\mathbf{1}]$ is 2. The average may be computed by the usual procedure of summing the numbers and dividing by the number of entries. Let's compute the average using Equation (1.35) instead. Since two of the seven outcomes are **3**'s, the probability of a **3** is $p(\mathbf{3}) = 2/7$. Similarly, three of the seven outcomes are **2**'s, so $p(\mathbf{2}) = 3/7$, and two of the seven outcomes are **1**'s, so $p(\mathbf{1}) = 2/7$. The average $\langle i\rangle$ is then

$$\begin{aligned}\langle i\rangle &= \sum_{i=1}^{3} ip(i) = 1p(\mathbf{1})+2p(\mathbf{2})+3p(\mathbf{3})\\ &= 1\left(\frac{2}{7}\right)+2\left(\frac{3}{7}\right)+3\left(\frac{2}{7}\right) = 2.\end{aligned} \tag{1.39}$$

Here are two useful and general properties of averages, derived from the definition given in Equation (1.38):

$$\begin{aligned}\langle af(x)\rangle &= \int af(x)p(x)\,dx = a\int f(x)p(x)\,dx\\ &= a\langle f(x)\rangle, \qquad \text{where } a \text{ is a constant.}\end{aligned} \tag{1.40}$$

$$\begin{aligned}\langle f(x)+g(x)\rangle &= \int [f(x)+g(x)]\,p(x)\,dx\\ &= \int f(x)p(x)\,dx + \int g(x)p(x)\,dx\\ &= \langle f(x)\rangle + \langle g(x)\rangle.\end{aligned} \tag{1.41}$$

Variance

The *variance* σ^2 is a measure of the width of a distribution. A broad, flat distribution has a large variance, while a narrow, peaked distribution has a small variance. The variance σ^2 is defined as the average square deviation from the mean,

$$\sigma^2 = \langle (x-a)^2\rangle = \langle x^2 - 2ax + a^2\rangle, \tag{1.42}$$

where $a = \langle x\rangle$ is the mean value, or first moment. We use a instead of $\langle x\rangle$ as a reminder here that this quantity is just a constant, not a variable. Using

Equation (1.41), Equation (1.42) becomes

$$\sigma^2 = \langle x^2 \rangle - \langle 2ax \rangle + \langle a^2 \rangle.$$

Using Equation (1.40),

$$\sigma^2 = \langle x^2 \rangle - 2a\langle x \rangle + a^2 = \langle x^2 \rangle - \langle x \rangle^2. \tag{1.43}$$

Second-moment quantities are important for understanding heat capacities (Chapter 12), random walks (Chapter 18), diffusion (Chapter 17), and polymer chain conformations (Chapters 32–34). The square root of the variance is σ, which is also called the *standard deviation*. Moments higher than the second describe asymmetries in the shape of the distribution. Examples 1.23–1.26 show calculations of means and variances for discrete and continuous probability distributions.

EXAMPLE 1.23 Coin flips: mean and variance. Compute the average number of heads $\langle n_H \rangle$ in $N = 4$ coin flips by using the distribution in Example 1.21:

$$\begin{aligned}\langle n_H \rangle &= \sum_{n_H=0}^{4} n_H p(n_H, N) \\ &= 0\left(\frac{1}{16}\right) + 1\left(\frac{4}{16}\right) + 2\left(\frac{6}{16}\right) + 3\left(\frac{4}{16}\right) + 4\left(\frac{1}{16}\right) = 2,\end{aligned}$$

and

$$\begin{aligned}\langle n_H^2 \rangle &= \sum_{n_H=0}^{4} n_H^2 p(n_H, N) \\ &= 0\left(\frac{1}{16}\right) + 1\left(\frac{4}{16}\right) + 4\left(\frac{6}{16}\right) + 9\left(\frac{4}{16}\right) + 16\left(\frac{1}{16}\right) = 5.\end{aligned}$$

According to Equation (1.43), the variance σ^2 is

$$\sigma^2 = \langle n_H^2 \rangle - \langle n_H \rangle^2 = 5 - 2^2 = 1.$$

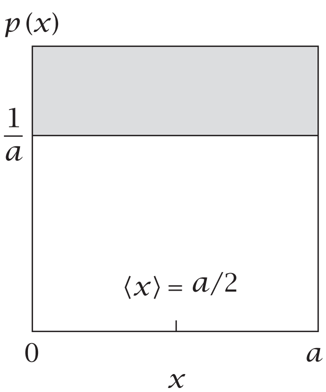


Figure 1.6 Flat distribution function, $0 \le x \le a$. The average value is $\langle x \rangle = a/2$ (see Example 1.24).

