CHAPTER 1

EXPERIMENTS, MODELS, AND PROBABILITIES

GETTING STARTED WITH PROBABILITY

You have read the "Hints on Studying Probability" in the Preface. Now you can begin. 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.

physical property

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 of our knowledge 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 you 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, you have to take account of all relevant facts.

Rabardast

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 you 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 applies 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, while axioms are facts that we have to 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, 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 will be 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 SET THEORY

The mathematical basis of probability is the theory of sets. Most people who study probability have already encountered set theory and are familiar with such terms as *set, element, union, intersection*, and *complement*. For them, the following paragraphs will review material already learned and introduce the notation and terminology we use here. For people who have no prior acquaintance with sets, this material introduces basic definitions and the properties of sets that are important in the study of probability.

A *set* is a collection of things. We use capital letters to denote sets. The things that together make up the set are *elements*. When we use mathematical notation to refer to set elements, we usually use small letters. Thus we can have a set A with elements x, y, and z. The symbol $\in$ denotes set inclusion. Thus $x \in A$ means "x is an element of set A." The symbol $\notin$ is the opposite of $\in$. Thus $c \notin A$ means "c is not an element of set A."

It is essential when working with sets to have a definition of each set. The definition allows someone to consider anything conceivable and determine whether that thing is an element of the set. There are many ways to define a set. One way is simply to name the elements:

$$A = \{\text{Rutgers University, Princeton University, the planet Mercury}\}$$

Note that in stating the definition, we write the name of the set on one side of = and the definition in curly brackets { } on the other side of =.

It follows that "the planet closest to the Sun $\in A$" is a true statement. It is also true that "Whitney Houston $\notin A$." Another way of writing the set is to give a mathematical rule for generating all of the elements of the set:

$$B = \left\{x^2 | x = 1, 2, 3, \ldots\right\}$$

This notation tells us to form a set by performing the operation to the left of the vertical bar, |, on the numbers to the right of the bar. The dots tell us to continue the sequence to the left of the dots. Since there is no number to the right of the dots, we continue the sequence indefinitely, forming an infinite set. The definition of B implies that $9 \in B$ and $10 \notin B$.

Yet another type of definition is a rule for testing something to determine whether it is a member of the set:

$$C = \{\text{all Rutgers juniors who weigh more than 170 pounds}\}$$

In addition to set inclusion, we also have the notion of a *subset*, which describes a possible relationship between two sets. By definition, A is a subset of B if every member of A is also a member of B. We use the symbol $\subset$ to denote subset. Thus $A \subset B$ is mathematical notation for the statement "the set A is a subset of the set B." If

$$I = \{\text{all positive integers, negative integers, and } 0\}$$

it follows that $B \subset I$, where B is defined above.

The definition of set equality

$$A = B$$

is

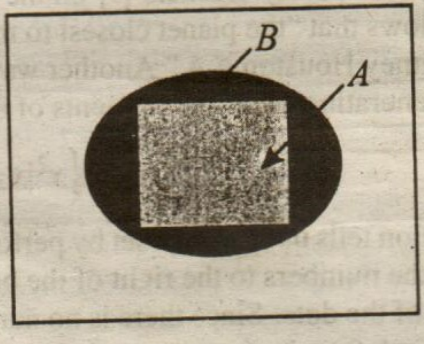
$$A = B \text{ if and only if } B \subset A \text{ and } A \subset B$$

This is the mathematical way of stating that A and B are identical if and only if every element of A is an element of B and every element of B is an element of A. This definition implies that a set is unaffected by the order of the elements in a definition. For example, $\{0,17,46\} = \{17,0,46\} = \{46,0,17\}$ are all the same set.

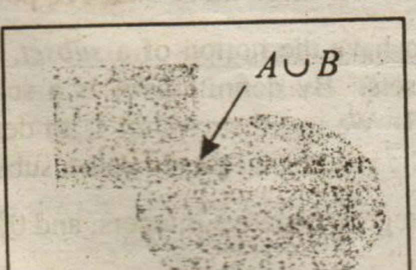
To work with sets mathematically it is necessary to define a *universal set*. This is the set of all things that we could possibly consider in a given context. In any study, all set operations relate to the universal set for that study. The members of the universal set include all of the elements of all of the sets in the study. We will use the letter S to denote the universal set. For example, the universal set for A could be $S = \{\text{all universities in New Jersey, all planets}\}$. The universal set for B could be $S = I = \{0,1,2,\ldots\}$. By definition, every set is a subset of the universal set. That is, for any set X, $X \subset S$.

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 ϕ. By definition ϕ is a subset of every set. For any set A, $\phi \subset A$.

It is customary to refer to Venn diagrams to display relationships among sets. By convention, the region enclosed by the large rectangle is the universal set S. Closed surfaces within this rectangle denote sets. A Venn diagram depicting the relationship $A \subset B$ is



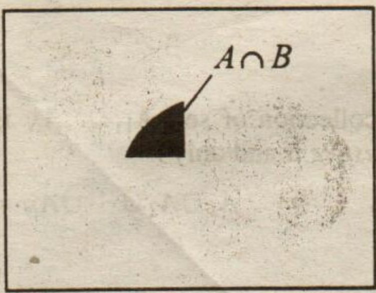
When we do set algebra, we form new sets from existing sets. There are three operations for doing this: *union*, *intersection*, and *complement*. Union and intersection combine two existing sets to produce a third set. The complement operation forms a new set from one existing set. The notation and definitions are



The *union of sets* A and B is the set of all elements that are either in A or in B, or in both. The union of A and B is denoted by $A \cup B$. In this Venn diagram, $A \cup B$ is the complete shaded area. Formally, the definition states

$$x \in A \cup B \text{ if and only if } x \in A \text{ or } x \in B$$

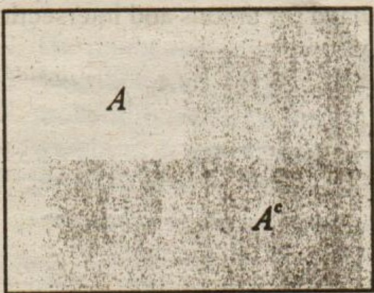
The set operation union corresponds to the logical "or" operation.



The *intersection* of two sets A and B is the set of all elements which are contained both in A and B. The intersection is denoted by $A \cap B$. Another notation for intersection is AB. Formally, the definition is

$$x \in A \cap B \text{ if and only if } x \in A \text{ and } x \in B$$

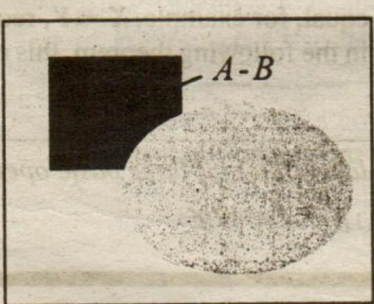
The set operation intersection corresponds to the logical "and" function.



The *complement of a set A*, denoted by A^c, is the set of all elements in S that are not in A. The complement of S is the null set ϕ. Formally,

$$x \in A^c \text{ if and only if } x \notin A$$

A fourth set operation is called the *difference*. It is a combination of intersection and complement.

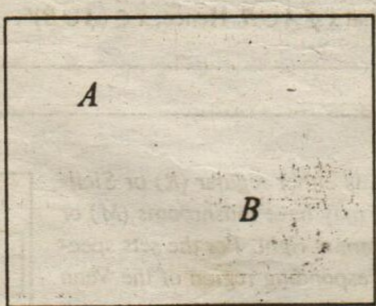


The *difference* between A and B is a set $A - B$ that contains all elements of A that are *not* elements of B. Formally,

$$x \in A - B \text{ if and only if } x \in A \text{ and } x \notin B$$

Note that $A - B = A \cap B^c$ and $A^c = S - A$.

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

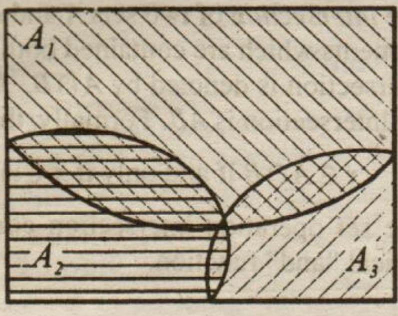


A collection of sets $A_1, \ldots, A_N$ is *mutually exclusive* if and only if

$$A_i \cap A_j = \phi \qquad i \neq j$$

When there are only two sets in the collection, we say that these sets are *disjoint*. Formally, A and B are disjoint if and only if

$$A \cap B = \phi$$



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

$$A_1 \cup A_2 \cup \cdots \cup A_N = S$$

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:

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

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

From the definition of set operations, we can derive many important relationships between sets and other sets derived from them. One example is

$$A - B \subset A$$

To prove that this is true, it is necessary to show that if $x \in A - B$, then it is also true that $x \in A$. A proof that two sets are equal, for example, $X = Y$, requires two separate proofs: $X \subset Y$ and $Y \subset X$. As we see in the following theorem, this can be complicated to show.

