1

Experiments, Models, and Probabilities

Getting Started with Probability

The title of this book is *Probability and Stochastic Processes.* We say and hear and read the word *probability* and its relatives (*possible, probable, probably*) in many contexts. Within the realm of applied mathematics, the meaning of *probability* is a question that has occupied mathematicians, philosophers, scientists, and social scientists for hundreds of years.

Everyone accepts that the probability of an event is a number between 0 and 1. Some people interpret probability as a physical property (like mass or volume or temperature) that can be measured. This is tempting when we talk about the probability that a coin flip will come up heads. This probability is closely related to the nature of the coin. Fiddling around with the coin can alter the probability of heads.

Another interpretation of probability relates to the knowledge that we have about something. We might assign a low probability to the truth of the statement, *It is raining now in Phoenix, Arizona,* because we know that Phoenix is in the desert. However, our knowledge changes if we learn that it was raining an hour ago in Phoenix. This knowledge would cause us to assign a higher probability to the truth of the statement, *It is raining now in Phoenix.*

Both views are useful when we apply probability theory to practical problems. Whichever view we take, we will rely on the abstract mathematics of probability, which consists of definitions, axioms, and inferences (theorems) that follow from the axioms. While the structure of the subject conforms to principles of pure logic, the terminology is not entirely abstract. Instead, it reflects the practical origins of probability theory, which was developed to describe phenomena that cannot be predicted with certainty. The point of view is different from the one we took when we started studying physics. There we said that if we do the same thing in the same way over and over again — send a space shuttle into orbit, for example —

the result will always be the same. To predict the result, we have to take account of all relevant facts.

The mathematics of probability begins when the situation is so complex that we just can't replicate everything important exactly, like when we fabricate and test an integrated circuit. In this case, repetitions of the same procedure yield different results. The situation is not totally chaotic, however. While each outcome may be unpredictable, there are consistent patterns to be observed when we repeat the procedure a large number of times. Understanding these patterns helps engineers establish test procedures to ensure that a factory meets quality objectives. In this repeatable procedure (making and testing a chip) with unpredictable outcomes (the quality of individual chips), the *probability* is a number between 0 and 1 that states the proportion of times we expect a certain thing to happen, such as the proportion of chips that pass a test.

As an introduction to probability and stochastic processes, this book serves three purposes:

- It introduces students to the logic of probability theory.
- It helps students develop intuition into how the theory relates to practical situations.
- It teaches students how to apply probability theory to solving engineering problems.

To exhibit the logic of the subject, we show clearly in the text three categories of theoretical material: definitions, axioms, and theorems. Definitions establish the logic of probability theory, and axioms are facts that we accept without proof. Theorems are consequences that follow logically from definitions and axioms. Each theorem has a proof that refers to definitions, axioms, and other theorems. Although there are dozens of definitions and theorems, there are only three axioms of probability theory. These three axioms are the foundation on which the entire subject rests. To meet our goal of presenting the logic of the subject, we could set out the material as dozens of definitions followed by three axioms followed by dozens of theorems. Each theorem would be accompanied by a complete proof.

While rigorous, this approach would completely fail to meet our second aim of conveying the intuition necessary to work on practical problems. To address this goal, we augment the purely mathematical material with a large number of examples of practical phenomena that can be analyzed by means of probability theory. We also interleave definitions and theorems, presenting some theorems with complete proofs, presenting others with partial proofs, and omitting some proofs altogether. We find that most engineering students study probability with the aim of using it to solve practical problems, and we cater mostly to this goal. We also encourage students to take an interest in the logic of the subject — it is very elegant — and we feel that the material presented is sufficient to enable these students to fill in the gaps we have left in the proofs.

Therefore, as you read this book you will find a progression of definitions, axioms, theorems, more definitions, and more theorems, all interleaved with examples and comments designed to contribute to your understanding of the theory. We also include brief quizzes that you should try to solve as you read the book. Each one

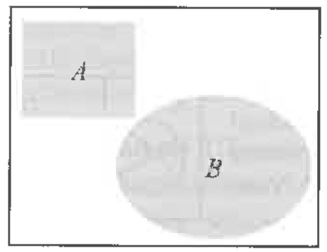
will help you decide whether you have grasped the material presented just before the quiz. The problems at the end of each chapter give you more practice applying the material introduced in the chapter. They vary considerably in their level of difficulty. Some of them take you more deeply into the subject than the examples and quizzes do.

1.1 Applying Set Theory to Probability

Probability is based on a repeatable experiment that consists of a procedure and observations. An *outcome* is an observation. An *event* is a set of outcomes.

The mathematical basis of probability is the theory of sets. We assume you have already encountered set theory and are familiar with such terms as *set, element, union, intersection* and *complement*. To work with sets mathematically it is necessary to define a *universal set*. This is the set of all things that we could possibly consider in a given context. We will use the letter S to denote the universal set. The *null set*, which is also important, may seem like it is not a set at all. By definition it has no elements. The notation for the null set is $\varnothing$. By definition $\varnothing$ is a subset of every set. For any set A, $\varnothing \subset A$.

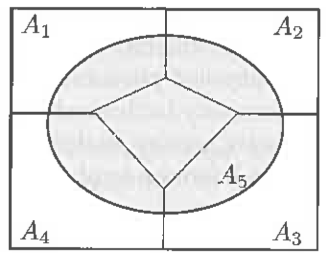
In working with probability we will often refer to two important properties of collections of sets. Here are the definitions.



A collection of sets $A_1, \ldots, A_n$ is *mutually exclusive* if and only if

$$A_i \cap A_j = \varnothing, \qquad i \neq j. \tag{1.1}$$

The word *disjoint* is sometimes used as a synonym for mutually exclusive.



A collection of sets $A_1, \ldots, A_n$ is *collectively exhaustive* if and only if

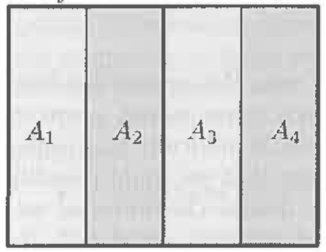
$$A_1 \cup A_2 \cup \cdots \cup A_n = S. \tag{1.2}$$

In the definition of *collectively exhaustive*, we used the somewhat cumbersome notation $A_1 \cup A_2 \cup \cdots \cup A_n$ for the union of N sets. Just as $\sum_{i=1}^{n} x_i$ is a shorthand for $x_1+x_2+\cdots+x_n$, we will use a shorthand for unions and intersections of n sets:

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n, \tag{1.3}$$

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n. \tag{1.4}$$

We will see that collections of sets that are both mutually exclusive and collectively exhaustive are sufficiently useful to merit a definition.

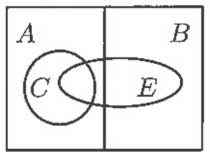


A collection of sets $A_1, \ldots, A_n$ is a *partition* if it is both mutually exclusive and collectively exhaustive.

Example 1.1

Phonesmart offers customers two kinds of smart phones, Apricot (A) and Banana (B). It is possible to buy a Banana phone with an optional external battery E. Apricot customers can buy a phone with an external battery (E) or an extra memory card (C) or both. Draw a Venn diagram that shows the relationship among the items A,B,C and E available to Phonesmart customers.

Since each phone is either Apricot or Banana, A and B form a partition. Since the external battery E is available for both kinds of phones, E intersects both A and B. However, since the memory card C is available only to Apricot customers, $C \subset A$. A Venn diagram representing these facts is shown on the right.



The mathematics we study is a branch of measure theory. Probability is a number that describes a set. The higher the number, the more probability there is. In this sense probability is like a quantity that measures a physical phenomenon; for example, a weight or a temperature. However, it is not necessary to think about probability in physical terms. We can do all the math abstractly, just as we defined sets and set operations in the previous paragraphs without any reference to physical phenomena.

Fortunately for engineers, the language of probability (including the word *probability* itself) makes us think of things that we experience. The basic model is a repeatable *experiment*. An experiment consists of a *procedure* and *observations*. There is uncertainty in what will be observed; otherwise, performing the experiment would be unnecessary. Some examples of experiments include

1. Flip a coin. Did it land with heads or tails facing up?
2. Walk to a bus stop. How long do you wait for the arrival of a bus?
3. Give a lecture. How many students are seated in the fourth row?
4. Transmit one of a collection of waveforms over a channel. What waveform arrives at the receiver?
5. Transmit one of a collection of waveforms over a channel. Which waveform does the receiver identify as the transmitted waveform?

For the most part, we will analyze *models* of actual physical experiments. We create models because real experiments generally are too complicated to analyze. For example, to describe *all* of the factors affecting your waiting time at a bus stop, you may consider

- The time of day. (Is it rush hour?)
- The speed of each car that passed by while you waited.
- The weight, horsepower, and gear ratios of each kind of bus used by the bus company.
- The psychological profile and work schedule of each bus driver. (Some drivers drive faster than others.)
- The status of all road construction within 100 miles of the bus stop.

It should be apparent that it would be difficult to analyze the effect of each of these factors on the likelihood that you will wait less than five minutes for a bus. Consequently, it is necessary to study a *model* of the experiment that captures the important part of the actual physical experiment. Since we will focus on the model of the experiment almost exclusively, we often will use the word *experiment* to refer to the model of an experiment.

Example 1.2

An experiment consists of the following procedure, observation, and model:

- Procedure: Monitor activity at a Phonesmart store.
- Observation: Observe which type of phone (Apricot or Banana) the next customer purchases.
- Model: Apricots and Bananas are equally likely. The result of each purchase is unrelated to the results of previous purchases.

As we have said, an experiment consists of both a procedure and observations. It is important to understand that two experiments with the same procedure but with different observations are different experiments. For example, consider these two experiments:

Example 1.3

Monitor the Phonesmart store until three customers purchase phones. Observe the sequence of Apricots and Bananas.

Example 1.4

Monitor the Phonesmart store until three customers purchase phones. Observe the number of Apricots.

These two experiments have the same procedure: monitor the Phonesmart store until three customers purchase phones. They are different experiments because they require different observations. We will describe models of experiments in terms of a set of possible experimental outcomes. In the context of probability, we give precise meaning to the word *outcome*.

Definition 1.1 Outcome

*An **outcome** of an experiment is any possible observation of that experiment.*

Implicit in the definition of an outcome is the notion that each outcome is distinguishable from every other outcome. As a result, we define the universal set of all possible outcomes. In probability terms, we call this universal set the *sample space*.

Definition 1.2 Sample Space

*The **sample space** of an experiment is the finest-grain, mutually exclusive, collectively exhaustive set of all possible outcomes.*

The *finest-grain* property simply means that all possible distinguishable outcomes are identified separately. The requirement that outcomes be mutually exclusive says that if one outcome occurs, then no other outcome also occurs. For the set of outcomes to be collectively exhaustive, every outcome of the experiment must be in the sample space.

Example 1.5

- The sample space in Example 1.2 is $S = \{a, b\}$ where a is the outcome "Apricot sold," and b is the outcome "Banana sold."
- The sample space in Example 1.3 is

$$S = \{aaa, aab, aba, abb, baa, bab, bba, bbb\} \tag{1.5}$$

- The sample space in Example 1.4 is $S = \{0, 1, 2, 3\}$.

In common speech, an event is something that occurs. In an experiment, we may say that an event occurs when a certain phenomenon is observed. To define an event mathematically, we must identify *all* outcomes for which the phenomenon is observed. That is, for each outcome, either the particular event occurs or it does not. In probability terms, we define an event in terms of the outcomes in the sample space.

Set Algebra	Probability
Set	Event
Universal set	Sample space
Element	Outcome

Table 1.1 The terminology of set theory and probability.

Definition 1.3 Event

*An **event** is a set of outcomes of an experiment.*

Table 1.1 relates the terminology of probability to set theory. All of this may seem so simple that it is boring. While this is true of the definitions themselves, applying them is a different matter. Defining the sample space and its outcomes are key elements of the solution of any probability problem. A probability problem arises from some practical situation that can be modeled as an experiment. To work on the problem, it is necessary to define the experiment carefully and then derive the sample space. Getting this right is a big step toward solving the problem.

Example 1.6

Observe the number of minutes a customer spends in the Phonesmart store. An outcome T is a nonnegative real number. The sample space is $S = \{T|T \geq 0\}$. The event "the customer stays longer than five minutes is $\{T|T > 5\}$.

Example 1.7

Monitor three customers in the Phonesmart store. Classify the behavior as buying (b) if a customer purchases a smartphone. Otherwise the behavior is no purchase (n). An outcome of the experiment is a sequence of three customer decisions. We can denote each outcome by a three-letter word such as bnb indicating that the first and third customers buy a phone and the second customer does not. We denote the event that customer i buys a phone by B_i and the event customer i does not buy a phone by N_i. The event $B_2 = \{nbn, nbb, bbn, bbb\}$. We can also express an outcome as an intersection of events B_i and N_j. For example the outcome $bnb = B_1N_2B_3$.

Quiz 1.1

Monitor three consecutive packets going through a Internet router. Based on the packet header, each packet can be classified as either video (v) if it was sent from a Youtube server or as ordinary data (d). Your observation is a sequence of three letters (each letter is either v or d). For example, two video packets followed by one data packet corresponds to vvd. Write the elements of the following sets:

$$A_1 = \{\text{second packet is video}\}, \qquad B_1 = \{\text{second packet is data}\},$$
$$A_2 = \{\text{all packets are the same}\}, \qquad B_2 = \{\text{video and data alternate}\},$$
$$A_3 = \{\text{one or more video packets}\}, \qquad B_3 = \{\text{two or more data packets}\}.$$

For each pair of events A_1 and B_1, A_2 and B_2, and so on, identify whether the pair of events is either mutually exclusive or collectively exhaustive or both.

1.2 Probability Axioms

A probability model assigns a number between 0 and 1 to every event. The probability of the union of mutually exclusive events is the sum of the probabilities of the events in the union.