EXAMPLE 1.24 The average and variance of a continuous function. Suppose you have a flat probability distribution $p(x) = 1/a$ (shown in Figure 1.6) for a variable $0 \le x \le a$. To compute $\langle x \rangle$, use Equation (1.36):

$$\langle x \rangle = \int_0^a x p(x)\, dx = \frac{1}{a}\int_0^a x\, dx = \left(\frac{1}{a}\right)\frac{x^2}{2}\bigg|_0^a = \frac{a}{2}.$$

Equation (1.34) gives the second moment $\langle x^2 \rangle$:

$$\langle x^2 \rangle = \int_0^a x^2 p(x)\, dx = \frac{1}{a}\int_0^a x^2\, dx = \left(\frac{1}{a}\right)\frac{x^3}{3}\bigg|_0^a = \frac{a^2}{3}.$$

The variance is $\langle x^2 \rangle - \langle x \rangle^2 = a^2/3 - a^2/4 = a^2/12$.

EXAMPLE 1.25 The average of an exponential distribution. Figure 1.7 shows a distribution function $g(x) = e^{-ax}$ over the range $0 \le x \le \infty$. First normalize $g(x)$ to make it a probability distribution. According to Equation (1.28), $p(x) = g(x)/g_0$. Integrate $g(x)$ to determine g_0:

$$g_0 = \int_0^\infty e^{-ax}\,dx = -\left(\frac{1}{a}\right)e^{-ax}\Big|_0^\infty = \frac{1}{a} \qquad \text{for } a > 0.$$

The normalized distribution function is $p(x) = g(x)/g_0 = ae^{-ax}$. Now, to compute $\langle x\rangle$ for this distribution, use Equation (1.34):

$$\langle x\rangle = \int_0^\infty xp(x)\,dx = a\int_0^\infty xe^{-ax}\,dx$$
$$= -\left[e^{-ax}\left(x+\frac{1}{a}\right)\right]\Big|_0^\infty = \frac{1}{a}.$$

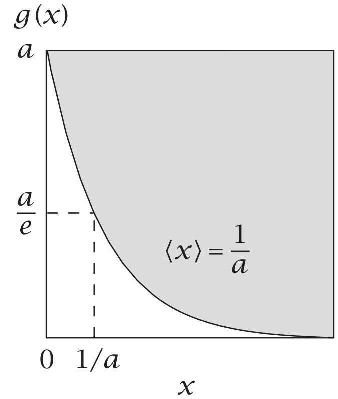


Figure 1.7 Exponential distribution function, $0 \le x \le \infty$. The average of $p(x) = ae^{-ax}$ is $\langle x\rangle = 1/a$ (see Example 1.25).

EXAMPLE 1.26 Averaging the orientations of a vector. For predicting the conformations of a polymer or spectroscopic properties, you may have a vector that is free to orient uniformly over all possible angles θ. If you want to compute its average projection on an axis, using quantities such as $\langle\cos\theta\rangle$ or $\langle\cos^2\theta\rangle$, put the beginning of the vector at the center of a sphere. If the vector orients uniformly, it points to any given patch on the surface of the sphere in proportion to the area of that patch.

The strip of area shown in Figure 1.8 has an angle θ with respect to the z-axis. The area of the strip is $(r\,d\theta)(2\pi\ell)$. Since $\ell = r\sin\theta$, the area of the strip is $2\pi r^2\sin\theta\,d\theta$. A strip has less area if θ is small than if θ approaches 90°. The fraction of vectors $p(\theta)$ that point to, or end in, this strip is

$$p(\theta) = \frac{2\pi r^2\sin\theta\,d\theta}{\displaystyle\int_0^\pi 2\pi r^2\sin\theta\,d\theta} = \frac{\sin\theta\,d\theta}{\displaystyle\int_0^\pi \sin\theta\,d\theta}. \tag{1.44}$$

The average $\langle\cos\theta\rangle$ over all vectors is

$$\langle\cos\theta\rangle = \int_0^\pi \cos\theta p(\theta)\,d\theta = \frac{\displaystyle\int_0^\pi \cos\theta\sin\theta\,d\theta}{\displaystyle\int_0^\pi \sin\theta\,d\theta}. \tag{1.45}$$

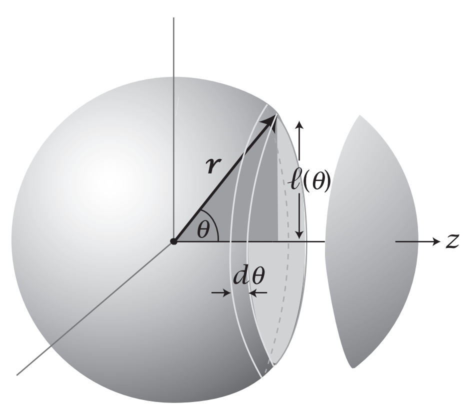


Figure 1.8 A vector that can orient in all directions can be represented as starting at the origin and ending on the surface of a sphere. The area $2\pi r\ell\,d\theta$ represents the relative proportion of all the vectors that land in the strip at an angle between θ and $\theta+d\theta$ relative to the z axis.