Theorem 1.1. *De Morgan's law relates all three basic operations:*

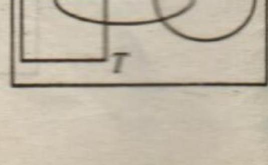
$$(A \cup B)^c = A^c \cap B^c$$

Proof There are two parts to the proof:

- To show $(A \cup B)^c \subset A^c \cap B^c$, suppose $x \in (A \cup B)^c$. That implies $x \notin A \cup B$. Hence, $x \notin A$ and $x \notin B$, which together imply $x \in A^c$ and $x \in B^c$. That is, $x \in A^c \cap B^c$.
- To show $A^c \cap B^c \subset (A \cup B)^c$, suppose $x \in A^c \cap B^c$. In this case, $x \in A^c$ and $x \in B^c$. Equivalently, $x \notin A$ and $x \notin B$ so that $x \notin A \cup B$. Hence, $x \in (A \cup B)^c$.

Quiz 1.1.

A slice of pizza sold by Gerlanda's Pizza is either regular (R) or Sicilian (T as in Thick) In addition, each slice may have mushrooms (M) or onions (O) as described by the Venn diagram at right. For the sets specified in parts (a)–(g) below, shade the corresponding region of the Venn diagram.



(a) R

(b) $M \cup O$

(c) $M \cap O$

(d) $R \cup M$

(e) $R \cap M$

(f) $T^c - M$

(g) $M - T^c$

(h) Are T and M mutually exclusive?

(i) Are R, T, and M collectively exhaustive?

(j) Are T and O mutually exclusive? State this condition in words.

1.2 APPLYING SET THEORY TO PROBABILITY

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 some 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 on heads or tails?
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 1.1. *An experiment consists of the following procedure, observation and model:*

- *Procedure: Flip a coin and let it land on a table.*
- *Observation: Observe which side (head or tail) faces you after the coin lands.*
- *Model: Heads and tails are equally likely. The result of each flip is unrelated to the results of previous flips.*

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 1.2. *Flip a coin three times. Observe the sequence of heads and tails.*

Example 1.3. *Flip a coin three times. Observe the number of heads.*

These two experiments have the same procedure: flip a coin three times. 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 any 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 1.4. *The sample space in Example 1.1 is $S = \{h, t\}$ where h is the outcome "observe head," and t is the outcome "observe tail." The sample space in Example 1.2 is*

$$S = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$$

The sample space in Example 1.3 is $S = \{0, 1, 2, 3\}$.

Example 1.5. *Manufacture an integrated circuit and test it to determine whether it meets quality objectives. The possible outcomes are "accepted" (a) and "rejected" (r). The sample space is $S = \{a, r\}$.*

In common speech, an event is just 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 of the sample space.

> ***Definition 1.3. Event:*** An *event* is a set of outcomes of an experiment.

The following table relates the terminology of probability to set theory:

Set Algebra	Probability
set	event
universal set	sample space
element	outcome

All of this is 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. *Suppose we roll a six sided die and observe the number of dots on the side facing upwards. We can label these outcomes $i = 1, \ldots, 6$ where i denotes the outcome that i dots appear on the up face. The sample space is $S = \{1, 2, \ldots, 6\}$. Each subset of S is an event. Examples of events are*

- *The event $E_1 = \{$Roll 4 or higher$\} = \{4, 5, 6\}$.*
- *The event $E_2 = \{$The roll is even$\} = \{2, 4, 6\}$.*
- *$E_3 = \{$The roll is the square of an integer$\} = \{1, 4\}$.*

Example 1.7. *Wait for someone to make a phone call and observe the duration of the call in minutes. An outcome x is a nonnegative real number. The sample space is $S = \{x | x \geq 0\}$. The event "the phone call lasts longer than five minutes" is $\{x | x > 5\}$.*

Example 1.8. *Consider three traffic lights encountered driving down a road. We say a light was red if the driver was required to come to a complete stop at that light; otherwise we call the light green. For the sake of simplicity, these definitions were carefully chosen to exclude the case of the yellow light. An outcome of the experiment is a description of whether each light was red or green. We can denote the outcome by a sequence of r and g such as rgr, the outcome that the first and third lights were red but the second light was green. We denote the event that light n was red or green* *by R_n or G_n. The event R_2 would be the set of outcomes $\{grg, grr, rrg, rrr\}$. We can also denote an outcome as an intersection of events R_i and G_j. For example, the event $R_1G_2R_3$ is the set containing the single outcome $\{rgr\}$.*

In Example 1.8, suppose we were interested only in the status of light 2. In this case, the set of events $\{G_2, R_2\}$ describes the events of interest. Moreover, for each possible outcome of the three light experiment, the second light was either red or green, so the set of events $\{G_2, R_2\}$ is both mutually exclusive and collectively exhaustive. However, $\{G_2, R_2\}$ is not a sample space for the experiment because the elements of the set do not completely describe the set of possible outcomes of the experiment. The set $\{G_2, R_2\}$ does not have the finest-grain property. Yet sets of this type are sufficiently useful to merit a name of their own.

> ***Definition 1.4. Event Space:*** An *event space* is a collectively exhaustive, mutually exclusive set of events.

An event space and a sample space have a lot in common. The members of both are mutually exclusive and collectively exhaustive. They differ in the finest-grain property that applies to a sample space but not to an event space. Because it possesses the finest-grain property, a sample space contains all the details of an experiment. The members of a sample space are *outcomes*. By contrast, the members of an event space are *events*. The event space is a set of events (sets), while the sample space is a set of outcomes (elements). Usually, a member of an event space contains many outcomes. Consider a simple example:

Example 1.9. *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 1.10. *For the four coins experiment of Example 1.9, let $B_i = \{\text{outcomes with } i \text{ heads}\}$ for $i = 0, 1, 2, 3, 4$. 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 an event space. 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 1.9 and Example 1.10 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 modem transmit four bits from one telephone 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 only interested in the number of times one of the letters occurs, it is sufficient to refer only to the event space B, which does not contain all of the information about the experiment but does contain all of the information we need. The event space is simpler to deal with than the sample space because it has fewer members (there are five events in the event space and 16 outcomes in the sample space). The simplification is much more significant when the complexity of the experiment is higher; for example, testing 10 circuits. The sample space has $2^{10} = 1024$ members, while the corresponding event space has only 11 members.

The concept of an event space is useful because it allows us to express any event as a union of mutually exclusive events. We will observe in the next section that the entire theory of probability is based on unions of mutually exclusive events.

> ***Theorem 1.2.*** *For an event space $B = \{B_1, B_2, \dots\}$ 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*
>
> $$A = C_1 \cup C_2 \cup \cdots$$

Example 1.11. *In the coin tossing experiment of Example 1.9, let A equal the set of outcomes with less than three heads.*

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

From Example 1.10, let $\{B_0, \dots, B_4\}$ denote the event space in which $B_i = \{\text{outcomes with } i \text{ heads}\}$. Theorem 1.2 states that

$$A = (A \cap B_0) \cup (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup (A \cap B_4)$$

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 $i = 4$, $A \cap B_i = \phi$ so that $A = B_0 \cup B_1 \cup B_2$, a union of disjoint 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 Example 1.11 and Theorem 1.2. 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 come from a fabrication machine.

Quiz 1.2. *Monitor three consecutive phone calls going through a telephone switching office. Classify each one as a voice call (v), if someone is speaking or a data call (d) if the call is carrying a modem or fax signal. Your observation is a sequence of three letters (each letter is either v or d). For example, two voice calls followed by one data call corresponds to vvd. Write the elements of the following sets:*

(a) $A_1 = \{\text{first call is a voice call}\}$ *(b) $B_1 = \{\text{first call is a data call}\}$*

(c) $A_2 = \{$*second call is a voice call*$\}$ (d) $B_2 = \{$*second call is a data call*$\}$

(e) $A_3 = \{$*all calls are the same*$\}$ (f) $B_3 = \{$*voice and data alternate*$\}$

(g) $A_4 = \{$*one or more voice calls*$\}$ (h) $B_4 = \{$*two or more data calls*$\}$

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

1.3 PROBABILITY AXIOMS

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.

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 analyze a complicated event in order to express it as the union of mutually exclusive events whose probabilities we can calculate. Then, we 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 disjoint events.

Theorem 1.3. *Given events A and B such that $A \cap B = \phi$, then*

$$P[A \cup B] = P[A] + P[B]$$

Although it may appear that Theorem 1.3 is a trivial special case of Axiom 3, this is not so. In fact, a simple proof of Theorem 1.3 may also use Axiom 2! If you are curious, Problem 1.4.7 gives the first steps toward a proof. It is a simple matter to extend Theorem 1.3 to any finite union of mutually exclusive sets.