Thus far our model of an experiment consists of a procedure and observations. This leads to a set-theory representation with a sample space (universal set S), outcomes (s that are elements of S), and events (A that are sets of elements). To complete the model, we assign a probability $P[A]$ to every event, A, in the sample space. With respect to our physical idea of the experiment, the probability of an event is the proportion of the time that event is observed in a large number of runs of the experiment. This is the *relative frequency* notion of probability. Mathematically, this is expressed in the following axioms.

Definition 1.4 Axioms of Probability

A probability measure $P[\cdot]$ *is a function that maps events in the sample space to real numbers such that*

Axiom 1 *For any event* A, $P[A] \geq 0$.

Axiom 2 $P[S] = 1$.

Axiom 3 *For any countable collection* $A_1, A_2, \ldots$ *of mutually exclusive events*

$$P[A_1 \cup A_2 \cup \cdots] = P[A_1] + P[A_2] + \cdots.$$

We will build our entire theory of probability on these three axioms. Axioms 1 and 2 simply establish a probability as a number between 0 and 1. Axiom 3 states that the probability of the union of mutually exclusive events is the sum of the individual probabilities. We will use this axiom over and over in developing the theory of probability and in solving problems. In fact, it is really all we have to work with. Everything else follows from Axiom 3. To use Axiom 3 to solve a practical problem, we will learn in Section 1.4 to analyze a complicated event as the union of mutually exclusive events whose probabilities we can calculate. Then, we

will add the probabilities of the mutually exclusive events to find the probability of the complicated event we are interested in.

A useful extension of Axiom 3 applies to the union of two mutually exclusive events.

Theorem 1.1

For mutually exclusive events A_1 *and* A_2,

$$P[A_1 \cup A_2] = P[A_1] + P[A_2].$$

Although it may appear that Theorem 1.1 is a trivial special case of Axiom 3, this is not so. In fact, a simple proof of Theorem 1.1 may also use Axiom 2! If you are curious, Problem 1.2.13 gives the first steps toward a proof. It is a simple matter to extend Theorem 1.1 to any finite union of mutually exclusive sets.

Theorem 1.2

If $A = A_1 \cup A_2 \cup \cdots \cup A_m$ *and* $A_i \cap A_j = \varnothing$ *for* $i \neq j$, *then*

$$P[A] = \sum_{i=1}^{m} P[A_i].$$

In Chapter 10, we show that the probability measure established by the axioms corresponds to the idea of relative frequency. The correspondence refers to a sequential experiment consisting of n repetitions of the basic experiment. We refer to each repetition of the experiment as a *trial*. In these n trials, $N_A(n)$ is the number of times that event A occurs. The relative frequency of A is the fraction $N_A(n)/n$. Theorem 10.7 proves that $\lim_{n\to\infty} N_A(n)/n = P[A]$.

Here we list some properties of probabilities that follow directly from the three axioms. While we do not supply the proofs, we suggest that students prove at least some of these theorems in order to gain experience working with the axioms.

Theorem 1.3

The probability measure $P[\cdot]$ *satisfies*

(a) $P[\varnothing] = 0$.

(b) $P[A^c] = 1 - P[A]$.

(c) *For any* A *and* B *(not necessarily mutually exclusive),*

$$P[A \cup B] = P[A] + P[B] - P[A \cap B].$$

(d) *If* $A \subset B$, *then* $P[A] \leq P[B]$.

Another consequence of the axioms can be expressed as the following theorem:

Theorem 1.4

The probability of an event $B = \{s_1, s_2, \ldots, s_m\}$ is the sum of the probabilities of the outcomes contained in the event:

$$P[B] = \sum_{i=1}^{m} P[\{s_i\}].$$

Proof Each outcome s_i is an event (a set) with the single element s_i. Since outcomes by definition are mutually exclusive, B can be expressed as the union of m mutually exclusive sets:

$$B = \{s_1\} \cup \{s_2\} \cup \cdots \cup \{s_m\} \tag{1.6}$$

with $\{s_i\} \cap \{s_j\} = \varnothing$ for $i \neq j$. Applying Theorem 1.2 with $B_i = \{s_i\}$ yields

$$P[B] = \sum_{i=1}^{m} P[\{s_i\}]. \tag{1.7}$$

Comments on Notation

We use the notation $P[\cdot]$ to indicate the probability of an event. The expression in the square brackets is an event. Within the context of one experiment, $P[A]$ can be viewed as a function that transforms event A to a number between 0 and 1.

Note that $\{s_i\}$ is the formal notation for a set with the single element s_i. For convenience, we will sometimes write $P[s_i]$ rather than the more complete $P[\{s_i\}]$ to denote the probability of this outcome.

We will also abbreviate the notation for the probability of the intersection of two events, $P[A \cap B]$. Sometimes we will write it as $P[A, B]$ and sometimes as $P[AB]$. Thus by definition, $P[A \cap B] = P[A, B] = P[AB]$.

Equally Likely Outcomes

A large number of experiments have a sample space $S = \{s_1, \ldots, s_n\}$ in which our knowledge of the practical situation leads us to believe that no one outcome is any more likely than any other. In these experiments we say that the n outcomes are *equally likely*. In such a case, the axioms of probability imply that every outcome has probability $1/n$.

Theorem 1.5

For an experiment with sample space $S = \{s_1, \ldots, s_n\}$ in which each outcome s_i is equally likely,

$$P[s_i] = 1/n \qquad 1 \leq i \leq n.$$

Proof Since all outcomes have equal probability, there exists p such that $P[s_i] = p$ for $i = 1, \ldots, n$. Theorem 1.4 implies

$$P[S] = P[s_1] + \cdots + P[s_n] = np. \tag{1.8}$$

Since Axiom 2 says $P[S] = 1$, $p = 1/n$.

Example 1.8

Roll a six-sided die in which all faces are equally likely. What is the probability of each outcome? Find the probabilities of the events: "Roll 4 or higher," "Roll an even number," and "Roll the square of an integer."

The probability of each outcome is $P[i] = 1/6$ for $i = 1, 2, \ldots, 6$. The probabilities of the three events are

- $P[\text{Roll 4 or higher}] = P[4] + P[5] + P[6] = 1/2$.
- $P[\text{Roll an even number}] = P[2] + P[4] + P[6] = 1/2$.
- $P[\text{Roll the square of an integer}] = P[1] + P[4] = 1/3$.

Quiz 1.2

A student's test score T is an integer between 0 and 100 corresponding to the experimental outcomes $s_0, \ldots, s_{100}$. A score of 90 to 100 is an A, 80 to 89 is a B, 70 to 79 is a C, 60 to 69 is a D, and below 60 is a failing grade of F. If all scores between 51 and 100 are equally likely and a score of 50 or less never occurs, find the following probabilities:

(a) $P[\{s_{100}\}]$ (b) $P[A]$

(c) $P[F]$ (d) $P[T < 90]$

(e) $P[\text{a } C \text{ grade or better}]$ (f) $P[\text{student passes}]$

1.3 Conditional Probability

Conditional probabilities correspond to a modified probability model that reflects partial information about the outcome of an experiment. The modified model has a smaller sample space than the original model.

As we suggested earlier, it is sometimes useful to interpret $P[A]$ as our knowledge of the occurrence of event A before an experiment takes place. If $P[A] \approx 1$, we have advance knowledge that A will almost certainly occur. $P[A] \approx 0$ reflects strong knowledge that A is unlikely to occur when the experiment takes place. With $P[A] \approx 1/2$, we have little knowledge about whether or not A will occur. Thus $P[A]$ reflects our knowledge of the occurrence of A *prior* to performing an experiment. Sometimes, we refer to $P[A]$ as the *a priori probability*, or the *prior probability*, of A.

In many practical situations, it is not possible to find out the precise outcome of an experiment. Rather than the outcome s_i, itself, we obtain information that the outcome is in the set B. That is, we learn that some event B has occurred, where B consists of several outcomes. Conditional probability describes our knowledge of A when we know that B has occurred but we still don't know the precise outcome. The notation for this new probability is $P[A|B]$. We read this as "the probability of A given B." Before going to the mathematical definition of conditional probability, we provide an example that gives an indication of how conditional probabilities can be used.

Example 1.9

Consider an experiment that consists of testing two integrated circuits (IC chips) that come from the same silicon wafer and observing in each case whether a chip is accepted (a) or rejected (r). The sample space of the experiment is $S = \{rr, ra, ar, aa\}$. Let B denote the event that the first chip tested is rejected. Mathematically, $B = \{rr, ra\}$. Similarly, let $A = \{rr, ar\}$ denote the event that the second chip is a failure.

The chips come from a high-quality production line. Therefore the prior probability $P[A]$ is very low. In advance, we are pretty certain that the second circuit will be accepted. However, some wafers become contaminated by dust, and these wafers have a high proportion of defective chips. When the first chip is a reject, the outcome of the experiment is in event B and $P[A|B]$, the probability that the second chip will also be rejected, is higher than the *a priori* probability $P[A]$ because of the likelihood that dust contaminated the entire wafer.

Definition 1.5 Conditional Probability

The conditional probability of the event A given the occurrence of the event B is

$$P[A|B] = \frac{P[AB]}{P[B]}.$$

Conditional probability is defined only when $P[B] > 0$. In most experiments, $P[B] = 0$ means that it is certain that B never occurs. In this case, it is illogical to speak of the probability of A given that B occurs. Note that $P[A|B]$ is a respectable probability measure relative to a sample space that consists of all the outcomes in B. This means that $P[A|B]$ has properties corresponding to the three axioms of probability.

Theorem 1.6

A conditional probability measure $P[A|B]$ has the following properties that correspond to the axioms of probability.

Axiom 1: $P[A|B] \geq 0$.

Axiom 2: $P[B|B] = 1$.

Axiom 3: If $A = A_1 \cup A_2 \cup \cdots$ with $A_i \cap A_j = \varnothing$ for $i \neq j$, then

$$P[A|B] = P[A_1|B] + P[A_2|B] + \cdots$$

You should be able to prove these statements using Definition 1.5.

Example 1.10

With respect to Example 1.9, consider the *a priori* probability model

$$P[rr] = 0.01, \quad P[ra] = 0.01, \quad P[ar] = 0.01, \quad P[aa] = 0.97. \tag{1.9}$$

Find the probability of A = "second chip rejected" and B = "first chip rejected." Also find the conditional probability that the second chip is a reject given that the first chip is a reject.

We saw in Example 1.9 that A is the union of two mutually exclusive events (outcomes) rr and ar. Therefore, the a priori probability that the second chip is rejected is

$$P[A] = P[rr] + P[ar] = 0.02 \tag{1.10}$$

This is also the a priori probability that the first chip is rejected:

$$P[B] = P[rr] + P[ra] = 0.02. \tag{1.11}$$

The conditional probability of the second chip being rejected given that the first chip is rejected is, by definition, the ratio of $P[AB]$ to $P[B]$, where, in this example,

$$P[AB] = P[\text{both rejected}] = P[rr] = 0.01 \tag{1.12}$$

Thus

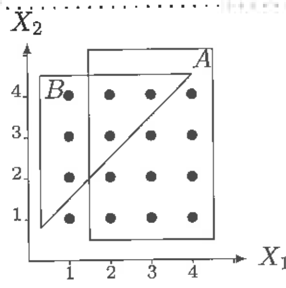
$$P[A|B] = \frac{P[AB]}{P[B]} = 0.01/0.02 = 0.5. \tag{1.13}$$

The information that the first chip is a reject drastically changes our state of knowledge about the second chip. We started with near certainty, $P[A] = 0.02$, that the second chip would not fail and ended with complete uncertainty about the quality of the second chip, $P[A|B] = 0.5$.

Example 1.11

Roll two fair four-sided dice. Let X_1 and X_2 denote the number of dots that appear on die 1 and die 2, respectively. Let A be the event $X_1 \geq 2$. What is $P[A]$? Let B

denote the event $X_2 > X_1$. What is $P[B]$? What is $P[A|B]$?



We begin by observing that the sample space has 16 elements corresponding to the four possible values of X_1 and the same four values of X_2. Since the dice are fair, the outcomes are equally likely, each with probability 1/16. We draw the sample space as a set of black circles in a two-dimensional diagram, in which the axes represent the events X_1 and X_2. Each outcome is a pair of values (X_1, X_2). The rectangle represents A. It contains 12 outcomes, each with probability 1/16.

To find $P[A]$, we add up the probabilities of outcomes in A, so $P[A] = 12/16 = 3/4$. The triangle represents B. It contains six outcomes. Therefore $P[B] = 6/16 = 3/8$. The event AB has three outcomes, $(2,3), (2,4), (3,4)$, so $P[AB] = 3/16$. From the definition of conditional probability, we write

$$P[A|B] = \frac{P[AB]}{P[B]} = \frac{1}{2}. \tag{1.14}$$

We can also derive this fact from the diagram by restricting our attention to the six outcomes in B (the conditioning event) and noting that three of the six outcomes in B (one-half of the total) are also in A.

Quiz 1.3

Monitor three consecutive packets going through an Internet router. Classify each one as either video (v) or data (d). Your observation is a sequence of three letters (each one is either v or d). For example, three video packets corresponds to vvv. The outcomes vvv and ddd each have probability 0.2 whereas each of the other outcomes vvd, vdv, vdd, dvv, dvd, and ddv has probability 0.1. Count the number of video packets N_V in the three packets you have observed. Describe in words and also calculate the following probabilities:

(a) $P[N_V = 2]$ (b) $P[N_V \geq 1]$

(c) $P[\{vvd\}|N_V = 2]$ (d) $P[\{ddv\}|N_V = 2]$

(e) $P[N_V = 2|N_V \geq 1]$ (f) $P[N_V \geq 1|N_V = 2]$

1.4 Partitions and the Law of Total Probability

A partition divides the sample space into mutually exclusive sets. The law of total probability expresses the probability of an event as the sum of the probabilities of outcomes that are in the separate sets of a partition.