This integration is simplified by noticing that $\sin\theta\, d\theta = -d\cos\theta$, by letting $x = \cos\theta$, and by replacing the limits 0 and π by 1 and -1. Then Equation (1.45) becomes

$$\langle \cos\theta \rangle = \frac{\int_1^{-1} x\, dx}{\int_1^{-1} dx} = \frac{\frac{1}{2}x^2\Big|_1^{-1}}{x\Big|_1^{-1}} = 0. \tag{1.46}$$

Physically, this says that the average projection on the z axis of uniformly distributed vectors is zero. You can also see this by symmetry: just as many vectors point forward ($0° < \theta \le 90°$) as backward ($90° < \theta \le 180°$), so the average is zero.

Later we will find the quantity $\langle \cos^2\theta \rangle$ to be useful. Following the same logic, you have

$$\langle \cos^2\theta \rangle = \frac{\int_0^{\pi} \cos^2\theta \sin\theta\, d\theta}{\int_0^{\pi} \sin\theta\, d\theta} = \frac{\int_1^{-1} x^2\, dx}{\int_1^{-1} dx} = \frac{\frac{1}{3}x^3\Big|_1^{-1}}{x\Big|_1^{-1}} = \frac{1}{3}. \tag{1.47}$$

Summary

Probabilities describe frequencies or incomplete knowledge. The addition and multiplication rules allow you to draw consistent inferences about probabilities of multiple events. Distribution functions describe collections of probabilities. Such functions have mean values and variances. Combined with combinatorics—the counting of arrangements of systems—probabilities provide the basis for reasoning about entropy, and about driving forces among molecules, described in the next chapter.

Examples of Distributions

Here are some probability distribution functions that commonly appear in statistical mechanics.

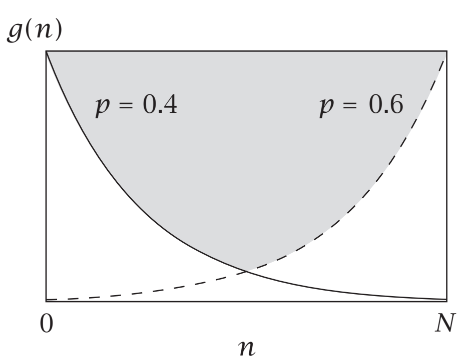


Figure 1.9 Bernoulli

$$g(n) = p^n(1-p)^{N-n},$$
$$n = 0, 1, 2, \ldots, N. \tag{1.48}$$

The Bernoulli distribution describes independent trials with two possible outcomes (see page 16). $g(n)$ is a distribution function, not a probability, because it is not normalized to sum to one.

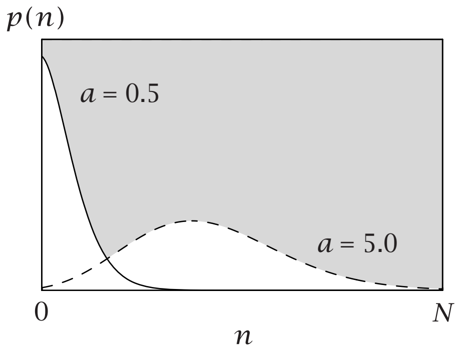


Figure 1.10 Poisson

$$p(n) = \frac{a^n e^{-a}}{n!},$$

$$n = 0, 1, 2, \ldots, N. \qquad (1.49)$$

The Poisson distribution approximates the binomial distribution when the number of trials is large and the probability of each one is small [3]. It is useful for describing radioactive decay, the number of vacancies in the Supreme Court each year [4], the numbers of dye molecules taken up by small particles, or the sizes of colloidal particles.