Theorem 1.4. *If $B = B_1 \cup B_2 \cup \cdots \cup B_m$ and $B_i \cap B_j = \phi$ for $i \neq j$, then*

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

In Chapter 7, 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. In these n trials, $N_A(n)$ is the number of times that event A occurs. The relative frequency of A is the fraction $N_A(n)/n$. In Chapter 7, we prove that $\lim_{n\to\infty} N_A(n)/n = P[A]$.

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

Theorem 1.5. *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 disjoint sets:

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

with $\{s_i\} \cap \{s_j\} = \phi$ for $i \neq j$. Theorem 1.4 leads to

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

which completes the proof.

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, $P[A \cap B] = P[A, B] = P[AB]$.

Example 1.12. *Let T_i denote the duration (in minutes) of the ith phone call you place today. The probability that your first phone call lasts less than five minutes and your second phone call lasts at least ten minutes is $P[T_1 < 5, T_2 \geq 10]$.*

EQUALLY LIKELY OUTCOMES

A large number of experiments can be modeled by a sample space $S = \{s_1, \ldots, s_n\}$ in which the n outcomes are equally likely. In such experiments, there are usually symmetry arguments that lead us to believe that no one outcome is any more likely than any other. In such a case, the axioms of probability imply that every outcome has probability $1/n$.

> ***Theorem 1.6.*** *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 1.5 implies

$$P[S] = P[s_1] + \cdots + P[s_n] = np$$

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

Example 1.13. *As in Example 1.6, 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

$$P[i] = 1/6 \qquad i = 1, 2, \ldots, 6$$

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 1.3. *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. Given that all scores between 51 and 100 are equally likely and a score of 50 or less cannot occur, please find the following probabilities:*

(a) $P[\{s_{79}\}]$ (b) $P[\{s_{100}\}]$

(c) $P[A]$ (d) $P[F]$

(e) $P[T \geq 80]$ (f) $P[T < 90]$

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

1.4 SOME CONSEQUENCES OF THE AXIOMS

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

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

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

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

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

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

The following theorem is a more complex consequence of the axioms. It is very useful. It refers to an event space $B_1, B_2, \ldots, B_m$ 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 event space $\{B_1, B_2, \ldots, B_m\}$,*

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

Proof The proof follows directly from Theorem 1.2 and Theorem 1.4. In this case, the disjoint sets are $C_i = \{A \cap B_i\}$.

Example 1.14. *A company has a model of telephone usage. It classifies all calls as either long (L), if they last more than three minutes, or brief (B). It also observes whether calls carry voice (V), data (D) or fax (F). This model implies an experiment in which the procedure is to monitor a call. The observation consists of the nature of the call, V, D, or F, and the length, L or B. The sample space has six outcomes $S = \{LV, BV, LD, BD, LF, BF\}$. The probabilities can be represented in the table in which the rows and columns are labeled by events and a table entry represents the probability of the intersection of the corresponding row and column events. In this case, the table is*

	V	F	D
L	0.3	0.15	0.12
B	0.2	0.15	0.08

For example, from the table we can read that the probability of a brief data call is $P[BD] = 0.08$. *Note that* $\{V, D, F\}$ *is an event space corresponding to* $\{B_1, B_2, B_3\}$ *in Theorem 1.8. Thus we can apply Theorem 1.8 to find*

$$P[L] = P[LV] + P[LD] + P[LF] = 0.57$$

Quiz 1.4. *Monitor a phone call. Classify the call as a voice call* (V), *if someone is speaking; or a data call* (D) *if the call is carrying a modem or fax signal. Classify the call as long* (L) *if if the call lasts for more than three minutes; otherwise classify the call as brief* (B). *Based on data collected by the telephone company, we use the following probability model:* $P[V] = 0.7$, $P[L] = 0.6$, $P[VL] = 0.35$. *Please find the following probabilities:*

(a) $P[DL]$	(b) $P[D \cup L]$
(c) $P[VB]$	(d) $P[V \cup L]$
(e) $P[V \cup D]$	(f) $P[LB]$

1.5 CONDITIONAL PROBABILITY

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 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.

$n_N(A), n_N(M), n_N(A,M) = n_N(A \cap M)$

$\frac{n_N(A,M)}{n(M)} = \frac{n_N(A \cap M)/N}{n_N(M)/N} = \frac{P(A \cap M)}{P(M)} = P(A|M)$

Relative frequency

Example 1.15. *Consider an experiment that consists of testing two integrated circuits that come from the same silicon wafer, and observing in each case whether a circuit is accepted (a) or rejected (r). There are four outcomes of the experiment rr, ra, ar, and aa. Let A denote the event that the second circuit is a failure. Mathematically,* $A = \{rr, ar\}$.

The circuits come from a high-quality production line. Therefore the prior probability of A is very low. In advance, we are pretty certain that the second circuit will be acceptable. However, some wafers become contaminated by dust, and these wafers have a high proportion of defective chips. Let $B = \{rr, ra\}$ *denote the event that the first chip tested is rejected.*

Given the knowledge that the first chip was rejected, our knowledge of the quality of the second chip changes. With the first chip a reject, the likelihood that the second chip will also be rejected is higher than it was a priori. Thus we would assign a higher probability to the rejection of the second chip when we know that the first chip is a reject.

The notation for this new probability is $P[A|B]$. We read this as "the probability of A given B." Mathematically, the definition of conditional probability is as follows.

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 makes no sense 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 1.9. *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 $A_i \cap A_j = \phi$ 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 1.16. *With respect to Example 1.15 for the "testing two chips," consider the a priori probability model:*

$$P[rr] = 0.01,\ P[ra] = 0.01,\ P[ar] = 0.01,\ P[aa] = 0.97$$

Find the probability of events 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 1.15 that A is the union of two disjoint events (outcomes) rr and ar. Therefore, the a priori probability that the second chip is rejected is

$$P[A] = P[rr] + P[ar] = 0.02$$

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

$$P[B] = P[rr] + P[ra] = 0.02.$$

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,

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

Thus

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

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 1.17. *Shuffle a deck of cards and observe the bottom card. What is the conditional probability the bottom card is the ace of clubs given that the the bottom card is a black card?*

. . . .

The sample space consists of the 52 cards that can appear on the bottom of the deck. Let A_C denote the event that the bottom card is the ace of clubs. Since all cards are equally likely to be at the bottom, the probability that a particular card, such as the ace of clubs, is at the bottom is $P[A_C] = 1/52$. Let B be the event that the bottom card is a black card. The event B occurs if the bottom card is one of the 26 clubs or spades, so that $P[B] = 26/52$. Given B, the conditional probability of the event A_C is

$$P[A_C|B] = \frac{P[A_C B]}{P[B]} = \frac{P[A_C]}{P[B]} = \frac{1/52}{26/52} = \frac{1}{26}$$

The key step was observing that $A_C B = A_C$ since if the bottom card is the ace of clubs then the bottom card must be a black card. Mathematically, this is an example of the fact that $A_C \subset B$ implies that $A_C B = A_C$.

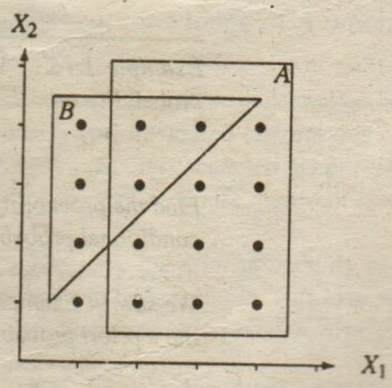
Example 1.18. *Roll two four-sided dice. Let X_1 and X_2 denote the number of dots that appear on die 1 and die 2, respectively. Draw the 4 by 4 sample space. Let A be the event $X_1 \geq 2$. What is $P[A]$? Let B denote the event $X_2 > X_1$. What is $P[B]$? What is $P[A|B]$?*

. . . .

Each outcome is a pair (X_1, X_2). To find $P[A]$, we add up the probabilities of the sample points in A.

From the sample space, we see that A has 12 points, each with probability $1/16$, so $P[A] = 12/16 = 3/4$. To find $P[B]$, we observe that B has 6 points and $P[B] = 6/16 = 3/8$. The compound event AB has exactly three points, $(2,3), (2,4), (3,4)$, so $P[AB] = 3/16$. From the definition of conditional probability, we write

$$P[A|B] = \frac{P[AB]}{P[B]} = 1/2$$



LAW OF TOTAL PROBABILITY

In many situations, we begin with information about conditional probabilities. Using these conditional probabilities, we would like to calculate unconditional probabilities. The law of total probability shows us how to do this.