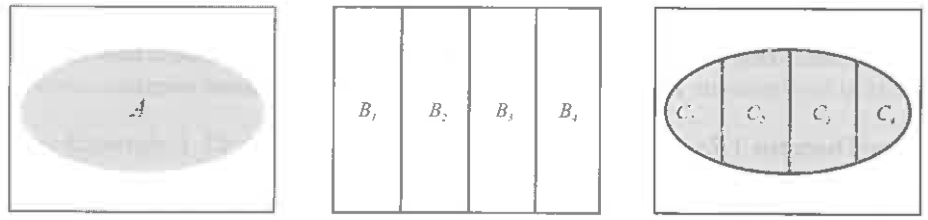


Figure 1.1 In this example of Theorem 1.7, the partition is $B = \{B_1, B_2, B_3, B_4\}$ and $C_i = A \cap B_i$ for $i = 1, \ldots, 4$. It should be apparent that $A = C_1 \cup C_2 \cup C_3 \cup C_4$.

Example 1.12

Flip four coins, a penny, a nickel, a dime, and a quarter. Examine the coins in order (penny, then nickel, then dime, then quarter) and observe whether each coin shows a head (h) or a tail (t). What is the sample space? How many elements are in the sample space?

The sample space consists of 16 four-letter words, with each letter either h or t. For example, the outcome $tthh$ refers to the penny and the nickel showing tails and the dime and quarter showing heads. There are 16 members of the sample space.

Example 1.13

Continuing Example 1.12, let $B_i = \{\text{outcomes with } i \text{ heads}\}$. Each B_i is an event containing one or more outcomes. For example, $B_1 = \{ttth, ttht, thtt, httt\}$ contains four outcomes. The set $B = \{B_0, B_1, B_2, B_3, B_4\}$ is a partition. Its members are mutually exclusive and collectively exhaustive. It is not a sample space because it lacks the finest-grain property. Learning that an experiment produces an event B_1 tells you that one coin came up heads, but it doesn't tell you which coin it was.

The experiment in Example 1.12 and Example 1.13 refers to a "toy problem," one that is easily visualized but isn't something we would do in the course of our professional work. Mathematically, however, it is equivalent to many real engineering problems. For example, observe a pair of modems transmitting four bits from one computer to another. For each bit, observe whether the receiving modem detects the bit correctly (c) or makes an error (e). Or test four integrated circuits. For each one, observe whether the circuit is acceptable (a) or a reject (r).

In all of these examples, the sample space contains 16 four-letter words formed with an alphabet containing two letters. If we are interested only in the number of times one of the letters occurs, it is sufficient to refer only to the partition B, which does not contain all of the information about the experiment but does contain all of the information we need. The partition is simpler to deal with than the sample space because it has fewer members (there are five events in the partition and 16 outcomes in the sample space). The simplification is much more significant when the complexity of the experiment is higher. For example, in testing 20 circuits the sample space has $2^{20} = 1{,}048{,}576$ members, while the corresponding partition has only 21 members.

We observed in Section 1.2 that the entire theory of probability is based on a union of mutually exclusive events. The following theorem shows how to use a partition to represent an event as a union of mutually exclusive events.

Theorem 1.7

For a partition $B = \{B_1, B_2, \ldots\}$ and any event A in the sample space, let $C_i = A \cap B_i$. For $i \neq j$, the events C_i and C_j are mutually exclusive and

$$A = C_1 \cup C_2 \cup \cdots.$$

Figure 1.1 is a picture of Theorem 1.7.

Example 1.14

In the coin-tossing experiment of Example 1.12, let A equal the set of outcomes with less than three heads:

$$A = \{tttt, httt, thtt, ttht, ttth, hhtt, htht, htth, tthh, thth, thht\}. \tag{1.15}$$

From Example 1.13, let $B_i = \{\text{outcomes with } i \text{ heads}\}$. Since $\{B_0, \ldots, B_4\}$ is a partition, Theorem 1.7 states that

$$A = (A \cap B_0) \cup (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup (A \cap B_4) \tag{1.16}$$

In this example, $B_i \subset A$, for $i = 0, 1, 2$. Therefore $A \cap B_i = B_i$ for $i = 0, 1, 2$. Also, for $i = 3$ and $i = 4$, $A \cap B_i = \varnothing$ so that $A = B_0 \cup B_1 \cup B_2$, a union of mutually exclusive sets. In words, this example states that the event "less than three heads" is the union of events "zero heads," "one head," and "two heads."

We advise you to make sure you understand Theorem 1.7 and Example 1.14. Many practical problems use the mathematical technique contained in the theorem. For example, find the probability that there are three or more bad circuits in a batch that comes from a fabrication machine.

The following theorem refers to a partition $\{B_1, B_2, \ldots, B_m\}$ and any event, A. It states that we can find the probability of A by adding the probabilities of the parts of A that are in the separate components of the event space.

Theorem 1.8

For any event A, and partition $\{B_1, B_2, \ldots, B_m\}$,

$$\mathrm{P}[A] = \sum_{i=1}^{m} \mathrm{P}[A \cap B_i].$$

Proof The proof follows directly from Theorem 1.7 and Theorem 1.2. In this case, the mutually exclusive sets are $C_i = \{A \cap B_i\}$.

Theorem 1.8 is often used when the sample space can be written in the form of a table. In this table, the rows and columns each represent a partition. This method is shown in the following example.

Example 1.15

A company has a model of email use. It classifies all emails as either long (l), if they are over 10 MB in size, or brief (b). It also observes whether the email is just text (t), has attached images (i), or has an attached video (v). This model implies an experiment in which the procedure is to monitor an email and the observation consists of the type of email, t, i, or v, and the length, l or b. The sample space has six outcomes: $S = \{lt, bt, li, bi, lv, bv\}$. In this problem, each email is classifed in two ways: by length and by type. Using L for the event that an email is long and B for the event that a email is brief, $\{L, B\}$ is a partition. Similarly, the text (T), image (I), and video (V) classification is a partition $\{T, I, V\}$. The sample space can be represented by a table in which the rows and columns are labeled by events and the intersection of each row and column event contains a single outcome. The corresponding table entry is the probability of that outcome. In this case, the table is

	T	I	V
L	0.3	0.12	0.15
B	0.2	0.08	0.15

(1.17)

For example, from the table we can read that the probability of a brief image email is $\mathrm{P}[bi] = \mathrm{P}[BI] = 0.08$. Note that $\{T, I, V\}$ is a partition corresponding to $\{B_1, B_2, B_3\}$ in Theorem 1.8. Thus we can apply Theorem 1.8 to find the probability of a long email:

$$\mathrm{P}[L] = \mathrm{P}[LT] + \mathrm{P}[LI] + \mathrm{P}[LV] = 0.57. \tag{1.18}$$

Law of Total Probability

In many applications, we begin with information about conditional probabilities and use the law of total probability to calculate unconditional probabilities.

Theorem 1.9 Law of Total Probability

For a partition $\{B_1, B_2, \ldots, B_m\}$ with $\mathrm{P}[B_i] > 0$ for all i,

$$\mathrm{P}[A] = \sum_{i=1}^{m} \mathrm{P}[A|B_i]\,\mathrm{P}[B_i].$$

Proof This follows from Theorem 1.8 and the identity $\mathrm{P}[AB_i] = \mathrm{P}[A|B_i]\,\mathrm{P}[B_i]$, which is a direct consequence of the definition of conditional probability.

The usefulness of the result can be seen in the next example.

Example 1.16

A company has three machines B_1, B_2, and B_3 making 1 kΩ resistors. Resistors within 50 Ω of the nominal value are considered acceptable. It has been observed that 80% of the resistors produced by B_1 and 90% of the resistors produced by B_2 are acceptable. The percentage for machine B_3 is 60%. Each hour, machine B_1 produces 3000 resistors, B_2 produces 4000 resistors, and B_3 produces 3000 resistors. All of the resistors are mixed together at random in one bin and packed for shipment. What is the probability that the company ships an acceptable resistor?

Let $A = \{\text{resistor is acceptable}\}$. Using the resistor accuracy information to formulate a probability model, we write

$$P[A|B_1] = 0.8, \quad P[A|B_2] = 0.9, \quad P[A|B_3] = 0.6. \tag{1.19}$$

The production figures state that $3000 + 4000 + 3000 = 10{,}000$ resistors per hour are produced. The fraction from machine B_1 is $P[B_1] = 3000/10{,}000 = 0.3$. Similarly, $P[B_2] = 0.4$ and $P[B_3] = 0.3$. Now it is a simple matter to apply the law of total probability to find the acceptable probability for all resistors shipped by the company:

$$\begin{aligned} P[A] &= P[A|B_1]P[B_1] + P[A|B_2]P[B_2] + P[A|B_3]P[B_3] \\ &= (0.8)(0.3) + (0.9)(0.4) + (0.6)(0.3) = 0.78. \end{aligned} \tag{1.20}$$

For the whole factory, 78% of resistors are within 50 Ω of the nominal value.

Bayes' Theorem

When we have advance information about $P[A|B]$ and need to calculate $P[B|A]$, we refer to the following formula:

Theorem 1.10 **Bayes' theorem**

$$P[B|A] = \frac{P[A|B]P[B]}{P[A]}.$$

Proof

$$P[B|A] = \frac{P[AB]}{P[A]} = \frac{P[A|B]P[B]}{P[A]}. \tag{1.21}$$

Bayes' theorem is a simple consequence of the definition of conditional probability. It has a name because it is extremely useful for making inferences about phenomena that cannot be observed directly. Sometimes these inferences are described as "reasoning about causes when we observe effects." For example, let $\{B_1, \ldots, B_m\}$ be a partition that includes all possible states of something that interests us but that we cannot observe directly (for example, the machine that made a particular resistor). For each possible state, B_i, we know the prior probability $P[B_i]$ and $P[A|B_i]$, the probability that an event A occurs (the resistor meets a quality criterion) if B_i is the actual state. Now we observe the actual event (either the resistor passes or fails a test), and we ask about the thing we are interested in (the machines that might have produced the resistor). That is, we use Bayes' theorem to find $P[B_1|A], P[B_2|A], \ldots, P[B_m|A]$. In performing the calculations, we use the law of total probability to calculate the denominator in Theorem 1.10. Thus for state B_i,

$$P[B_i|A] = \frac{P[A|B_i]P[B_i]}{\sum_{i=1}^{m} P[A|B_i]P[B_i]}. \tag{1.22}$$

Example 1.17

In Example 1.16 about a shipment of resistors from the factory, we learned that:

- The probability that a resistor is from machine B_3 is $P[B_3] = 0.3$.
- The probability that a resistor is *acceptable* — i.e., within 50 Ω of the nominal value — is $P[A] = 0.78$.
- Given that a resistor is from machine B_3, the conditional probability that it is acceptable is $P[A|B_3] = 0.6$.

What is the probability that an acceptable resistor comes from machine B_3?

Now we are given the event A that a resistor is within 50 Ω of the nominal value, and we need to find $P[B_3|A]$. Using Bayes' theorem, we have

$$P[B_3|A] = \frac{P[A|B_3]P[B_3]}{P[A]}. \tag{1.23}$$

Since all of the quantities we need are given in the problem description, our answer is

$$P[B_3|A] = (0.6)(0.3)/(0.78) = 0.23. \tag{1.24}$$

Similarly we obtain $P[B_1|A] = 0.31$ and $P[B_2|A] = 0.46$. Of all resistors within 50 Ω of the nominal value, only 23% come from machine B_3 (even though this machine produces 30% of all resistors). Machine B_1 produces 31% of the resistors that meet the 50 Ω criterion and machine B_2 produces 46% of them.

Quiz 1.4

Monitor customer behavior in the Phonesmart store. Classify the behavior as buying (B) if a customer purchases a smartphone. Otherwise the behavior is no purchase (N). Classify the time a customer is in the store as long (L) if the customer stays more than three minutes; otherwise classify the amount of time as rapid (R). Based on experience with many customers, we use the probability model $P[N] = 0.7$, $P[L] = 0.6$, $P[NL] = 0.35$. Find the following probabilities:

(a) $P[B \cup L]$ (b) $P[N \cup L]$

(c) $P[N \cup B]$ (d) $P[LR]$

1.5 Independence

Two events are independent if observing one event does not change the probability of observing the other event.

Definition 1.6 Two Independent Events

Events A and B are ***independent*** *if and only if*

$$P[AB] = P[A]P[B].$$

When events A and B have nonzero probabilities, the following formulas are equivalent to the definition of independent events:

$$P[A|B] = P[A], \qquad P[B|A] = P[B]. \tag{1.25}$$

To interpret independence, consider probability as a description of our knowledge of the result of the experiment. $P[A]$ describes our prior knowledge (before the experiment is performed) that the outcome is included in event A. The fact that the outcome is in B is partial information about the experiment. $P[A|B]$ reflects our knowledge of A when we learn that B occurs. $P[A|B] = P[A]$ states that learning that B occurs does not change our information about A. It is in this sense that the events are independent.

Problem 1.5.11 asks the reader to prove that if A and B are independent, then A and B^c are also independent. The logic behind this conclusion is that if learning that event B occurs does not alter the probability of event A, then learning that B does not occur also should not alter the probability of A.

Keep in mind that **independent and mutually exclusive are *not* synonyms**. In some contexts these words can have similar meanings, but this is not the case in probability. Mutually exclusive events A and B have no outcomes in common and therefore $P[AB] = 0$. In most situations independent events are not mutually exclusive! Exceptions occur only when $P[A] = 0$ or $P[B] = 0$. When we have to calculate probabilities, knowledge that events A and B are *mutually exclusive* is very helpful. Axiom 3 enables us to *add* their probabilities to obtain the probability of the *union*. Knowledge that events C and D are *independent* is also very useful. Definition 1.6 enables us to *multiply* their probabilities to obtain the probability of the *intersection*.

Example 1.18

Suppose that for the experiment monitoring three purchasing decisions in Example 1.7, each outcome (a sequence of three decisions, each either buy or not buy) is equally likely. Are the events B_2 that the second customer purchases a phone and N_2 that the second customer does not purchase a phone independent? Are the events B_1 and B_2 independent?