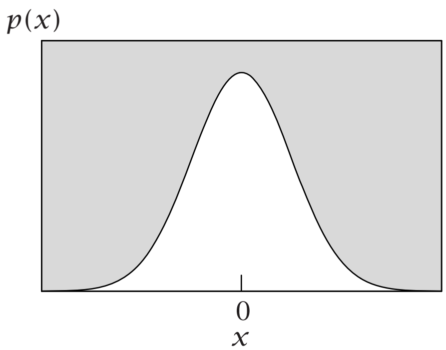


Figure 1.11 Gaussian

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2},$$

$$-\infty \le x \le \infty. \qquad (1.50)$$

The Gaussian distribution is derived from the binomial distribution for large N [5]. It is important for statistics, error analysis, diffusion, conformations of polymer chains, and the Maxwell–Boltzmann distribution law of gas velocities.

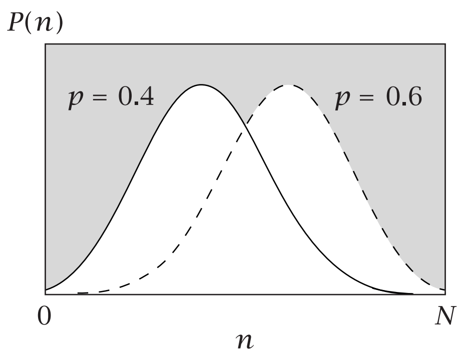


Figure 1.12 Binomial

$$P(n) = p^n (1-p)^{N-n} \times \left(\frac{N!}{n!(N-n)!}\right),$$

$$n = 0, 1, 2, \ldots, N. \qquad (1.51)$$

The binomial distribution for collections of Bernoulli trials is derived on pages 16–17.

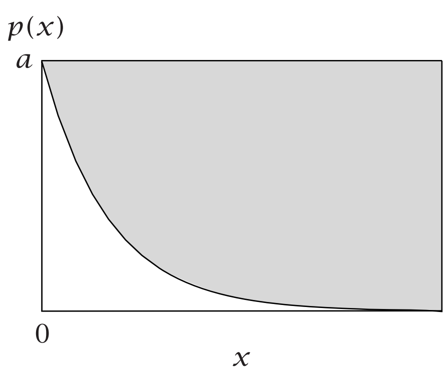


Figure 1.13 Exponential (Boltzmann)

$$p(x) = ae^{-ax},$$

$$0 \le x \le \infty. \qquad (1.52)$$

The exponential, or Boltzmann distribution, is central to statistical thermodynamics (see Chapters 5 and 10).

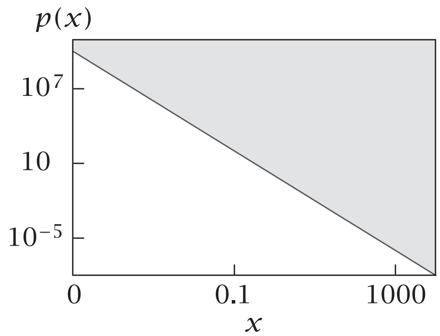


Figure 1.14 Power law

$$p(x) = 1/x^q, \tag{1.53}$$

where q is a constant called the *power law exponent*. Power laws describe the frequencies of earthquakes, the numbers of links to World Wide Web sites, the distribution of incomes ('the rich get richer'), and the noise spectrum in some electronic devices.

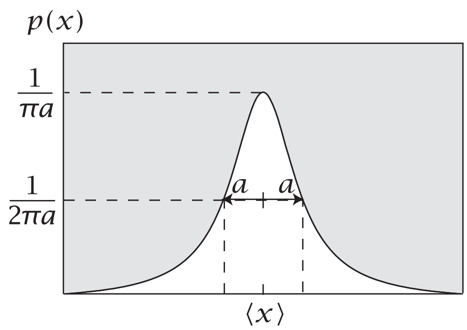


Figure 1.15 Lorentzian

$$p(x) = \frac{1}{\pi}\frac{a}{(x-\langle x\rangle)^2+a^2}, \quad -\infty \le x \le \infty. \tag{1.54}$$

$2a$ is the width of the Lorentzian curve at the level of half the maximum probability. Lorentzian distributions are useful in spectroscopy [3].

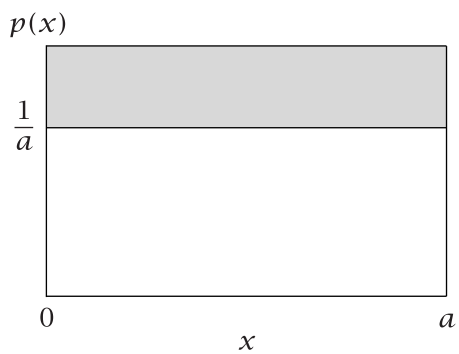


Figure 1.16 Flat

$$p(x) = 1/a, \tag{1.55}$$

where a is a constant independent of x (see Example 1.24).