Theorem 1.10. Law of Total Probability. *If $B_1, B_2, \ldots, B_m$ is an event space and $P[B_i] > 0$ for $i = 1, \ldots, m$, then*

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

This follows from Theorem 1.8 and the identity

$$P[AB] = P[A|B]P[B]. \tag{1.1}$$

which is a direct consequence of the definition of conditional probability. The usefulness of the result can be seen in the next example.

✓ ***Example 1.19.*** *A company has three machines B_1, B_2, and B_3 for making 1 kΩ resistors. It has been observed that 80% of resistors produced by B_1 are within 50 Ω of the nominal value. Machine B_2 produces 90% of resistors within 50 Ω of the nominal value. 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 a resistor that is within 50 Ω of the nominal value?*

....

Let $A = \{$resistor is within 50 Ω of the nominal value$\}$. Using the resistor accuracy information to formulate a probability model, we write

$$P[A|B_1] = 0.8, \quad P[A|B_2] = 0.9, \quad P[A|B_3] = 0.6$$

The production figures state that $3000 + 4000 + 3000 = 10,000$ resistors per hour are produced. The fraction from machine B_1 is $P[B_1] = 3000/10000 = 0.3$. 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 accuracy probability for all resistors shipped by the company:

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

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

BAYES' THEOREM

In many situations, we have advance information about $P[A|B]$ and need to calculate $P[B|A]$. To do so we have the following formula:

Theorem 1.11. Bayes' theorem.

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

The proof just combines the definition $P[B|A] = P[AB]/P[A]$ with Equation (1.1).

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 an event space that includes all possible states of something that interests us but 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 1.11. Thus for state B_i,

$$P[B_i|A] = \frac{P[A|B_i]P[B_i]}{\sum_{i=1}^{m} P[A|B_i]P[B_i]} \tag{1.2}$$

Example 1.20. *In a shipment of resistors from the factory, we learned in Example 1.19 the following facts:*

- *The probability a resistor is from machine B_3 is $P[B_3] = 0.3$.*
- *The probability a resistor is acceptable, i.e. within 50 Ω of the nominal value, is $P[A] = 0.78$.*
- *Given a resistor is from machine B_3, the conditional probability 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 need to find $P[B_3|A]$. Using Bayes' theorem, we have

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

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

$$P[B_3|A] = (0.6)(0.3)/(0.78) = 0.23$$

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 coming from the factory). Machine B_1 produces 31% of the resistors that meet the 50 Ω criterion and machine B_2 produces 46% of them.

Quiz 1.5. *Monitor three consecutive phone calls going through a telephone switching office. Classify each one as a voice call (v), if someone is speaking; or a data call (d) if the call is carrying a modem or fax signal. Your observation is a sequence of three letters (each one is either v or d). For example, three voice calls corresponds to vvv. The outcomes vvv and ddd have probability 0.2 while the other outcomes vvd, vdv, vdd, dvv, dvd, and ddv each have probability 0.1. Count the number of voice calls N_V in the three calls you have observed. Consider the four events $N_V = 0$, $N_V = 1$, $N_V = 2$, $N_V = 3$. 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) $P[N_V \geq 1|N_V = 2]$

1.6 INDEPENDENCE

Definition 1.6. 2 Independent Events: Events A and B are *independent* if and only if

$$P[AB] = P[A]P[B] \tag{1.3}$$

The following formulas are equivalent to the definition of independent events A and B.

$$P[A|B] = P[A] \qquad P[B|A] = P[B] \tag{1.4}$$

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 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 1.6.7 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 should not alter the probability of A.

Keep in mind that **independent and disjoint are *not* synonyms.** In some contexts these words can have similar meanings, but this is not the case in probability. Disjoint events have no outcomes in common and therefore $P[AB] = 0$. In most situations independent events are not disjoint! Exceptions occur only when $P[A] = 0$ or $P[B] = 0$. When we have to calculate probabilities, knowledge that events A and B are *disjoint* 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 1.21. *Suppose that for the three traffic lights of Example 1.8, each outcome (a sequence of three lights, each either red or green) is equally likely. Are the events R_2 that the second light was red and G_2 that the second light was green independent? Are the events R_1 and R_2 independent?*

....

Each element of the sample space

$$S = \{rrr, rrg, rgr, rgg, grr, grg, ggr, ggg\}$$

has probability 1/8. The events $R_2 = \{rrr, rrg, grr, grg\}$ and $G_2 = \{rgr, rgg, ggr, ggg\}$ each contain four outcomes so $P[R_2] = P[G_2] = 4/8$. However, $R_2 \cap G_2 = \phi$ and $P[R_2G_2] = 0$. That is, R_2 and

G_2 must be disjoint since the second light cannot have been both red and green. Since $P[R_2G_2] \neq P[R_2]P[G_2]$, R_2 and G_2 are not independent.

The events $R_1 = \{rgg, rgr, rrg, rrr\}$ and $R_2 = \{rrg, rrr, grg, grr\}$ each have four outcomes so $P[R_1] = P[R_2] = 4/8$. In this case, the intersection $R_1 \cap R_2 = \{rrg, rrr\}$ has probability $P[R_1R_2] = 2/8$. Since $P[R_1R_2] = P[R_1]P[R_2]$, events R_1 and R_2 are independent.

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 1.22. *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 test?*

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

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

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 in the sample space as

$$P[(ma, ea)] = P[ME] = P[M]P[E] = 0.95 \times 0.8 = 0.76$$
$$P[(ma, er)] = P[ME^c] = P[M]P[E^c] = 0.95 \times 0.2 = 0.19$$
$$P[(mr, ea)] = P[M^cE] = P[M^c]P[E] = 0.05 \times 0.8 = 0.04$$
$$P[(mr, er)] = P[M^cE^c] = P[M^c]P[E^c] = 0.05 \times 0.2 = 0.01$$

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. 3 Independent Events: A_1, A_2, and A_3 are *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 1.23. *In an experiment with equiprobable outcomes, the event space 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 $A_3 = \phi$ independent?*

....

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

$$P[A_1 \cap A_2 \cap A_3] = P[A_1]P[A_2]P[A_3] = 0$$

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

$$P[A_1 \cap A_2] = P[2,3] = 1/2 \neq P[A_1]P[A_2] = 3/4 \times 3/4$$

Hence the three events are dependent.

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

Definition 1.8. More than Two Independent Events: If $n \geq 3$, the sets $A_1, A_2, \ldots, A_n$ are independent if and only if

(a) Every set of $n-1$ sets taken from $A_1, A_2, \ldots A_n$ is independent.

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

Quiz 1.6. *Monitor two consecutive phone calls going through a telephone switching office. Classify each one as a voice call (v), if someone is speaking; or a data call (d) if the call is carrying a modem or fax signal. Your observation is a sequence of two letters (either v or d). For example, two voice calls corresponds to vv. The two calls are independent and the probability that any one of them is a voice call is 0.8. Denote the identity of call i by C_i. If call i is a voice call, then $C_i = v$; otherwise, $C_i = d$. Count the number of voice calls in the two calls you have observed. N_V is the number of voice calls. Consider the three events $N_V = 0$, $N_V = 1$, $N_V = 2$. 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 $\{N_V$ is even$\}$

1.7 SEQUENTIAL EXPERIMENTS AND TREE DIAGRAMS

Many experiments consist of a sequence of *subexperiments*. The procedure followed for each subexperiment may depend on the results of the previous subexperiments. We can use a tree diagram to represent the sequential nature of the subexperiments. Following the procedure and recording the observations of the experiment is equivalent to following a sequence of branches from the root of the tree to a leaf. Each leaf corresponds to an outcome of the experiment.

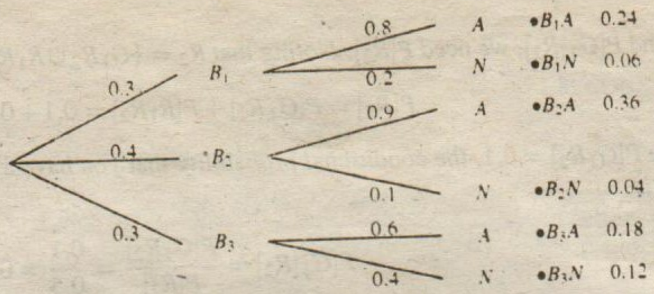
It is natural to model conditional probabilities in terms of sequential experiments and to illustrate them with tree diagrams. At the root of the tree, the probability of a particular event is described by our a priori knowledge. If the possible results of the first subexperiment are described by the events $B_1, \ldots, B_m$, then $\{B_1, \ldots, B_m\}$ is an event space. From the root, we draw branches to each event B_i. Following a branch from the root corresponds to observing the result of the first subexperiment. We label the branches with the prior probabilities $P[B_1], \ldots, P[B_m]$. For each event B_i, we have conditional probabilities describing the result of the second subexperiment. Thus from each of the first set of branches, we draw a new branch and label it with the conditional probability. Following a sequence of branches from the root to a leaf (a right endpoint of the tree) specifies the result of each subexperiment. Thus the leaves represent outcomes of the complete experiment. The probability of each outcome is the product of the probabilities on the branches between the root of the tree and the leaf corresponding to the outcome. Generally, we label each leaf with the corresponding outcome and its probability.