Each element of the sample space $S = \{bbb, bbn, bnb, bnn, nbb, nbn, nnb, nnn\}$ has probability $1/8$. Each of the events

$$B_2 = \{bbb, bbn, nbb, nbn\} \quad \text{and} \quad N_2 = \{bnb, bnn, nnb, nnn\} \tag{1.26}$$

contains four outcomes, so $P[B_2] = P[N_2] = 4/8$. However, $B_2 \cap N_2 = \varnothing$ and $P[B_2N_2] = 0$. That is, B_2 and N_2 are mutually exclusive because the second customer cannot both purchase a phone and not purchase a phone. Since $P[B_2N_2] \neq P[B_2]P[N_2]$, B_2 and N_2 are not independent. Learning whether or not the event B_2 (second customer buys a phone) occurs drastically affects our knowledge of whether or not the event N_2 (second customer does not buy a phone) occurs. Each of the events $B_1 = \{bnn, bnb, bbn, bbb\}$ and $B_2 = \{bbn, bbb, nbn, nbb\}$ has four outcomes, so $P[B_1] = P[B_2] = 4/8 = 1/2$. In this case, the intersection $B_1 \cap B_2 = \{bbn, bbb\}$ has probability $P[B_1B_2] = 2/8 = 1/4$. Since $P[B_1B_2] = P[B_1]P[B_2]$, events B_1 and B_2 are independent. Learning whether or not the event B_2 (second customer buys a phone) occurs does not affect our knowledge of whether or not the event B_1 (first customer buys a phone) occurs.

In this example we have analyzed a probability model to determine whether two events are independent. In many practical applications we reason in the opposite direction. Our knowledge of an experiment leads us to *assume* that certain pairs of events are independent. We then use this knowledge to build a probability model for the experiment.

Example 1.19

Integrated circuits undergo two tests. A mechanical test determines whether pins have the correct spacing, and an electrical test checks the relationship of outputs to inputs. We *assume* that electrical failures and mechanical failures occur independently. Our information about circuit production tells us that mechanical failures occur with probability 0.05 and electrical failures occur with probability 0.2. What is the probability model of an experiment that consists of testing an integrated circuit and observing the results of the mechanical and electrical tests?

To build the probability model, we note that the sample space contains four outcomes:

$$S = \{(ma, ea), (ma, er), (mr, ea), (mr, er)\} \tag{1.27}$$

where m denotes mechanical, e denotes electrical, a denotes accept, and r denotes reject. Let M and E denote the events that the mechanical and electrical tests are acceptable. Our prior information tells us that $P[M^c] = 0.05$, and $P[E^c] = 0.2$. This implies $P[M] = 0.95$ and $P[E] = 0.8$. Using the independence assumption and

Definition 1.6, we obtain the probabilities of the four outcomes:

$$P[(ma,ea)] = P[ME] = P[M]P[E] = 0.95 \times 0.8 = 0.76, \tag{1.28}$$
$$P[(ma,er)] = P[ME^c] = P[M]P[E^c] = 0.95 \times 0.2 = 0.19, \tag{1.29}$$
$$P[(mr,ea)] = P[M^cE] = P[M^c]P[E] = 0.05 \times 0.8 = 0.04, \tag{1.30}$$
$$P[(mr,er)] = P[M^cE^c] = P[M^c]P[E^c] = 0.05 \times 0.2 = 0.01. \tag{1.31}$$

Thus far, we have considered independence as a property of a pair of events. Often we consider larger sets of independent events. For more than two events to be *independent*, the probability model has to meet a set of conditions. To define mutual independence, we begin with three sets.

Definition 1.7 Three Independent Events

A_1, A_2, and A_3 are ***mutually independent*** *if and only if*

(a) A_1 and A_2 are independent,

(b) A_2 and A_3 are independent,

(c) A_1 and A_3 are independent,

(d) $P[A_1 \cap A_2 \cap A_3] = P[A_1]P[A_2]P[A_3]$.

The final condition is a simple extension of Definition 1.6. The following example shows why this condition is insufficient to guarantee that "everything is independent of everything else," the idea at the heart of independence.

Example 1.20

In an experiment with equiprobable outcomes, the partition is $S = \{1,2,3,4\}$. $P[s] = 1/4$ for all $s \in S$. Are the events $A_1 = \{1,3,4\}$, $A_2 = \{2,3,4\}$, and $A_3 = \emptyset$ mutually independent?

These three sets satisfy the final condition of Definition 1.7 because $A_1 \cap A_2 \cap A_3 = \emptyset$, and

$$P[A_1 \cap A_2 \cap A_3] = P[A_1]P[A_2]P[A_3] = 0. \tag{1.32}$$

However, A_1 and A_2 are not independent because, with all outcomes equiprobable,

$$P[A_1 \cap A_2] = P[\{3,4\}] = 1/2 \neq P[A_1]P[A_2] = 3/4 \times 3/4. \tag{1.33}$$

Hence the three events are not mutually independent.

The definition of an arbitrary number of mutually independent events is an extension of Definition 1.7.

Definition 1.8 More than Two Independent Events

If $n \geq 3$, the events $A_1, A_2, \ldots, A_n$ are mutually independent if and only if

(a) all collections of $n-1$ events chosen from $A_1, A_2, \ldots A_n$ are mutually independent,

(b) $P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1]P[A_2]\cdots P[A_n]$.

This definition and Example 1.20 show us that when $n > 2$ it is a complex matter to determine whether or not n events are mutually independent. On the other hand, if we know that n events are mutually independent, it is a simple matter to determine the probability of the intersection of any subset of the n events. Just multiply the probabilities of the events in the subset.

Quiz 1.5

Monitor two consecutive packets going through a router. Classify each one as video (v) if it was sent from a Youtube server or as ordinary data (d) otherwise. Your observation is a sequence of two letters (either v or d). For example, two video packets corresponds to vv. The two packets are independent and the probability that any one of them is a video packet is 0.8. Denote the identity of packet i by C_i. If packet i is a video packet, then $C_i = v$; otherwise, $C_i = d$. Count the number N_V of video packets in the two packets you have observed. Determine whether the following pairs of events are independent:

(a) $\{N_V = 2\}$ and $\{N_V \geq 1\}$

(b) $\{N_V \geq 1\}$ and $\{C_1 = v\}$

(c) $\{C_2 = v\}$ and $\{C_1 = d\}$

(d) $\{C_2 = v\}$ and $\{N_V \text{ is even}\}$

1.6 MATLAB

The MATLAB programming environment can be used for studying probability models by performing numerical calculations, simulating experiments, and drawing graphs. Simulations make extensive use of the MATLAB random number generator `rand`. In addition to introducing aspects of probability theory, each chapter of this book concludes with a section that uses MATLAB to demonstrate with numerical examples the concepts presented in the chapter. All of the MATLAB programs in this book can be downloaded from the companion website. On the other hand, the MATLAB sections are not essential to understanding the theory. You can use this text to learn probability without using MATLAB.

Engineers studied and applied probability theory long before the invention of MATLAB. Nevertheless, MATLAB provides a convenient programming environment for solving probability problems and for building models of probabilistic systems. Versions of MATLAB, including a low-cost student edition, are available for most computer systems.

At the end of each chapter, we include a MATLAB section (like this one) that introduces ways that MATLAB can be applied to the concepts and problems of the chapter. We assume you already have some familiarity with the basics of running MATLAB. If you do not, we encourage you to investigate the built-in tutorial, books dedicated to MATLAB, and various Web resources.

MATLAB can be used two ways to study and apply probability theory. Like a sophisticated scientific calculator, it can perform complex numerical calculations and draw graphs. It can also simulate experiments with random outcomes. To simulate experiments, we need a source of randomness. MATLAB uses a computer algorithm, referred to as a *pseudorandom number generator*, to produce a sequence of numbers between 0 and 1. Unless someone knows the algorithm, it is impossible to examine some of the numbers in the sequence and thereby calculate others. The calculation of each random number is similar to an experiment in which all outcomes are equally likely and the sample space is all binary numbers of a certain length. (The length depends on the machine running MATLAB.) Each number is interpreted as a fraction, with a binary point preceding the bits in the binary number. To use the pseudorandom number generator to simulate an experiment that contains an event with probability p, we examine one number, r, produced by the MATLAB algorithm and say that the event occurs if $r < p$; otherwise it does not occur.

A MATLAB simulation of an experiment starts with `rand`: the random number generator `rand(m,n)` returns an $m \times n$ array of pseudorandom numbers. Similarly, `rand(n)` produces an $n \times n$ array and `rand(1)` is just a scalar random number. Each number produced by `rand(1)` is in the interval $(0,1)$. Each time we use `rand`, we get new, seemingly unpredictable numbers. Suppose p is a number between 0 and 1. The comparison `rand(1) < p` produces a 1 if the random number is less than p; otherwise it produces a zero. Roughly speaking, the function `rand(1) < p` simulates a coin flip with $P[\text{tail}] = p$.

Example 1.21

```
>> X=rand(1,4)
X =
  0.0879  0.9626  0.6627  0.2023
>> X<0.5
ans =
     1     0     0     1
```

Since `rand(1,4) < 0.5` compares four random numbers against 0.5, the result is a random sequence of zeros and ones that simulates a sequence of four flips of a fair coin. We associate the outcome 1 with {head} and 0 with {tail}.

MATLAB also has some convenient variations on `rand`. For example, `randi(k)` generates a random integer from the set $\{1, 2, \ldots, k\}$ and `randi(k,m,n)` generates an $m \times n$ array of such random integers.

Example 1.22

Use MATLAB to generate 12 random student test scores T as described in Quiz 1.2.

Since `randi(50,1,12)` generates 12 test scores from the set $\{1, \ldots, 50\}$, we need only to add 50 to each score to obtain test scores in the range $\{51, \ldots, 100\}$.

```
>> 50+randi(50,1,12)
ans =
   69   78   60   68   93   99   77   95   88   57   51   90
```

Finally, we note that MATLAB's random numbers are only seemingly unpredictable. In fact, MATLAB maintains a seed value that determines the subsequent "random" numbers that will be returned. This seed is controlled by the `rng` function; `s=rng` saves the current seed and `rng(s)` restores a previously saved seed. Initializing the random number generator with the same seed always generates the same sequence:

Example 1.23

```
>> s=rng;
>> 50+randi(50,1,12)
ans =
   89   76   80   80   72   92   58   56   77   78   59   58
>> rng(s);
>> 50+randi(50,1,12)
ans =
   89   76   80   80   72   92   58   56   77   78   59   58
```

When you run a simulation that uses `rand`, it normally doesn't matter how the `rng` seed is initialized. However, it can be instructive to use the same repeatable sequence of `rand` values when you are debugging your simulation.

Quiz 1.6

The number of characters in a tweet is equally likely to be any integer between 1 and 140. Simulate an experiment that generates 1000 tweets and counts the number of "long" tweets that have over 120 characters. Repeat this experiment 5 times.

Problems

Difficulty: • Easy ■ Moderate ◆ Difficult ◆◆ Experts Only

1.1.1• Ricardo's offers customers two kinds of pizza crust, Roman (R) and Neapolitan (N). All pizzas have cheese but not all pizzas have tomato sauce. Roman pizzas can have tomato sauce or they can be white (W); Neapolitan pizzas always have tomato sauce. It is possible to order a Roman pizza with mushrooms (M) added. A Neapolitan pizza can contain mushrooms or onions (O) or both, in addition to the tomato sauce and cheese. Draw a Venn diagram that shows the relationship among the ingredients N, M, O, T, and W in the menu of Ricardo's pizzeria.

1.1.2 A hypothetical wi-fi transmission can take place at any of three speeds depending on the condition of the radio channel between a laptop and an access point. The speeds are high (h) at 54 Mb/s, medium (m) at 11 Mb/s, and low (l) at 1 Mb/s. A user of the wi-fi connection can transmit a short signal corresponding to a mouse click (c), or a long signal corresponding to a tweet (t). Consider the experiment of monitoring wi-fi signals and observing the transmission speed and the length. An observation is a two-letter word, for example, a high-speed, mouse-click transmission is hm.

(a) What is the sample space?

(b) A_1 is the event "medium speed connection." What outcomes are in A_1?

(c) A_2 is the event "mouse click." What outcomes are in A_2?

(d) Let A_3 be the event "high speed connection or low speed connection." What are the outcomes in A_3?

(e) Are A_1, A_2, A_3 mutually exclusive?

(f) Are A_1, A_2, A_3 collectively exhaustive?

1.1.3 An integrated circuit factory has three machines X, Y, and Z. Test one integrated circuit produced by each machine. Either a circuit is acceptable (a) or it fails (f). An observation is a sequence of three test results corresponding to the circuits from machines X, Y, and Z, respectively. For example, aaf is the observation that the circuits from X and Y pass the test and the circuit from Z fails the test.

(a) What are the elements of the sample space of this experiment?

(b) What are the elements of the sets

$$Z_F = \{\text{circuit from } Z \text{ fails}\},$$
$$X_A = \{\text{circuit from } X \text{ is acceptable}\}.$$

(c) Are Z_F and X_A mutually exclusive?

(d) Are Z_F and X_A collectively exhaustive?

(e) What are the elements of the sets

$$C = \{\text{more than one circuit acceptable}\},$$
$$D = \{\text{at least two circuits fail}\}.$$

(f) Are C and D mutually exclusive?

(g) Are C and D collectively exhaustive?

1.1.4 Shuffle a deck of cards and turn over the first card. What is the sample space of this experiment? How many outcomes are in the event that the first card is a heart?

1.1.5 Find out the birthday (month and day but not year) of a randomly chosen person. What is the sample space of the experiment? How many outcomes are in the event that the person is born in July?

1.1.6 The sample space of an experiment consists of all undergraduates at a university. Give four examples of partitions.

1.1.7 The sample space of an experiment consists of the measured resistances of two resistors. Give four examples of partitions.

1.2.1 Find $P[B]$ in each case:

(a) Events A and B are a partition and $P[A] = 3P[B]$.

(b) For events A and B, $P[A \cup B] = P[A]$ and $P[A \cap B] = 0$.

(c) For events A and B, $P[A \cup B] = P[A] - P[B]$.

1.2.2 You roll two fair six-sided dice; one die is red, the other is white. Let R_i be the event that the red die rolls i. Let W_j be the event that the white die rolls j.

(a) What is $P[R_3W_2]$?

(b) What is the $P[S_5]$ that the sum of the two rolls is 5?

1.2.3 You roll two fair six-sided dice. Find the probability $P[D_3]$ that the absolute value of the difference of the dice is 3.

1.2.4 Indicate whether each statement is true or false.

(a) If $P[A] = 2P[A^c]$, then $P[A] = 1/2$.

(b) For all A and B, $P[AB] \leq P[A]P[B]$.

(c) If $P[A] < P[B]$, then $P[AB] < P[B]$.

(d) If $P[A \cap B] = P[A]$, then $P[A] \geq P[B]$.

1.2.5 Computer programs are classified by the length of the source code and by the execution time. Programs with more than 150 lines in the source code are big (B). Programs with $\leq$ 150 lines are little (L). Fast programs (F) run in less than 0.1 seconds. Slow programs (W) require at least 0.1 seconds. Monitor a program executed by a computer. Observe the length of the source code and the run time. The probability model for this experiment contains the following information: $P[LF] = 0.5$, $P[BF] = 0.2$, and $P[BW] = 0.2$. What is the sample space of the experiment? Calculate the following probabilities: $P[W]$, $P[B]$, and $P[W \cup B]$.

1.2.6 There are two types of cellular phones, handheld phones (H) that you carry and mobile phones (M) that are mounted in vehicles. Phone calls can be classified by the traveling speed of the user as fast (F) or slow (W). Monitor a cellular phone call and observe the type of telephone and the speed of the user. The probability model for this experiment has the following information: $P[F] = 0.5$, $P[HF] = 0.2$, $P[MW] = 0.1$. What is the sample space of the experiment? Find the following probabilities $P[W]$, $P[MF]$, and $P[H]$.

1.2.7 Shuffle a deck of cards and turn over the first card. What is the probability that the first card is a heart?

1.2.8 You have a six-sided die that you roll once and observe the number of dots facing upwards. What is the sample space? What is the probability of each sample outcome? What is the probability of E, the event that the roll is even?

1.2.9 A student's score on a 10-point quiz is equally likely to be any integer between 0 and 10. What is the probability of an A, which requires the student to get a score of 9 or more? What is the probability the student gets an F by getting less than 4?

1.2.10 Use Theorem 1.3 to prove the following facts:

(a) $P[A \cup B] \geq P[A]$

(b) $P[A \cup B] \geq P[B]$

(c) $P[A \cap B] \leq P[A]$

(d) $P[A \cap B] \leq P[B]$

1.2.11 Use Theorem 1.3 to prove by induction the *union bound*: For any collection of events $A_1, \ldots, A_n$,

$$P[A_1 \cup A_2 \cup \cdots \cup A_n] \leq \sum_{i=1}^{n} P[A_i].$$

1.2.12 Using *only* the three axioms of probability, prove $P[\varnothing] = 0$.

1.2.13 Using the three axioms of probability and the fact that $P[\varnothing] = 0$, prove Theorem 1.2. Hint: Define $A_i = B_i$ for $i = 1, \ldots, m$ and $A_i = \varnothing$ for $i > m$.

1.2.14 For each fact stated in Theorem 1.3, determine which of the three axioms of probability are needed to prove the fact.

1.3.1 Mobile telephones perform *handoffs* as they move from cell to cell. During a call, a telephone either performs zero handoffs (H_0), one handoff (H_1), or more than one handoff (H_2). In addition, each call is either long (L), if it lasts more than three minutes, or brief (B). The following table describes the probabilities of the possible types of calls.

	H_0	H_1	H_2
L	0.1	0.1	0.2
B	0.4	0.1	0.1

(a) What is the probability that a brief call will have no handoffs?

(b) What is the probability that a call with one handoff will be long?

(c) What is the probability that a long call will have one or more handoffs?

1.3.2 You have a six-sided die that you roll once. Let R_i denote the event that the roll is i. Let G_j denote the event that the roll is greater than j. Let E denote the event that the roll of the die is even-numbered.

(a) What is $P[R_3|G_1]$, the conditional probability that 3 is rolled given that the roll is greater than 1?

(b) What is the conditional probability that 6 is rolled given that the roll is greater than 3?

(c) What is $P[G_3|E]$, the conditional probability that the roll is greater than 3 given that the roll is even?

(d) Given that the roll is greater than 3, what is the conditional probability that the roll is even?

1.3.3 You have a shuffled deck of three cards: 2, 3, and 4. You draw one card. Let C_i denote the event that card i is picked. Let E denote the event that the card chosen is a even-numbered card.

(a) What is $P[C_2|E]$, the probability that the 2 is picked given that an even-numbered card is chosen?

(b) What is the conditional probability that an even-numbered card is picked given that the 2 is picked?

1.3.4 Phonesmart is having a sale on Bananas. If you buy one Banana at full price, you get a second at half price. When couples come in to buy a pair of phones, sales of Apricots and Bananas are equally likely. Moreover, given that the first phone sold is a Banana, the second phone is twice as likely to be a Banana rather than an Apricot. What is the probability that a couple buys a pair of Bananas?

1.3.5 The basic rules of genetics were discovered in mid-1800s by Mendel, who found that each characteristic of a pea plant, such as whether the seeds were green or yellow, is determined by two genes, one from each parent. In his pea plants, Mendel found that yellow seeds were a dominant trait over green seeds. A yy pea with two yellow genes has yellow seeds; a gg pea with two recessive genes has green seeds; a hybrid gy or yg pea has yellow seeds. In one of Mendel's experiments, he started with a parental generation in which half the pea plants were yy and half the plants were gg. The two groups were crossbred so that each pea plant in the first generation was gy. In the second generation, each pea plant was equally likely to inherit a y or a g gene from each first-generation parent. What is the probability $P[Y]$ that a randomly chosen pea plant in the second generation has yellow seeds?

1.3.6 From Problem 1.3.5, what is the conditional probability of yy, that a pea plant has two dominant genes given the event Y that it has yellow seeds?

1.3.7 You have a shuffled deck of three cards: 2, 3, and 4, and you deal out the three cards. Let E_i denote the event that ith card dealt is even numbered.

(a) What is $P[E_2|E_1]$, the probability the second card is even given that the first card is even?

(b) What is the conditional probability that the first two cards are even given that the third card is even?

(c) Let O_i represent the event that the ith card dealt is odd numbered. What is $P[E_2|O_1]$, the conditional probability that the second card is even given that the first card is odd?

(d) What is the conditional probability that the second card is odd given that the first card is odd?

1.3.8 Deer ticks can carry both Lyme disease and human granulocytic ehrlichiosis (HGE). In a study of ticks in the Midwest, it was found that 16% carried Lyme disease, 10% had HGE, and that 10% of the ticks that had either Lyme disease or HGE carried both diseases.

(a) What is the probability $P[LH]$ that a tick carries both Lyme disease (L) and HGE (H)?

(b) What is the conditional probability that a tick has HGE given that it has Lyme disease?

1.4.1 Given the model of handoffs and call lengths in Problem 1.3.1,

(a) What is the probability $P[H_0]$ that a phone makes no handoffs?

(b) What is the probability a call is brief?

(c) What is the probability a call is long or there are at least two handoffs?

1.4.2 For the telephone usage model of Example 1.15, let B_m denote the event that a call is billed for m minutes. To generate a phone bill, observe the duration of the call in integer minutes (rounding up). Charge for M minutes $M = 1, 2, 3, \ldots$ if the exact duration T is $M - 1 < t \leq M$. A more complete probability model shows that for $m = 1, 2, \ldots$ the probability of each event B_m is

$$P[B_m] = \alpha(1-\alpha)^{m-1}$$

where $\alpha = 1 - (0.57)^{1/3} = 0.171$.

(a) Classify a call as long, L, if the call lasts more than three minutes. What is $P[L]$?

(b) What is the probability that a call will be billed for nine minutes or less?

1.4.3 Suppose a cellular telephone is equally likely to make zero handoffs (H_0), one handoff (H_1), or more than one handoff (H_2). Also, a caller is either on foot (F) with probability 5/12 or in a vehicle (V).

(a) Given the preceding information, find three ways to fill in the following probability table:

	H_0	H_1	H_2
F			
V			

(b) Suppose we also learn that 1/4 of all callers are on foot making calls with no handoffs and that 1/6 of all callers are vehicle users making calls with a single handoff. Given these additional facts, find all possible ways to fill in the table of probabilities.

1.5.1 Is it possible for A and B to be independent events yet satisfy $A = B$?

1.5.2 Events A and B are equiprobable, mutually exclusive, and independent. What is $P[A]$?

1.5.3 At a Phonesmart store, each phone sold is twice as likely to be an Apricot as a Banana. Also each phone sale is independent of any other phone sale. If you monitor the sale of two phones, what is the probability that the two phones sold are the same?

1.5.4 Use a Venn diagram in which the event areas are proportional to their probabilities to illustrate two events A and B that are independent.

1.5.5 In an experiment, A and B are mutually exclusive events with probabilities $P[A] = 1/4$ and $P[B] = 1/8$.

(a) Find $P[A \cap B]$, $P[A \cup B]$, $P[A \cap B^c]$, and $P[A \cup B^c]$.

(b) Are A and B independent?

1.5.6 In an experiment, C and D are independent events with probabilities $P[C] = 5/8$ and $P[D] = 3/8$.

(a) Determine the probabilities $P[C \cap D]$, $P[C \cap D^c]$, and $P[C^c \cap D^c]$.

(b) Are C^c and D^c independent?

1.5.7 In an experiment, A and B are mutually exclusive events with probabilities $P[A \cup B] = 5/8$ and $P[A] = 3/8$.

(a) Find $P[B]$, $P[A \cap B^c]$, and $P[A \cup B^c]$.

(b) Are A and B independent?

1.5.8 In an experiment, C, and D are independent events with probabilities $P[C \cap D] = 1/3$, and $P[C] = 1/2$.

(a) Find $P[D]$, $P[C \cap D^c]$, and $P[C^c \cap D^c]$.

(b) Find $P[C \cup D]$ and $P[C \cup D^c]$.

(c) Are C and D^c independent?

1.5.9 In an experiment with equiprobable outcomes, the sample space is $S = \{1, 2, 3, 4\}$ and $P[s] = 1/4$ for all $s \in S$. Find three events in S that are pairwise independent but are not independent. (Note: Pairwise independent events meet the first three conditions of Definition 1.7).

1.5.10 (Continuation of Problem 1.3.5) One of Mendel's most significant results was the conclusion that genes determining different characteristics are transmitted independently. In pea plants, Mendel found that round peas (r) are a dominant trait over wrinkled peas (w). Mendel crossbred a group of (rr, yy) peas with a group of (ww, gg) peas. In this notation, rr denotes a pea with two "round" genes and ww denotes a pea with two "wrinkled" genes. The first generation were either (rw, yg), (rw, gy), (wr, yg), or (wr, gy) plants with both hybrid shape and hybrid color. Breeding among the first generation yielded second-generation plants in which genes for each characteristic were equally likely to be either dominant or recessive. What is the probability $P[Y]$ that a second-generation pea plant has yellow seeds? What is the probability $P[R]$ that a second-generation plant has round peas? Are R and Y independent events? How many visibly different kinds of pea plants would Mendel observe in the second generation? What are the probabilities of each of these kinds?

1.5.11 For independent events A and B, prove that

(a) A and B^c are independent.

(b) A^c and B are independent.

(c) A^c and B^c are independent.

1.5.12 Use a Venn diagram in which the event areas are proportional to their probabilities to illustrate three events A, B, and C that are independent.

1.5.13 Use a Venn diagram in which event areas are in proportion to their probabilities to illustrate events A, B, and C that are pairwise independent but not independent.

1.6.1 Following Quiz 1.2, use MATLAB, but not the `randi` function, to generate a vector $\mathbf{T}$ of 200 independent test scores such that all scores between 51 and 100 are equally likely.

2 Sequential Experiments

Many applications of probability refer to sequential experiments in which the procedure consists of many actions performed in sequence, with an observation taken after each action. Each action in the procedure together with the outcome associated with it can be viewed as a separate experiment with its own probability model. In analyzing sequential experiments we refer to the separate experiments in the sequence as *subexperiments*.

2.1 Tree Diagrams

Tree diagrams display the outcomes of the subexperiments in a sequential experiment. The labels of the branches are probabilities and conditional probabilities. The probability of an outcome of the entire experiment is the product of the probabilities of branches going from the root of the tree to a leaf.

Many experiments consist of a sequence of *subexperiments*. The procedure followed for each subexperiment may depend on the results of the previous subexperiments. We often find it useful to use a type of graph referred to as a *tree diagram* to represent the sequence of subexperiments. To do so, we assemble the outcomes of each subexperiment into sets in a partition. Starting at the root of the tree,[1] we represent each event in the partition of the first subexperiment as a branch and we label the branch with the probability of the event. Each branch leads to a node. The events in the partition of the second subexperiment appear as branches growing from every node at the end of the first subexperiment. The labels of the branches

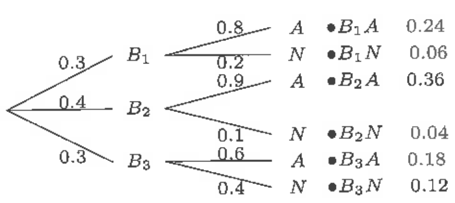
[1]Unlike biological trees, which grow from the ground up, probabilities usually grow from left to right. Some of them have their roots on top and leaves on the bottom.

of the second subexperiment are the *conditional* probabilities of the events in the second subexperiment. We continue the procedure taking the remaining subexperiments in order. The nodes at the end of the final subexperiment are the leaves of the tree. Each leaf corresponds to an outcome of the entire sequential experiment. The probability of each outcome is the product of the probabilities and conditional probabilities on the path from the root to the leaf. We usually label each leaf with a name for the event and the probability of the event.