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

√ ***Example 1.24.*** *For the resistors of Example 1.19, we have used A to denote the event that a randomly chosen resistor is "within 50 Ω of the nominal value." This could mean "acceptable." Let us use the notation N to be the complement of A: "not acceptable." 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. Sketch a sequential tree for this experiment. What is the probability of choosing a resistor from machine B_2 that is not acceptable?*

....

These two steps correspond to the following tree:



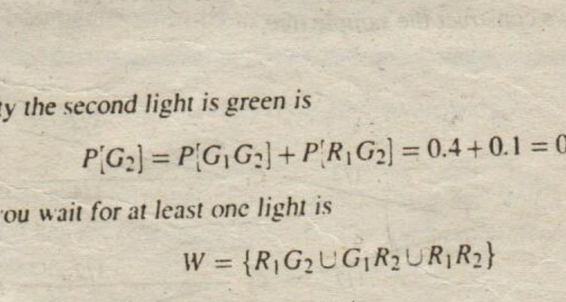
To use the tree to find the probability of the event B_2N, a nonacceptable resistor from machine B_2, we start at the left and find the probability of reaching B_2 is $P[B_2] = 0.4$. We then move to the right to B_2N and multiply $P[B_2]$ by $P[N|B_2] = 0.1$ to obtain $P[B_2N] = (0.4)(0.1) = 0.04$.

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 1.9).

Example 1.25. *Suppose 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 light? Lastly, what is $P[G_1|R_2]$, the conditional probability of a green first light given a red second light?*

....

In the case of the two-light experiment, the complete tree is



The probability the second light is green is

$$P[G_2] = P[G_1G_2] + P[R_1G_2] = 0.4 + 0.1 = 0.5$$

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

$$W = \{R_1G_2 \cup G_1R_2 \cup R_1R_2\}$$

The probability that you wait for at least one light is

$$P[W] = P[R_1G_2] + P[G_1R_2] + P[R_1R_2] = 0.1 + 0.1 + 0.4 = 0.6$$

To find $P[G_1|R_2]$, we need $P[R_2]$. Noting that $R_2 = \{G_1R_2 \cup R_1R_2\}$, we have

$$P[R_2] = P[G_1R_2] + P[R_1R_2] = 0.1 + 0.4 = 0.5$$

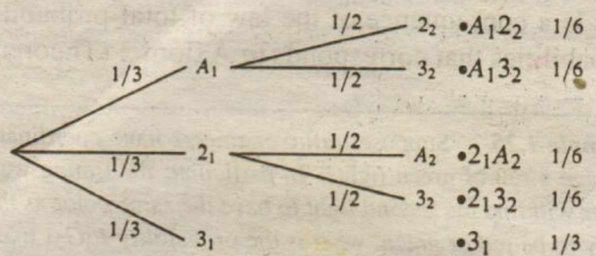
Since $P[G_1R_2] = 0.1$, the conditional probability that you have a green first light given a red second light is

$$P[G_1|R_2] = \frac{P[G_1R_2]}{P[R_2]} = \frac{0.1}{0.5} = 0.2$$

Example 1.26. *Consider the game of Three. You shuffle a deck of three cards: ace, 2, 3. With the ace worth 1 point, you draw cards until your total is 3 or more. You win if your total is 3. What is $P[W]$, the probability that you win?*

....

Let C_i denote the event that card C is the ith card drawn. For example, 3_2 is the event that the 3 was the second card drawn. The tree is



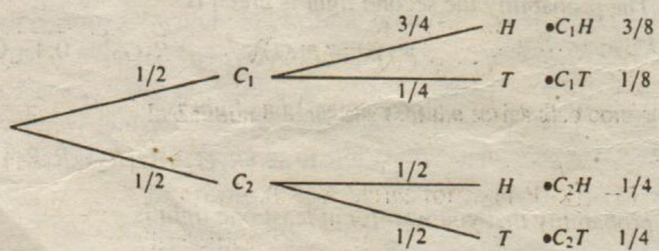
You win if A_12_2, 2_1A_2, or 3_1 occurs. Hence, the probability that you win is

$$P[W] = P[A_12_2] + P[2_1A_2] + P[3_1] = \left(\frac{1}{3}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right) + \frac{1}{3} = \frac{2}{3}$$

Example 1.27. *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 will flip 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?*

....

First, we construct the sample tree:



To find the conditional probabilities, we see

$$P[C_1|H] = \frac{P[C_1H]}{P[H]} = \frac{P[C_1H]}{P[C_1H]+P[C_2H]} = \frac{3/8}{3/8+1/4} = \frac{3}{5}$$

Similarly,

$$P[C_1|T] = \frac{P[C_1T]}{P[T]} = \frac{P[C_1T]}{P[C_1T]+P[C_2T]} = \frac{1/8}{1/8+1/4} = \frac{1}{3}$$

As we would expect, we are more likely to have chosen coin 1 when the first flip is heads but we are more likely to have chosen coin 2 when the first flip is tails.

Quiz 1.7. *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 receive the paging signal clearly. Consequently, the system will page a phone up to three times before giving up. If 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 is found?*

1.8 COUNTING METHODS

Suppose we have a shuffled full deck and we deal seven cards. What is the probability that we draw no queens? In theory, we can draw the sample space tree for the seven cards drawn. However, the resulting tree is so large, this is impractical. In short, it is too difficult to enumerate all 133 million combinations of seven cards. (In fact, you may wonder if 133 million is even approximately the number of such combinations.) To solve this problem, we need to develop procedures that permit us to count how many seven card combinations there are and how many of them do not have a queen.

The results we will derive all follow from the fundamental principle of counting:

> ***Fundamental Principle of Counting:*** If experiment A has n possible outcomes, and experiment B has k possible outcomes, then there are nk possible outcomes when you perform both experiments.

This principle is easily demonstrated by a few examples.

---

e.g.

Example 1.28. *Let A be the experiment "Flip a coin." Let B be "Roll a die." Then, A has two outcomes, H and T, and B has six outcomes, $1, \ldots, 6$. The joint experiment, called "Flip a coin and roll a die," has 12 outcomes:*

$$(H,1), \ldots, (H,6), (T,1), \ldots, (T,6)$$

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.

Do

Example 1.29. *Shuffle the deck and deal out all the cards. The outcome of the experiment is a sequence of cards of the deck. How many possible outcomes are there?*

Let subexperiment k be "Deal the kth card." The first subexperiment has 52 possible outcomes corresponding to the 52 cards that could be drawn. After the first card is drawn, the second subexperiment has 51 possible outcomes corresponding to the 51 remaining cards. The total number of outcomes is

$$52 \times 51 \times \cdots \times 1 = 52!$$

A second way to think of the experiment is to say that we will number 52 empty slots from 1 to 52. We will start with the deck in some order and we will choose a numbered slot for each card. In this case, there are 52 slots and each card is matched with a slot. The outcome of each subexperiment is a numbered slot position. There are 52 possible positions for the first card, 51 for the second card, and so on.

Example 1.30. *Shuffle the deck and choose three cards in order. How many outcomes are there?*

Ask?

In this experiment, there are 52 possible outcomes for the first card, 51 for the second card, and 50 for the third card. The total number of outcomes is $52 \times 51 \times 50$.

def

In Example 1.30, we chose an ordered sequence of three 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 to pick, $n-1$ choices for the second object, etc. Therefore, the total number of possibilities is

$$(n)_k = n(n-1)(n-2)\cdots(n-k+1)$$

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

Theorem 1.12. *The number of k-permutations of n distinguishable objects is*

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

Choosing objects from a collection is also 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 is also called *sampling without replacement*. A second type of sampling occurs when an object can be chosen repeatedly. In this case, when we remove the object from the collection, we replace the object with a duplicate. This is known as *sampling with replacement*.

Example 1.31. A laptop computer has PCMCIA expansion card slots A and B. Each slot can be filled with either a modem card (m), a SCSI interface (i), or a GPS card (g). From the set $\{m, i, g\}$ of possible cards, what is the set of possible ways to fill the two slots when we sample with replacement? In other words, how many ways can we fill the two card slots when we allow both slots to hold the same type of card?

....