This is a complicated description of a simple procedure as we see in the following five examples.

Example 2.1

For the resistors of Example 1.16, we used A to denote the event that a randomly chosen resistor is "within 50 Ω of the nominal value." This could mean "acceptable." We use the notation N ("not acceptable") for the complement of A. The experiment of testing a resistor can be viewed as a two-step procedure. First we identify which machine (B_1, B_2, or B_3) produced the resistor. Second, we find out if the resistor is acceptable. Draw a tree for this sequential experiment. What is the probability of choosing a resistor from machine B_2 that is not acceptable?

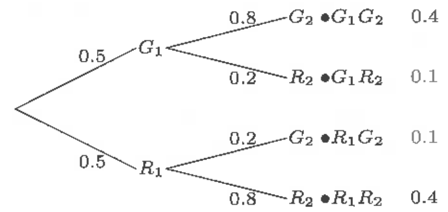


This two-step procedure is shown in the tree on the left. To use the tree to find the probability of the event B_2N, a nonacceptable resistor from machine B_2, we start at the left and find that the probability of reaching B_2 is $P[B_2] = 0.4$. We then move to the right to B_2N and multiply $P[B_2]$ by $P[N|B_2] = 0.1$ to obtain $P[B_2N] = (0.4)(0.1) = 0.04$.

We observe in this example a general property of all tree diagrams that represent sequential experiments. The probabilities on the branches leaving any node add up to 1. This is a consequence of the law of total probability and the property of conditional probabilities that corresponds to Axiom 3 (Theorem 1.6). Moreover, Axiom 2 implies that the probabilities of all of the leaves add up to 1.

Example 2.2

Traffic engineers have coordinated the timing of two traffic lights to encourage a run of green lights. In particular, the timing was designed so that with probability 0.8 a driver will find the second light to have the same color as the first. Assuming the first light is equally likely to be red or green, what is the probability $P[G_2]$ that the second light is green? Also, what is $P[W]$, the probability that you wait for at least one of the first two lights? Lastly, what is $P[G_1|R_2]$, the conditional probability of a green first light given a red second light?



The tree for the two-light experiment is shown on the left. The probability that the second light is green is

$$P[G_2] = P[G_1G_2] + P[R_1G_2] = 0.4 + 0.1 = 0.5. \qquad (2.1)$$

The event W that you wait for at least one light is the event that at least one light is red.

$$W = \{R_1G_2 \cup G_1R_2 \cup R_1R_2\}. \qquad (2.2)$$

The probability that you wait for at least one light is

$$P[W] = P[R_1G_2] + P[G_1R_2] + P[R_1R_2] = 0.1 + 0.1 + 0.4 = 0.6. \qquad (2.3)$$

An alternative way to the same answer is to observe that W is also the complement of the event that both lights are green. Thus,

$$P[W] = P[(G_1G_2)^c] = 1 - P[G_1G_2] = 0.6. \qquad (2.4)$$

To find $P[G_1|R_2]$, we need $P[R_2] = 1 - P[G_2] = 0.5$. Since $P[G_1R_2] = 0.1$, the conditional probability that you have a green first light given a red second light is

$$P[G_1|R_2] = \frac{P[G_1R_2]}{P[R_2]} = \frac{0.1}{0.5} = 0.2. \qquad (2.5)$$

The next example is the "Monty Hall" game, a famous problem with a solution that many regard as counterintuitive. Tree diagrams provide a clear explanation of the answer.

Example 2.3 Monty Hall

In the Monty Hall game, a new car is hidden behind one of three closed doors while a goat is hidden behind each of the other two doors. Your goal is to select the door that hides the car. You make a preliminary selection and then a final selection. The game proceeds as follows:

- You select a door.
- The host, Monty Hall (who knows where the car is hidden), opens one of the two doors you didn't select to reveal a goat.
- Monty then asks you if you would like to switch your selection to the other unopened door.
- After you make your choice (either staying with your original door, or switching doors), Monty reveals the prize behind your chosen door.

To maximize your probability $P[C]$ of winning the car, *is switching to the other door either (a) a good idea, (b) a bad idea or (c) makes no difference?*

To solve this problem, we will consider the "switch" and "do not switch" policies separately. That is, we will construct two different tree diagrams: The first describes what happens if you switch doors while the second describes what happens if you do not switch.

First we describe what is the same no matter what policy you follow. Suppose the doors are numbered 1, 2, and 3. Let H_i denote the event that the car is hidden behind door i. Also, let's assume you first choose door 1. (Whatever door you do choose, that door can be labeled door 1 and it would not change your probability of winning.) Now let R_i denote the event that Monty opens door i that hides a goat. If the car is behind door 1 Monty can choose to open door 2 or door 3 because both hide goats. He chooses door 2 or door 3 by flipping a fair coin. If the car is behind door 2, Monty opens door 3 and if the car is behind door 3, Monty opens door 2. Let C denote the event that you win the car and G the event that you win a goat. After Monty opens one of the doors, you decide whether to change your choice or stay with your choice of door 1. Finally, Monty opens the door of your final choice, either door 1 or the door you switched to.

The tree diagram in Figure 2.1(a) applies to the situation in which you change your choice. From this tree we learn that when the car is behind door 1 (event H_1) and Monty opens door 2 (event R_2), you switch to door 3 and then Monty opens door 3 to reveal a goat (event G). On the other hand, if the car is behind door 2, Monty reveals the goat behind door 3 and you switch to door 2 and win the car. Similarly, if the car is behind door 3, Monty reveals the goat behind door 2 and you switch to door 3 and win the car. For always switch, we see that

$$P[C] = P[H_2R_3C] + P[H_3R_2C] = 2/3. \tag{2.6}$$

If you do not switch, the tree is shown in Figure 2.1(b). In this tree, when the car is behind door 1 (event H_1) and Monty opens door 2 (event R_2), you stay with door 1 and then Monty opens door 1 to reveal the car. On the other hand, if the car is behind door 2, Monty will open door 3 to reveal the goat. Since your final choice was door 1, Monty opens door 1 to reveal the goat. For do not switch,

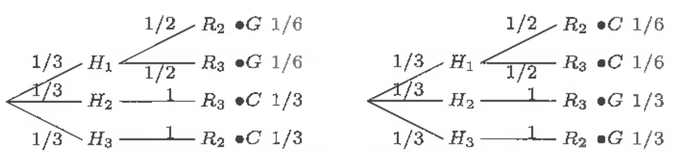
$$P[C] = P[H_1R_2C] + P[H_1R_3C] = 1/3.$$

Thus switching is better; if you don't switch, you win the car only if you initially guessed the location of the car correctly, an event that occurs with probability 1/3. If you switch, you win the car when your initial guess was wrong, an event with probability 2/3.

Note that the two trees look largely the same because the key step where you make a choice is somewhat hidden because it is implied by the first two branches followed in the tree.

Quiz 2.1

In a cellular phone system, a mobile phone must be paged to receive a phone call. However, paging attempts don't always succeed because the mobile phone may not



(a) Switch (b) Do Not Switch

Figure 2.1 Tree Diagrams for Monty Hall

receive the paging signal clearly. Consequently, the system will page a phone up to three times before giving up. If the results of all paging attempts are independent and a single paging attempt succeeds with probability 0.8, sketch a probability tree for this experiment and find the probability $P[F]$ that the phone receives the paging signal clearly.

2.2 Counting Methods

In all applications of probability theory it is important to understand the sample space of an experiment. The methods in this section determine the number of outcomes in the sample space of a sequential experiment.

Understanding the sample space is a key step in formulating and solving a probability problem. To begin, it is often useful to know the number of outcomes in the sample space. This number can be enormous as in the following simple example.

Example 2.4

Choose 7 cards at random from a deck of 52 different cards. Display the cards in the order in which you choose them. How many different sequences of cards are possible?

The procedure consists of seven subexperiments. In each subexperiment, the observation is the identity of one card. The first subexperiment has 52 possible outcomes corresponding to the 52 cards that could be drawn. For each outcome of the first subexperiment, the second subexperiment has 51 possible outcomes corresponding to the 51 remaining cards. Therefore there are 52×51 outcomes of the first two subexperiments. The total number of outcomes of the seven subexperiments is

$$52 \times 51 \times \cdots \times 46 = 674{,}274{,}182{,}400. \tag{2.7}$$

Although many practical experiments are more complicated, the techniques for determining the size of a sample space all follow from the fundamental principle of counting in Theorem 2.1:

Theorem 2.1 Fundamental Principle of Counting

An experiment consists of two subexperiments. If one subexperiment has k outcomes and the other subexperiment has n outcomes, then the experiment has nk outcomes.

Example 2.5

There are two subexperiments. The first subexperiment is "Flip a coin and observe either heads H or tails T." The second subexperiment is "Roll a six-sided die and observe the number of spots." It has six outcomes, $1, 2, \ldots, 6$. The experiment, "Flip a coin and roll a die," has $2 \times 6 = 12$ outcomes:

$$\begin{array}{cccccc} (H,1), & (H,2), & (H,3), & (H,4), & (H,5), & (H,6), \\ (T,1), & (T,2), & (T,3), & (T,4), & (T,5), & (T,6). \end{array}$$

Generally, if an experiment E has k subexperiments $E_1, \ldots, E_k$ where E_i has n_i outcomes, then E has $\prod_{i=1}^{k} n_i$ outcomes.

In Example 2.4, we chose an ordered sequence of seven objects out of a set of 52 *distinguishable objects*. In general, an ordered sequence of k distinguishable objects is called a *k-permutation*. We will use the notation $(n)_k$ to denote the number of possible k-permutations of n distinguishable objects. To find $(n)_k$, suppose we have n distinguishable objects, and the experiment is to choose a sequence of k of these objects. There are n choices for the first object, $n-1$ choices for the second object, etc. Therefore, the total number of possibilities is

$$(n)_k = n(n-1)(n-2)\cdots(n-k+1). \tag{2.8}$$

Multiplying the right side by $(n-k)!/(n-k)!$ yields our next theorem.

Theorem 2.2

The number of k-permutations of n distinguishable objects is

$$(n)_k = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

Sampling without Replacement

Sampling without replacement corresponds to a sequential experiment in which the sample space of each subexperiment depends on the outcomes of previous subexperiments. Choosing objects randomly from a collection is called *sampling*, and the chosen objects are known as a *sample*. A k-permutation is a type of sample obtained by specific rules for selecting objects from the collection. In particular, once we choose an object for a k-permutation, we remove the object from the collection and we cannot choose it again. Consequently, this procedure is called *sampling without replacement*.

Different outcomes in a k-permutation are distinguished by the order in which objects arrive in a sample. By contrast, in many practical problems, we are concerned only with the identity of the objects in a sample, not their order. For example, in many card games, only the set of cards received by a player is of interest. The order in which they arrive is irrelevant.

Example 2.6

Suppose there are four objects, A, B, C, and D, and we define an experiment in which the procedure is to choose two objects without replacement, arrange them in alphabetical order, and observe the result. In this case, to observe AD we could choose A first or D first or both A and D simultaneously. The possible outcomes of the experiment are AB, AC, AD, BC, BD, and CD.

In contrast to this example with six outcomes, the next example shows that the k-permutation corresponding to an experiment in which the observation is the sequence of two letters has $4!/2! = 12$ outcomes.

Example 2.7

Suppose there are four objects, A, B, C, and D, and we define an experiment in which the procedure is to choose two objects without replacement and observe the result. The 12 possible outcomes of the experiment are AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, and DC.

In Example 2.6, each outcome is a subset of the outcomes of a k-permutation. Each subset is called a *k-combination*. We want to find the number of k-combinations.

We use the notation $\binom{n}{k}$ to denote the number of k-combinations. The words for this number are "n choose k," the number of k-combinations of n objects. To find $\binom{n}{k}$, we perform the following two subexperiments to assemble a k-permutation of n distinguishable objects:

1. Choose a k-combination out of the n objects.
2. Choose a k-permutation of the k objects in the k-combination.

Theorem 2.2 tells us that the number of outcomes of the combined experiment is $(n)_k$. The first subexperiment has $\binom{n}{k}$ possible outcomes, the number we have to derive. By Theorem 2.2, the second experiment has $(k)_k = k!$ possible outcomes. Since there are $(n)_k$ possible outcomes of the combined experiment,

$$(n)_k = \binom{n}{k} \cdot k! \tag{2.9}$$

Rearranging the terms yields our next result.

Theorem 2.3

The number of ways to choose k objects out of n distinguishable objects is

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}.$$

We encounter $\binom{n}{k}$ in other mathematical studies. Sometimes it is called a *binomial coefficient* because it appears (as the coefficient of $x^k y^{n-k}$) in the expansion of the binomial $(x+y)^n$. In addition, we observe that

$$\binom{n}{k} = \binom{n}{n-k}. \tag{2.10}$$

The logic behind this identity is that choosing k out of n elements to be part of a subset is equivalent to choosing $n-k$ elements to be excluded from the subset.

In most contexts, $\binom{n}{k}$ is undefined except for integers n and k with $0 \le k \le n$. Here, we adopt the following definition that applies to all nonnegative integers n and all real numbers k:

Definition 2.1 n **choose** k

For an integer $n \ge 0$, we define

$$\binom{n}{k} = \begin{cases} \dfrac{n!}{k!(n-k)!} & k = 0, 1, \ldots, n, \\ 0 & \textit{otherwise.} \end{cases}$$

This definition captures the intuition that given, say, $n = 33$ objects, there are zero ways of choosing $k = -5$ objects, zero ways of choosing $k = 8.7$ objects, and zero ways of choosing $k = 87$ objects. Although this extended definition may seem unnecessary, and perhaps even silly, it will make many formulas in later chapters more concise and easier for students to grasp.

Example 2.8

- The number of combinations of seven cards chosen from a deck of 52 cards is

$$\binom{52}{7} = \frac{52 \times 51 \times \cdots \times 46}{2 \times 3 \times \cdots \times 7} = 133{,}784{,}560, \tag{2.11}$$

  which is the number of 7-combinations of 52 objects. By contrast, we found in Example 2.4 674,274,182,400 7-permutations of 52 objects. (The ratio is $7! = 5040$).
- There are 11 players on a basketball team. The starting lineup consists of five players. There are $\binom{11}{5} = 462$ possible starting lineups.
- There are $\binom{120}{60} \approx 10^{36}$. ways of dividing 120 students enrolled in a probability course into two sections with 60 students in each section.

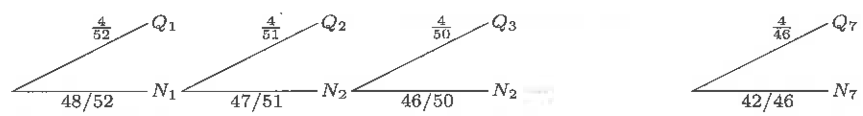
- A baseball team has 15 field players and ten pitchers. Each field player can take any of the eight nonpitching positions. The starting lineup consists of one pitcher and eight field players. Therefore, the number of possible starting lineups is $N = \binom{10}{1}\binom{15}{8} = 64{,}350$. For each choice of starting lineup, the manager must submit to the umpire a batting order for the 9 starters. The number of possible batting orders is $N \times 9! = 23{,}351{,}328{,}000$ since there are N ways to choose the 9 starters, and for each choice of 9 starters, there are $9! = 362{,}880$ possible batting orders.

Example 2.9

There are four queens in a deck of 52 cards. You are given seven cards at random from the deck. What is the probability that you have no queens?

Consider an experiment in which the procedure is to select seven cards at random from a set of 52 cards and the observation is to determine if there are one or more queens in the selection. The sample space contains $H = \binom{52}{7}$ possible combinations of seven cards, each with probability $1/H$. There are $H_{NQ} = \binom{52-4}{7}$ combinations with no queens. The probability of receiving no queens is the ratio of the number of outcomes with no queens to the number of outcomes in the sample space. $H_{NQ}/H = 0.5504$.

Another way of analyzing this experiment is to consider it as a sequence of seven subexperiments. The first subexperiment consists of selecting a card at random and observing whether it is a queen. If it is a queen, an outcome with probability $4/52$ (because there are 52 outcomes in the sample space and four of them are in the event {queen}), stop looking for queens. Otherwise, with probability $48/52$, select another card from the remaining 51 cards and observe whether it is a queen. This outcome of this subexperiment has probability $4/51$. If the second card is not a queen, an outcome with probability $47/51$, continue until you select a queen or you have seven cards with no queen. Using Q_i and N_i to indicate a "Queen" or "No queen" on subexperiment i, the tree for this experiment is



The probability of the event N_7 that no queen is received in your seven cards is the product of the probabilities of the branches leading to N_7:

$$(48/52) \times (47/51) \cdots \times (42/46) = 0.5504. \tag{2.12}$$

Sampling with Replacement

Consider selecting an object from a collection of objects, replacing the selected object, and repeating the process several times, each time replacing the selected object before making another selection. We refer to this situation as *sampling with replacement*. Each selection is the procedure of a subexperiment. The subexperiments are referred to as *independent trials*. In this section we consider the number of possible outcomes that result from sampling with replacement. In the next section we derive probability models for for experiments that specify sampling with replacement.

Example 2.10

There are four queens in a deck of 52 cards. You are given seven cards at random from the deck. After receiving each card you return it to the deck and receive another card at random. Observe whether you have not received any queens among the seven cards you were given. What is the probability that you have received no queens?

The sample space contains 52^7 outcomes. There are 48^7 outcomes with no queens. The ratio is $(48/52)^7 = 0.5710$, the probability of receiving no queens. If this experiment is considered as a sequence of seven subexperiments, the tree looks the same as the tree in Example 2.9, except that all the horizontal branches have probability $48/52$ and all the diagonal branches have probability $4/52$.

The result of Example 2.10 generalizes naturally when we want to choose with replacement a sample of n objects out of a collection of m distinguishable objects. The experiment consists of a sequence of n identical subexperiments with m outcomes in the sample space of each subexperiment. Hence there are m^n ways to choose with replacement a sample of n objects.

Theorem 2.4

Given m distinguishable objects, there are m^n ways to choose with replacement an ordered sample of n objects.

Example 2.11

There are $2^{10} = 1024$ binary sequences of length 10.

Sampling with replacement corresponds to performing n repetitions of an identical subexperiment. Using x_i to denote the outcome of the ith subexperiment, the result for n repetitions of the subexperiment is a sequence $x_1, \ldots, x_n$.

Example 2.12

A chip fabrication facility produces microprocessors. Each microprocessor is tested to determine whether it runs reliably at an acceptable clock speed. A subexperiment to test a microprocessor has sample space $S_{\text{sub}} = \{0, 1\}$ to indicate whether the test was a failure (0) or a success (1). For test i, we record $x_i = 0$ or $x_i = 1$ to indicate the result. In testing four microprocessors, the observation sequence, $x_1x_2x_3x_4$, is one of 16 possible outcomes:

$$S = \left\{ \begin{array}{cccccccc} 0000, & 0001, & 0010, & 0011, & 0100, & 0101, & 0110, & 0111, \\ 1000, & 1001, & 1010, & 1011, & 1100, & 1101, & 1110, & 1111 \end{array} \right\}.$$

Note that we can think of the observation sequence $x_1, \ldots, x_n$ as the result of sampling with replacement n times from a sample space S_{sub}. For sequences of identical subexperiments, we can express Theorem 2.4 as

Theorem 2.5

For n repetitions of a subexperiment with sample space $S_{sub} = \{s_0, \ldots, s_{m-1}\}$, the sample space S of the sequential experiment has m^n outcomes.

Example 2.13

There are ten students in a probability class; each earns a grade $s \in S_{\text{sub}} = \{A, B, C, F\}$. We use the notation x_i to denote the grade of the ith student. For example, the grades for the class could be

$$x_1 x_2 \cdots x_{10} = CBBACFBACF \tag{2.13}$$

The sample space S of possible sequences contains $4^{10} = 1{,}048{,}576$ outcomes.

In Example 2.13, repeating a subexperiment n times and recording the observation consists of constructing a word with n letters. In general, n repetitions of the same subexperiment consists of choosing symbols from the alphabet $\{s_0, \ldots, s_{m-1}\}$. In Example 2.12, $m = 2$ and we have a binary alphabet with symbols $s_0 = 0$ and $s_1 = 1$.

A more challenging problem than finding the number of possible combinations of m objects sampled with replacement from a set of n objects is to calculate the number of observation sequences such that each object appears a specified number of times. We start with the case in which each subexperiment is a trial with sample space $S_{\text{sub}} = \{0, 1\}$ indicating failure or success.

Example 2.14

For five subexperiments with sample space $S_{\text{sub}} = \{0, 1\}$, what is the number of observation sequences in which 0 appears $n_0 = 2$ times and 1 appears $n_1 = 3$ times?

The 10 five-letter words with 0 appearing twice and 1 appearing three times are:

$$\{00111, 01011, 01101, 01110, 10011, 10101, 10110, 11001, 11010, 11100\}.$$

Example 2.14 determines the number of outcomes in the sample space of an experiment with five subexperiments by listing all of the outcomes. Even in this simple example it is not a simple matter to determine all of the outcomes, and in most practical applications of probability there are far more then ten outcomes in the sample space of an experiment and listing them all is out of the question. On the other hand, the counting methods covered in this chapter provide formulas for quickly calculating the number of outcomes in a sample space.

In Example 2.14 each outcome corresponds to the position of three ones in a five-letter binary word. That is, each outcome is completely specified by choosing three positions that contain 1. There are $\binom{5}{3} = 10$ ways to choose three positions in a word. More generally, for length n binary words with n_1 1's, we choose $\binom{n}{n_1}$ slots to hold a 1.

Theorem 2.6

The number of observation sequences for n subexperiments with sample space $S = \{0, 1\}$ with 0 appearing n_0 times and 1 appearing $n_1 = n - n_0$ times is $\binom{n}{n_1}$.

Theorem 2.6 can be generalized to subexperiments with $m > 2$ elements in the sample space. For n trials of a subexperiment with sample space $S_{\text{sub}} = \{s_0, \ldots, s_{m-1}\}$, we want to find the number of outcomes in which s_0 appears n_0 times, s_1 appears n_1 times, and so on. Of course, there are no such outcomes unless $n_0 + \cdots + n_{m-1} = n$. The notation for the number of outcomes is

$$\binom{n}{n_0, \ldots, n_{m-1}}.$$

It is referred to as the *multinomial coefficient*. To derive a formula for the multinomial coefficient, we generalize the logic used in deriving the formula for the binomial coefficient. With n subexperiments, representing the observation sequence by n slots, we first choose n_0 positions in the observation sequence to hold s_0, then n_1 positions to hold s_1, and so on. The details can be found in the proof of the following theorem:

Theorem 2.7

For n repetitions of a subexperiment with sample space $S = \{s_0, \ldots, s_{m-1}\}$, the number of length $n = n_0 + \cdots + n_{m-1}$ observation sequences with s_i appearing n_i times is

$$\binom{n}{n_0, \ldots, n_{m-1}} = \frac{n!}{n_0! n_1! \cdots n_{m-1}!}.$$

Proof Let $M = \binom{n}{n_0, \ldots, n_{m-1}}$. Start with n empty slots and perform the following sequence of subexperiments:

Subexperiment	Procedure
0	Label n_0 slots as s_0.
1	Label n_1 slots as s_1.
$\vdots$	$\vdots$
$m-1$	Label the remaining n_{m-1} slots as s_{m-1}.

There are $\binom{n}{n_0}$ ways to perform subexperiment 0. After n_0 slots have been labeled, there are $\binom{n-n_0}{n_1}$ ways to perform subexperiment 1. After subexperiment $j-1$, $n_0 + \cdots + n_{j-1}$ slots have already been filled, leaving $\binom{n-(n_0+\cdots+n_{j-1})}{n_j}$ ways to perform subexperiment j.

From the fundamental counting principle,

$$\begin{aligned} M &= \binom{n}{n_0}\binom{n-n_0}{n_1}\binom{n-n_0-n_1}{n_2}\cdots\binom{n-n_0-\cdots-n_{m-2}}{n_{m-1}} \\ &= \frac{n!}{(n-n_0)!n_0!}\frac{(n-n_0)!}{(n-n_0-n_1)!n_1!}\cdots\frac{(n-n_0-\cdots-n_{m-2})!}{(n-n_0-\cdots-n_{m-1})!n_{m-1}!}. \end{aligned} \tag{2.14}$$

Canceling the common factors, we obtain the formula of the theorem.

Note that a binomial coefficient is the special case of the multinomial coefficient for an alphabet with $m = 2$ symbols. In particular, for $n = n_0 + n_1$,

$$\binom{n}{n_0, n_1} = \binom{n}{n_0} = \binom{n}{n_1}. \tag{2.15}$$

Lastly, in the same way that we extended the definition of the binomial coefficient, we will employ an extended definition for the multinomial coefficient.

Definition 2.2 Multinomial Coefficient

For an integer $n \geq 0$, we define

$$\binom{n}{n_0, \ldots, n_{m-1}} = \begin{cases} \dfrac{n!}{n_0! n_1! \cdots n_{m-1}!} & \begin{array}{l} n_0 + \cdots + n_{m-1} = n; \\ n_i \in \{0, 1, \ldots, n\}, i = 0, 1, \ldots, m-1, \end{array} \\ 0 & \textit{otherwise.} \end{cases}$$

Example 2.15

In Example 2.13, the professor uses a *curve* in determining student grades. When there are ten students in a probability class, the professor always issues two grades of A, three grades of B, three grades of C and two grades of F. How many different ways can the professor assign grades to the ten students?

In Example 2.13, we determine that with four possible grades there are $4^{10} = 1{,}048{,}576$ ways of assigning grades to ten students. However, now we are limited to choosing $n_0 = 2$ students to receive an A, $n_1 = 3$ students to receive a B, $n_2 = 3$ students to receive a C and $n_3 = 4$ students to receive an F. The number of ways that fit the curve is the multinomial coefficient

$$\binom{n}{n_0, n_1, n_2, n_3} = \binom{10}{2, 3, 3, 2} = \frac{10!}{2!3!3!2!} = 25{,}200. \tag{2.16}$$

Quiz 2.2

Consider a binary code with 4 bits (0 or 1) in each code word. An example of a code word is 0110.

(a) How many different code words are there?

(b) How many code words have exactly two zeroes?

(c) How many code words begin with a zero?

(d) In a constant-ratio binary code, each code word has N bits. In every word, M of the N bits are 1 and the other $N - M$ bits are 0. How many different code words are in the code with $N = 8$ and $M = 3$?

2.3 Independent Trials

Independent trials are identical subexperiments in a sequential experiment. The probability models of all the subexperiments are identical and independent of the outcomes of previous subexperiments. Sampling with replacement is one category of experiments with independent trials.

We now apply the counting methods of Section 2.2 to derive probability models for experiments consisting of independent repetitions of a subexperiment. We start with a simple subexperiment in which there are two outcomes: a success (1) occurs with probability p; otherwise, a failure (0) occurs with probability $1 - p$. The results of all trials of the subexperiment are mutually independent. An outcome of the complete experiment is a sequence of successes and failures denoted by a sequence of ones and zeroes. For example, 10101... is an alternating sequence of successes and failures. Let E_{n_0,n_1} denote the event n_0 failures and n_1 successes in $n = n_0 + n_1$ trials. To find $P[E_{n_0,n_1}]$, we first consider an example.

Example 2.16

What is the probability $P[E_{2,3}]$ of two failures and three successes in five independent trials with success probability p.