In the experiment to choose with replacement a sample of two cards, let xy denote the outcome that card type x is used in slot A and card type y is used in slot B. The possible outcomes are

$$S = \{mm, mi, mg, im, ii, ig, gm, gi, gg\}$$

As we see from S, the number of possible outcomes is nine.

The fact that Example 1.31 had nine possible outcomes should not be surprising. Since we were sampling with replacement, there were always three possible outcomes for the each of the subexperiments to choose a PCMCIA card. Hence, by the fundamental theorem of counting, Example 1.31 must have $3^2 = 9$ possible outcomes.

This result generalizes naturally when we want to choose with replacement a sample of k objects out of a collection of n distinguishable objects. Sampling with replacement ensures that in each subexperiment needed to choose one of the k objects, there are n possible objects to choose. Hence there must be n^k ways to choose with replacement a sample of k objects.

*

with replacement

Theorem 1.13. *Given n distinguishable objects, there are n^k ways to choose with replacement a sample of k objects.*

Both in choosing a k-permutation or in sampling with replacement, different outcomes are distinguished by the order in which we choose objects. In Example 1.31, mi and im are distinct outcomes. However, in many practical problems, the order in which the objects were chosen makes no difference. For example, in a bridge hand, it does not matter in what order the cards are dealt. Suppose there are four objects, A, B, C, and D, and we define an experiment in which the procedure is to choose two objects, 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. What we are doing is picking a subset of the collection of objects. Each subset is called a *k-combination*. We want to find the number of k-combinations.

We will use $\binom{n}{k}$, which is read as "n choose k," to denote 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 1.12 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

1 2 3 4 (12, 13, 14, 23, 24, 34)

(12, 21) (13, 31)

Theorem 1.12, the second experiment has $(k)_k = k!$ possible outcomes. Since there are $(n)_k$ possible outcomes of the combined experiment,

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

Rearranging the terms yields our next result.

Theorem 1.14. *The number of ways to choose k objects out of n distinguishable objects is*

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!} \qquad (1.5)$$

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 form $(x+y)^n$.

Example 1.32.

- *The number of five card poker hands is*

$$\binom{52}{5} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{2 \cdot 3 \cdot 4 \cdot 5} = 2{,}598{,}960$$

- *The number of ways of picking 60 out of 120 students is* $\binom{120}{60}$.
- *The number of ways of choosing 5 starters for a basketball team with 11 players is* $\binom{11}{5} = 462$.
- *A baseball team has 15 field players and 10 pitchers. Each field player can take any of the eight nonpitching positions. Therefore, the number of possible starting lineups is* $N = \binom{10}{1}\binom{15}{8} = 64{,}350$ *since you must choose 1 of the 10 pitchers and you must choose 8 out of the 15 field players. 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* $N(9!) = 23{,}351{,}328{,}000$ *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.*

N = ( )( ) =

K = 9!

NK = —

Example 1.33. *To return to our original question of this section, suppose we draw seven cards. What is the probability of getting a hand without any queens?*

There are $H = \binom{52}{7}$ *possible hands. All H hands have probability* $1/H$. *There are* $H_{NQ} = \binom{48}{7}$ *hands that have no queens since we must choose 7 cards from a deck of 48 cards that has no queens. Since all hands are equally likely, the probability of drawing no queens is* $H_{NQ}/H = 0.5504$.

Quiz 1.8. 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?*

1.9 INDEPENDENT TRIALS

Suppose we perform the same experiment over and over. Each time, a success occurs with probability p; otherwise, a failure occurs with probability $1-p$. In addition, the result of each trial is independent of the results of all other trials. The outcome of the experiment is a sequence of successes and failures denoted by a sequence of ones and zeroes. For example, $10101\ldots$ is an alternating sequence of successes and failures. Let $S_{k,n}$ denote the event that there were k successes in n trials. To find $P[S_{k,n}]$, we consider an example.

Example 1.34. *What is the probability $P[S_{3,5}]$ of three successes in five independent trials with success probability p.*

To find $P[S_{3,5}]$, we observe that the outcomes with three successes in five trials are 11100, 11010, 11001, 10110, 10101, 10011, 01110, 01101, 01011, and 00111. Each outcome with three successes has probability $p^3(1-p)^2$. The number of such sequences is the number of ways to choose three slots out of five slots in which to place the three ones. There are $\binom{5}{3}=10$ possible ways. To find $P[S_{3,5}]$, we add up the probabilities associated with the 10 outcomes with 3 successes, yielding

$$P[S_{3,5}] = \binom{5}{3} p^3(1-p)^2$$

In general, for n independent trials we observe that

- Each outcome with k successes has probability $p^k(1-p)^{n-k}$.
- There are $\binom{n}{k}$ outcomes that have k successes.

To further confirm the second fact, note that out of n trials, there are $\binom{n}{k}$ ways to choose k of the trials to call successes. Summing over the $\binom{n}{k}$ outcomes with k successes, the probability of k successes in n independent trials is

$$P[S_{k,n}] = \binom{n}{k} p^k(1-p)^{n-k} \tag{1.6}$$

Example 1.35. *In Example 1.19, we found 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 found to be acceptable. Thus for $0 \le i \le 100$,

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

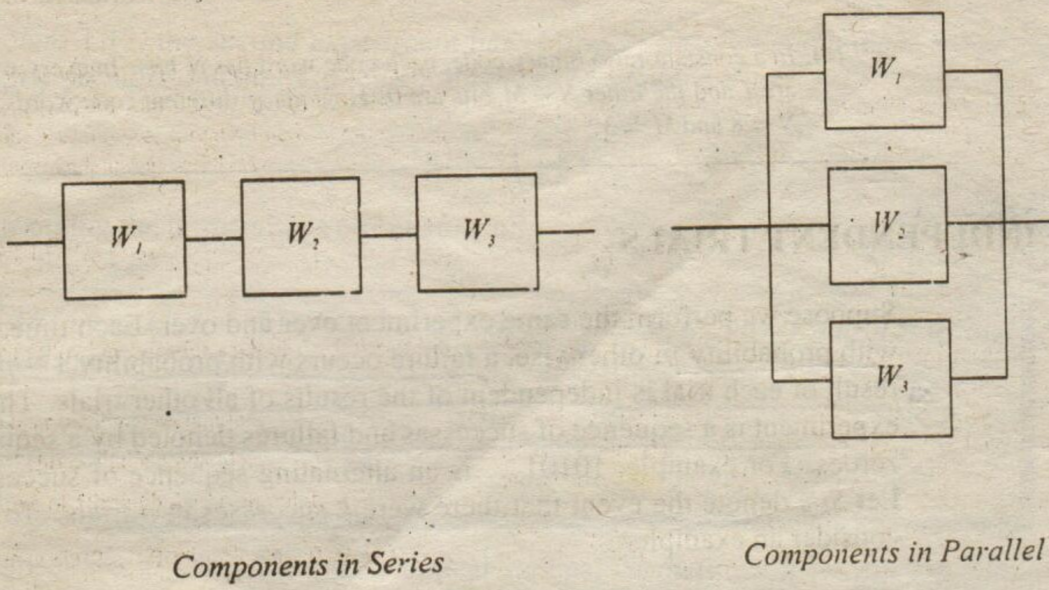


Figure 1.1 Serial and parallel devices.

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.

The next example describes how cellular phones use repeated trials to transmit data accurately.

Example 1.36. *To communicate one bit of information reliably, cellular phones transmit the same binary symbol five times. Thus the information "zero" is transmitted as 00000 and "one" is 11111. The receiver detects the correct information if three or more binary symbols are received correctly. What is the information error probability $P[E]$, if the binary symbol error probability is $q = 0.1$?*

In this case, we have five trials corresponding to the five times the binary symbol is sent. On each trial, a success occurs when a binary symbol is received correctly. The probability of a success is $p = 1 - q = 0.9$. The error event E occurs when the number of successes is strictly less than three:

$$P[E] = P[S_{0,5}] + P[S_{1,5}] + P[S_{2,5}]$$
$$= q^5 + 5pq^4 + 10p^2q^3 = 0.0081$$

By increasing the number of binary symbols per information bit from 1 to 5, the cellular phone reduces the probability of error by more than one order of magnitude from 0.1 to 0.0081.