To find $P[E_{2,3}]$, we observe that the outcomes with three successes in five trials are 11100, 11010, 11001, 10110, 10101, 10011, 01110, 01101, 01011, and 00111. We note that the probability of each outcome is a product of five probabilities, each related to one subexperiment. In outcomes with three successes, three of the probabilities are p and the other two are $1 - p$. Therefore each outcome with three successes has probability $(1 - p)^2 p^3$.

From Theorem 2.6, we know that the number of such sequences is $\binom{5}{3}$. To find $P[E_{2,3}]$, we add up the probabilities associated with the 10 outcomes with 3 successes, yielding

$$P[E_{2,3}] = \binom{5}{3}(1-p)^2 p^3. \tag{2.17}$$

In general, for $n = n_0 + n_1$ independent trials we observe that

- Each outcome with n_0 failures and n_1 successes has probability $(1-p)^{n_0} p^{n_1}$.

- There are $\binom{n}{n_0} = \binom{n}{n_1}$ outcomes that have n_0 failures and n_1 successes.

Therefore the probability of n_1 successes in n independent trials is the sum of $\binom{n}{n_1}$ terms, each with probability $(1-p)^{n_0} p^{n_1} = (1-p)^{n-n_1} p^{n_1}$.

Theorem 2.8

The probability of n_0 failures and n_1 successes in $n = n_0 + n_1$ independent trials is

$$P[E_{n_0,n_1}] = \binom{n}{n_1}(1-p)^{n-n_1} p^{n_1} = \binom{n}{n_0}(1-p)^{n_0} p^{n-n_0}.$$

The second formula in this theorem is the result of multiplying the probability of n_0 failures in n trials by the number of outcomes with n_0 failures.

Example 2.17

In Example 1.16, we found that a randomly tested resistor was acceptable with probability $P[A] = 0.78$. If we randomly test 100 resistors, what is the probability of T_i, the event that i resistors test acceptable?

Testing each resistor is an independent trial with a success occurring when a resistor is acceptable. Thus for $0 \le i \le 100$,

$$P[T_i] = \binom{100}{i}(0.78)^i(1-0.78)^{100-i} \tag{2.18}$$

We note that our intuition says that since 78% of the resistors are acceptable, then in testing 100 resistors, the number acceptable should be near 78. However, $P[T_{78}] \approx 0.096$, which is fairly small. This shows that although we might expect the number acceptable to be close to 78, that does not mean that the probability of exactly 78 acceptable is high.

Now suppose we perform n independent repetitions of a subexperiment for which there are m possible outcomes for any subexperiment. That is, the sample space for each subexperiment is $(s_0, \ldots, s_{m-1})$ and every event in one subexperiment is independent of the events in all the other subexperiments. Therefore, in every subexperiment the probabilities of corresponding events are the same and we can use the notation $P[s_k] = p_k$ for all of the subexperiments.

An outcome of the experiment consists of a sequence of n subexperiment outcomes. In the probability tree of the experiment, each node has m branches and branch i has probability p_i. The probability of an outcome of the sequential experiment is just the product of the n branch probabilities on a path from the root of the tree to the leaf representing the outcome. For example, with $n = 5$, the outcome $s_2 s_0 s_3 s_2 s_4$ occurs with probability $p_2 p_0 p_3 p_2 p_4$. We want to find the probability of the event

$$E_{n_0,\ldots,n_{m-1}} = \{s_0 \text{ occurs } n_0 \text{ times}, \ldots, s_{m-1} \text{ occurs } n_{m-1} \text{ times}\} \tag{2.19}$$

Note that the notation $E_{n_0,\ldots,n_{m-1}}$ implies that the experiment consists of a sequence of $n = n_0 + \cdots + n_{m-1}$ trials.

To calculate $\mathrm{P}[E_{n_0,\ldots,n_{m-1}}]$, we observe that the probability of the outcome

$$\underbrace{s_0 \cdots s_0}_{n_0 \text{ times}} \underbrace{s_1 \cdots s_1}_{n_1 \text{ times}} \cdots \underbrace{s_{m-1} \cdots s_{m-1}}_{n_{m-1} \text{ times}} \tag{2.20}$$

is

$$p_0^{n_0} p_1^{n_1} \cdots p_{m-1}^{n_{m-1}}. \tag{2.21}$$

Next, we observe that any other experimental outcome that is a reordering of the preceding sequence has the same probability because on each path through the tree to such an outcome there are n_i occurrences of s_i. As a result,

$$\mathrm{P}\left[E_{n_0,\ldots,n_{m-1}}\right] = M p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r} \tag{2.22}$$

where M, the number of such outcomes, is the multinomial coefficient $\binom{n}{n_0,\ldots,n_{m-1}}$ of Definition 2.2. Applying Theorem 2.7, we have the following theorem:

Theorem 2.9

A subexperiment has sample space $S_{sub} = \{s_0, \ldots, s_{m-1}\}$ with $\mathrm{P}[s_i] = p_i$. For $n = n_0 + \cdots + n_{m-1}$ independent trials, the probability of n_i occurences of s_i, $i = 0, 1, \ldots, m-1$, is

$$\mathrm{P}\left[E_{n_0,\ldots,n_{m-1}}\right] = \binom{n}{n_0, \ldots, n_{m-1}} p_0^{n_0} \cdots p_{m-1}^{n_{m-1}}.$$

Example 2.18

A packet processed by an Internet router carries either audio information with probability 7/10, video, with probability 2/10, or text with probability 1/10. Let $E_{a,v,t}$ denote the event that the router processes a audio packets, v video packets, and t text packets in a sequence of 100 packets. In this case,

$$\mathrm{P}\left[E_{a,v,t}\right] = \binom{100}{a, v, t} \left(\frac{7}{10}\right)^a \left(\frac{2}{10}\right)^v \left(\frac{1}{10}\right)^t \tag{2.23}$$

Keep in mind that by the extended definition of the multinomial coefficient, $\mathrm{P}[E_{a,v,t}]$ is nonzero only if $a + v + t = 100$ and a, v, and t are nonnegative integers.

Quiz 2.3

Data packets containing 100 bits are transmitted over a communication link. A transmitted bit is received in error (either a 0 sent is mistaken for a 1, or a 1 sent


```
Y =
  Columns 1 through 12
    47    52    48    46    54    48    47    48    59    44    49    48
  Columns 13 through 24
    42    52    40    40    47    48    48    48    53    49    45    61
  Columns 25 through 36
    60    59    49    47    49    45    48    51    48    53    52    53
  Columns 37 through 48
    56    54    60    53    52    51    58    47    50    48    44    49
  Columns 49 through 60
    50    46    52    50    51    51    57    50    49    56    44    56
```

Figure 2.2 The simulation output of 60 repeated experiments of 100 coin flips.

is mistaken for a 0) with probability $\epsilon = 0.01$, independent of the correctness of any other bit. The packet has been coded in such a way that if three or fewer bits are received in error, then those bits can be corrected. If more than three bits are received in error, then the packet is decoded with errors.

(a) Let $E_{k,100-k}$ denote the event that a received packet has k bits in error and $100-k$ correctly decoded bits. What is $\mathrm{P}[E_{k,100-k}]$ for $k = 0, 1, 2, 3$?

(b) Let C denote the event that a packet is decoded correctly. What is $\mathrm{P}[C]$?

2.4 MATLAB

Two or three lines of MATLAB code are sufficient to simulate an arbitrary number of sequential trials.

We recall from Section 1.6 that `rand(1,m)<p` simulates m coin flips with $\mathrm{P}[\text{heads}] = p$. Because MATLAB can simulate these coin flips much faster than we can actually flip coins, a few lines of MATLAB code can yield quick simulations of many experiments.

Example 2.19

Using MATLAB, perform 60 experiments. In each experiment, flip a coin 100 times and record the number of heads in a vector $\mathbf{Y}$ such that the jth element Y_j is the number of heads in subexperiment j.

```
>> X=rand(100,60)<0.5;
>> Y=sum(X,1)
```

The MATLAB code for this task appears on the left. The 100×60 matrix $\mathbf{X}$ has i,jth element `X(i,j)=0` (tails) or `X(i,j)=1` (heads) to indicate the result of flip i of subexperiment j. Since Y sums X across the first dimension, Y(j) is the number of heads in the jth subexperiment. Each Y(j) is between 0 and 100 and generally in the neighborhood of 50. The output of a sample run is shown in Figure 2.2.

Example 2.20

In testing a microprocessor all four grades, 1, 2, 3 and 4, have probability 0.25, independent of any other microprocessor. Simulate the testing of 100 microprocessors. Your output should be a 4×1 vector $\mathbf{X}$ such that X_i is the number of grade i microprocessors.

```
%chiptest.m
G=ceil(4*rand(1,100));
T=1:4;
X=hist(G,T);
```

The first line generates a row vector G of random grades for 100 microprocessors. The possible test scores are in the vector T. Lastly, X=hist(G,T) returns a histogram vector X such that X(j) counts the number of elements G(i) that equal T(j).

Note that "help hist" will show the variety of ways that the hist function can be called. Morever, X=hist(G,T) does more than just count the number of elements of G that equal each element of T. In particular, hist(G,T) creates bins centered around each T(j) and counts the number of elements of G that fall into each bin.

Note that in MATLAB all variables are assumed to be matrices. In writing MATLAB code, X may be an $n \times m$ matrix, an $n \times 1$ column vector, a $1 \times m$ row vector, or a 1×1 scalar. In MATLAB, we write X(i,j) to index the i,jth element. By contrast, in this text, we vary the notation depending on whether we have a scalar X, or a vector or matrix $\mathbf{X}$. In addition, we use $X_{i,j}$ to denote the i,jth element. Thus, $\mathbf{X}$ and X (in a MATLAB code fragment) may both refer to the same variable.

Quiz 2.4

The flip of a thick coin yields heads with probability 0.4, tails with probability 0.5, or lands on its edge with probability 0.1. Simulate 100 thick coin flips. Your output should be a 3×1 vector $\mathbf{X}$ such that X_1, X_2, and X_3 are the number of occurrences of heads, tails, and edge.

Problems

Difficulty: • Easy ▪ Moderate ◆ Difficult ◆◆ Experts Only

2.1.1 • Suppose you flip a coin twice. On any flip, the coin comes up heads with probability 1/4. Use H_i and T_i to denote the result of flip i.

(a) What is the probability, $P[H_1|H_2]$, that the first flip is heads given that the second flip is heads?

(b) What is the probability that the first flip is heads and the second flip is tails?

2.1.2 • For Example 2.2, suppose $P[G_1] = 1/2$, $P[G_2|G_1] = 3/4$, and $P[G_2|R_1] = 1/4$. Find $P[G_2]$, $P[G_2|G_1]$, and $P[G_1|G_2]$.

2.1.3 • At the end of regulation time, a basketball team is trailing by one point and a player goes to the line for two free throws. If the player makes exactly one free throw, the game goes into overtime. The probability that the first free throw is good is 1/2. However, if the first attempt is good, the player relaxes and the second attempt is good with probability 3/4. However, if the player misses the first attempt, the added pressure reduces the success probability to 1/4. What is the probability that the game goes into overtime?

2.1.4 • You have two biased coins. Coin A comes up heads with probability 1/4. Coin B comes up heads with probability 3/4. However, you are not sure which is which, so you choose a coin randomly and you flip it. If the flip is heads, you guess that the flipped coin is B; otherwise, you guess that the flipped coin is A. What is the probability $P[C]$ that your guess is correct?

2.1.5 • Suppose you have two coins, one biased, one fair, but you don't know which coin is which. Coin 1 is biased. It comes up heads with probability 3/4, while coin 2 comes up heads with probability 1/2. Suppose you pick a coin at random and flip it. Let C_i denote the event that coin i is picked. Let H and T denote the possible outcomes of the flip. Given that the outcome of the flip is a head, what is $P[C_1|H]$, the probability that you picked the biased coin? Given that the outcome is a tail, what is the probability $P[C_1|T]$ that you picked the biased coin?

2.1.6 ▪ Suppose that for the general population, 1 in 5000 people carries the human immunodeficiency virus (HIV). A test for the presence of HIV yields either a positive (+) or negative (−) response. Suppose the test gives the correct answer 99% of the time. What is $P[-|H]$, the conditional probability that a person tests negative given that the person does have the HIV virus? What is $P[H|+]$, the conditional probability that a randomly chosen person has the HIV virus given that the person tests positive?

2.1.7 ▪ A machine produces photo detectors in pairs. Tests show that the first photo detector is acceptable with probability 3/5. When the first photo detector is acceptable, the second photo detector is acceptable with probability 4/5. If the first photo detector is defective, the second photo detector is acceptable with probability 2/5.

(a) Find the probability that exactly one photo detector of a pair is acceptable.

(b) Find the probability that both photo detectors in a pair are defective.

2.1.8 ▪ You have two biased coins. Coin A comes up heads with probability 1/4. Coin B comes up heads with probability 3/4. However, you are not sure which is which so you flip each coin once, choosing the first coin randomly. Use H_i and T_i to denote the result of flip i. Let A_1 be the event that coin A was flipped first. Let B_1 be the event that coin B was flipped first. What is $P[H_1 H_2]$? Are H_1 and H_2 independent? Explain your answer.

2.1.9 ▪ A particular birth defect of the heart is rare; a newborn infant will have the defect D with probability $P[D] = 10^{-4}$. In the general exam of a newborn, a particular heart arrhythmia A occurs with probability 0.99 in infants with the defect. However, the arrhythmia also appears with probability 0.1 in infants without the defect. When the arrhythmia is present, a lab test for the defect is performed. The result of the lab test is either positive (event T^+) or negative (event T^-). In a newborn with the defect, the lab test is positive with probability $p = 0.999$ independent from test to test. In a newborn without the defect, the lab test is negative with probability $p = 0.999$. If the arrhythmia is present and the test is positive, then heart surgery (event H) is performed.

(a) Given the arrythmia A is present, what is the probability the infant has the defect D?

(b) Given that an infant has the defect, what is the probability $P[H|D]$ that heart surgery is performed?

(c) Given that the infant does not have the defect, what is the probability