RELIABILITY PROBLEMS

Independent trials can also be used to describe reliability problems in which we would like to calculate the probability that a particular operation succeeds. The operation

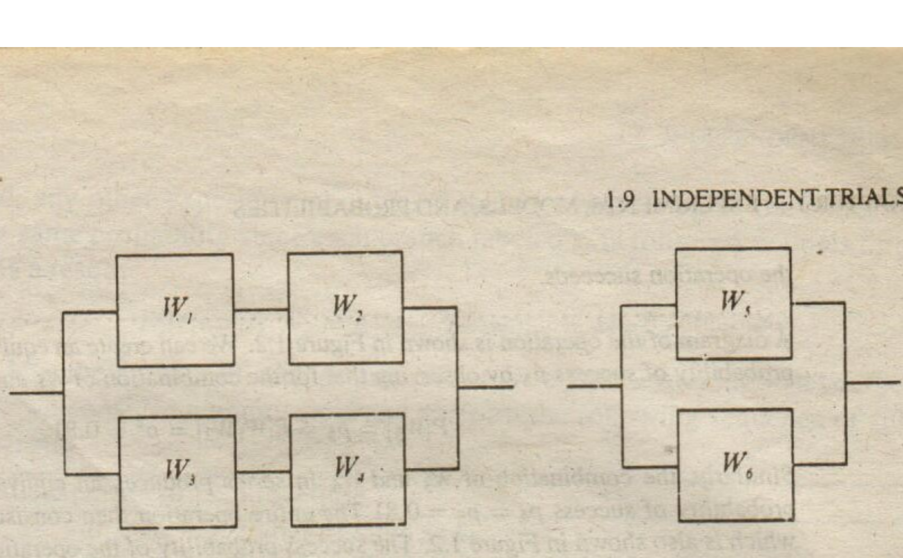


Figure 1.2 The operation described in Example 1.37. On the left is the original operation. On the right is the equivalent operation with each pair of series components replaced with an equivalent component.

consists of n components and each component succeeds with probability p, independent of any other component. Let W_i denote the event that component i succeeds. As depicted in Figure 1.1, there are two basic types of operations.

- *Components in series.* The operation succeeds if *all* of its components succeed. One example of such an operation is a sequence of computer programs, in which each program after the first one uses the result of the previous program. Therefore, the complete operation fails if any of the component programs fail. Whenever the operation consists of k components in series, we need all k components to succeed in order to have a successful operation. The probability the operation succeeds is

$$P[W] = P[W_1 W_2 \cdots W_n] = p \times p \times \cdots \times p = p^n$$

- *Components in parallel.* The operation succeeds if *any* of its components work. This operation occurs when introduce redundancy to promote reliability. In a redundant system, such as a space shuttle, there are n computers on board so that the shuttle can continue to function as long as at least one computer operates successfully. If the components are in parallel, the operation fails when all elements fail, so we have

$$P[W^c] = P[W_1^c W_2^c \cdots W_n^c] = (1-p)^n$$

The probability that the parallel operation succeeds is

$$P[W] = 1 - P[W^c] = 1 - (1-p)^n$$

We can analyze complicated combinations of components in series and in parallel by reducing several components in parallel or components in series to a single equivalent component.

Example 1.37. *An operation consists of two redundant parts. The first part has two components in series (W_1 and W_2) and the second part has two components in series (W_3 and W_4). All components succeed with probability $p = 0.9$. Draw a diagram of the operation and calculate the probability that* *the operation succeeds.*

A diagram of the operation is shown in Figure 1.2. We can create an equivalent component, W_5, with probability of success p_5 by observing that for the combination of W_1 and W_2,

$$P[W_5] = p_5 = P[W_1 W_2] = p^2 = 0.81$$

Similarly, the combination of W_3 and W_4 in series produces an equivalent component, W_6, with probability of success $p_6 = p_5 = 0.81$. The entire operation then consists of W_5 and W_6 in parallel which is also shown in Figure 1.2. The success probability of the operation is

$$P[W] = 1 - (1 - p_5)^2 = 0.964$$

We could consider the combination of W_5 and W_6 to be an equivalent component W_7 with success probability $p_7 = 0.964$ and then analyze a more complex operation that contains W_7 as a component.

Working on these reliability problems leads us to the observation that in calculating probabilities of events involving independent trials, it is easy to find the probability of an intersection and difficult to find directly the probability of a union. Specifically, for the device with components in series, it is difficult to calculate directly the probability that the device fails. Similarly, when the components are in parallel, calculating the probability the device works is hard. However, De Morgan's law (Theorem 1.1) allows us to express a union as the complement of an intersection and vice versa. Therefore when it is difficult to calculate directly the probability we need, we can often calculate the probability of the complementary event first and then subtract this probability from one to find the answer. This is how we calculated the probability that the parallel device works.

MULTIPLE OUTCOMES

Suppose we perform n independent repetitions of a subexperiment for which there are r possible outcomes for any trial. That is, the sample space for each trial is $(s_1, \ldots, s_r)$ and that on any trial, $P[s_k] = p_k$, independent of the result of any other trial.

An outcome of the experiment consists of a sequence of n trial outcomes. Consider the experiment as following a probability tree where on the ith branch, we choose the branch labeled s_i with probability p_i. The probability of an experimental outcome is just the product of the branch probabilities. For example, the experimental outcome $s_1 s_1 s_3 s_4 s_5$ occurs with probability $p_1 p_1 p_3 p_4 p_5$. Let N_i denote the number of times that outcome s_i occurs out of the n trials. We want to find

$$P[N_1 = n_1, N_2 = n_2, \ldots, N_r = n_r]$$

First, the probability of the outcome

$$\underbrace{s_1 \cdots s_1}_{n_1 \text{ times}} \underbrace{s_2 \cdots s_2}_{n_2 \text{ times}} \cdots \underbrace{s_r \cdots s_r}_{n_r \text{ times}}$$

is

$$p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

Second, any other experimental outcome that is a reordering of the above sequence has the same probability since each branch labeled s_i is followed n_i times through the tree. As a result,

$$P[N_1 = n_1, N_2 = n_2, \cdots, N_r = n_r] = M p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

where M is the number of such sequences. M is called the multinomial coefficient. To find M, we consider n empty slots and perform the following sequence of subexperiments:

subexperiment	Procedure
1	Label n_1 slots as s_1.
2	Label n_2 slots as s_2.
$\vdots$	$\vdots$
r	Label remaining $n_r = n - (n_1 + \cdots + n_{r-1})$ slots as s_r.

There are $\binom{n}{n_1}$ ways to perform the first subexperiment. Similarly, there are $\binom{n-n_1}{n_2}$ ways to perform the second subexperiment. After $j-1$ subexperiments, $n_1 + \cdots + n_{j-1}$ slots have already been filled, leaving $\binom{n-(n_1+\cdots+n_{j-1})}{n_j}$ ways to perform the jth subexperiment. From our basic counting principle,

$$\begin{aligned} M &= \binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-(n_1+n_2)}{n_3}\cdots\binom{n_r}{n_r} \\ &= \frac{n!}{(n-n_1)!n_1!}\frac{(n-n_1)!}{(n-n_1-n_2)!n_2!}\cdots\frac{(n-(n_1+\cdots n_{r-1}))!}{(n-(n_1+\cdots n_r))!n_r!} \end{aligned}$$

If we cancel the common factors, we have

$$M = \frac{n!}{n_1!n_2!\cdots n_r!}$$

so that

$$\begin{aligned} &P[N_1 = n_1, \cdots N_r = n_r] \\ &= \begin{cases} \frac{n!}{n_1!\cdots n_r!} p_1^{n_1} \cdots p_r^{n_r} & n_1 + \cdots + n_r = n;\ n_i \ge 0, i = 1, \ldots, r \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Example 1.38. *Each call arriving at a telephone switch is independently either a voice call with probability 7/10, a fax call with probability 2/10, or a modem call with probability 1/10. Let X, Y, and Z denote the number of voice, fax, and modem calls out of 100 observed calls. In this case,*

$$\begin{aligned} &P[X = x, Y = y, Z = z] \\ &= \begin{cases} \frac{100!}{x!y!z!}\left(\frac{7}{10}\right)^x\left(\frac{2}{10}\right)^y\left(\frac{1}{10}\right)^z & x+y+z = 100, x \ge 0, y \ge 0, z \ge 0 \\ 0 & \textit{otherwise} \end{cases} \end{aligned}$$

Quiz 1.9. *A memory module consists of 9 chips. The device is designed with redundancy so that it works even if one of its chips is defective. Each chip contains n transistors and functions properly if all of its transistors work. A transistor works with probability p independent of any other transistor.*

(a) *What is the probability $P[C]$ that a chip works?*

(b) *What is the probability $P[M]$ that the memory module works?*

CHAPTER SUMMARY

This chapter introduces the model of an experiment consisting of a procedure and observations. Outcomes of the experiment are elements of a sample space. Probability is a number assigned to every set in the sample space. Three axioms contain the fundamental properties of probability. The rest of this book uses these axioms to develop methods of working on practical problems.

1. *Sample space*, *event*, and *outcome* are probability terms for the set theory concepts of universal set, set, and element.
2. A *probability measure* $P[A]$ is a function that assigns a number between 0 and 1 to every event A in the sample space. The assigned probabilities conform to the three axioms presented in Section 1.3.
3. A *conditional probability* $P[A|B]$ describes the likelihood of A given that B has occurred. The conditional probability $P[A|B]$ also satisfies the three axioms of probability.
4. *Tree diagrams* illustrate experiments that consist of a sequence of steps. The labels on the tree branches can be used to calculate the probabilities of outcomes of the combined experiment.
5. *Counting methods* determine the number of outcomes of complicated experiments.

PROBLEMS

1.2.1. ● A fax transmission can take place at any of three speeds depending on the condition of the phone connection between the two fax machines. The speeds are high (h) at 14,400 b/s, medium (m) at 9600 b/s, and low (l) at 4800 b/s. In response to requests for information a company sends either short faxes of two (t) pages, or long faxes of four (f) pages. Consider the experiment of monitoring a fax transmission and observing the transmission speed and the length. An observation is a two-letter word, for example, a high-speed, two-page fax is ht.

(a) What is the sample space of the experiment?

(b) Let A_1 be the event "medium speed fax." What are the outcomes in A_1?

(c) Let A_2 be the event "short (two-page) fax." What are the outcomes in A_2?

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

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

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

1.2.2. ● 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 machine 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?

1.2.3. ● 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?

1.2.4. ● 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?

1.2.5. ● Let the sample space of an experiment consist of all the undergraduates at a university. Give four examples of event spaces.

1.2.6. ● Let the sample space of the experiment consist of the measured resistances of two resistors. Give four examples of event spaces.

1.3.1. ● 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 ≤ 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:

(a) $P[W]$

(b) $P[B]$

(c) $P[W \cup B]$

1.3.2. ● 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? Calculate the following probabilities:

(a) $P[W]$

(b) $P[MF]$

(c) $P[H]$

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

1.3.4. ● 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?

1.3.5. ● 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?

1.4.1. ● Cellular telephones perform *handoffs* as they move from cell to cell. During a call, 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 3 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

What is the probability $P[H_0]$ that a phone makes no handoffs? What is the probability a call is brief? 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 1.14, 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 9 minutes or less?

1.4.3. ■ 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. Each gene is either dominant d or recessive r. Mendel's experiment is to select a plant and observe whether the genes are both dominant d, both recessive, r, or one of each (hybrid) h. 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; while 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?

1.4.4. ■ Use Theorem 1.7 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) $P[A \cap B] \leq P[B]$

1.4.5. ♦ 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 above 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.

1.4.6. ♦ Using *only* the three axioms of probability, prove $P[\phi] = 0$.

1.4.7. ♦ Using the three axioms of probability and the fact that $P[\phi] = 0$, prove Theorem 1.4. Hint: Define $A_i = B_i$ for $i = 1, \ldots, m$ and $A_i = \phi$ for $i > m$.

1.4.8. ♦♦ For each fact stated in Theorem 1.7, determine which of the three axioms of probability are needed to prove the fact.

1.5.1. ● Given the model of handoffs and call length in Problem 1.4.1,

(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?

1.5.2. ● 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?

1.5.3. ● You have a shuffled deck of three clubs: 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 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?

1.5.4. ■ From Problem 1.4.3, what is the conditional probability of yy, that a pea plant has two dominant genes given the event Y that it has yellow seeds?

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

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

1.5.6. ♦ 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?

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

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

1.6.3. ■ In an experiment, A, B, C and D are events with probabilities $P[A] = 1/4$, $P[B] = 1/8$, $P[C] = 5/8$ and $P[D] = 3/8$. Furthermore, A and B are disjoint while C and D are independent.

(a) What is $P[A \cap B]$?

(b) What is $P[A \cup B]$?

(c) What is $P[A \cap B^c]$?

(d) What is $P[A \cup B^c]$?

(e) Are A and B independent?

(f) What is $P[C \cap D]$?

(g) What is $P[C \cup D]$?

(h) What is $P[C|D]$?

(i) What is $P[C \cap D^c]$?

(j) What is $P[C \cup D^c]$?

(k) What is $P[C^c \cap D^c]$?

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

1.6.4. ■ In an experiment, A, B, C, and D are events with probabilities $P[A \cup B] = 5/8$. $P[A] = 3/8$, $P[C \cap D] = 1/3$ and $P[C] = 1/2$. Furthermore, A and B are disjoint, while C and D are independent.

(a) What is $P[A \cap B]$? 0

(b) What is $P[B]$? 1/4

(c) What is $P[A \cap B^c]$? 3/8

(d) What is $P[A \cup B^c]$? 3/4

(e) Are A and B independent? No

(f) What is $P[D]$? 2/3

(g) What is $P[C \cup D]$? 5/6

(h) What is $P[C|D]$? 1/2

(i) What is $P[C \cap D^c]$? 1/6

(j) What is $P[C \cup D^c]$? 2/3

(k) What is $P[C^c \cap D^c]$? 1/6

(l) Are C and D^c independent? Yes

1.6.5. ■ In an experiment with equiprobable outcomes, the event 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).

1.6.6. ■ (Continuation of Problem 1.4.3) 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 are a dominant trait over wrinkled peas. 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?

1.6.7. ♦ 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.

1.6.8. ♦ 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.

1.6.9. ♦ For a Venn diagram in which the event areas are proportional to their probabilities to illustrate three events A, B, and C that are pairwise independent but not independent.

1.7.1. ● Suppose you flip a coin twice. On any flip, the coin comes up heads with probability $1/4$. Use H_i and T_i 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? 3/16

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

1.7.2. ● For Example 1.25, 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 $P[G_1|G_2]$. 1/2 3/4

1.7.3. ● 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? 1/4

1.7.4. ● 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. Let events A and B designate which coin was picked. What is the probability $P[C]$ that your guess is correct? 3/4

1.7.5. ■ 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? 0.0194

1.7.6. ■ 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 $4/5$. Otherwise, if the first photo detector is defective, the second photo detector is acceptable with probability $2/5$.

(a) What is the probability that exactly one photo detector of a pair is acceptable? 7/25

(b) What is the probability that both photo detectors in a pair are defective? 6/25

1.7.7. ■ 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 where the first coin you flip is chosen 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 $P[H_1H_2]$? Are H_1 and H_2 independent? Please explain your answer.

1.7.8. ■ Suppose Dagwood (Blondie's husband) wants to eat a sandwich but needs to go on a diet. So, Dagwood decides to let the flip of the coin determine whether he eats. Using an unbiased coin, Dagwood will postpone the diet (and go directly to the refrigerator) if either (a) he flips heads on his first flip or (b) he flips tails on the first flip but then proceeds to get two heads out of the next three flips. Note that the first flip is *not* counted in the attempt to win two of three and that Dagwood never performs any unnecessary flips. Let H_i be the event that Dagwood flips heads on try i. Let T_i be the event that tails occurs on flip i.

(a) Sketch the tree for this experiment. Please label the probabilities of all outcomes carefully.

(b) What are $P[H_3]$ and $P[T_3]$?

(c) Let D be the event that Dagwood must diet. What is $P[D]$? What is $P[H_1|D]$?

(d) Are H_3 and H_2 independent events?

1.7.9. ■ The quality of each pair of diodes produced by the machine in Problem 1.7.6 is independent of the quality of every other pair of diodes.

(a) What is the probability of finding no good diodes in a collection of n pairs produced by the machine?

(b) How many pairs of diodes must the machine produce to reach a probability of 0.99 that there will be at least one acceptable diode?

1.7.10. ■ Each time a fisherman casts his line, a fish is caught with probability p, independent of whether a fish is caught on any other cast of a line. The fisherman will fish all day until a fish is caught and then he quits and goes home. Let C_i denote the event that on cast i the fisherman catches a fish. Draw the tree for this experiment and find the following probabilities:

(a) $P[C_1]$

(b) $P[C_2]$

(c) $P[C_n]$

1.8.1. ● Consider a binary code with 5 bits (0 or 1) in each code word. An example of a code word is 01010. How many different code words are there? How many code words have exactly three 0's?

1.8.2. ● Consider a language containing four letters: A, B, C, D. How many three-letter words can you form in this language? How many four-letter words can you form if each letter only appears once in each word?

can repeat

1.8.3. ■ Shuffle a deck of cards and pick two cards at random. Observe the sequence of the two cards in the order in which they were chosen.

(a) How many outcomes are in the sample space? ${}^nP_r$

(b) How many outcomes are in the event that the two cards are the same type but different suits?

(c) What is the probability that the two cards are the same type but different suits?

(d) Suppose the experiment specifies observing the set of two cards regardless without considering the order in which they are selected and redo parts (a)–(c).

1.8.4. ● On an American League baseball team with 15 field players and 10 pitchers, the manager must select for the starting lineup, 8 field players, 1 pitcher, and 1 designated hitter. A starting lineup specifies the players for these positions and the positions in a batting order for the 8 field players and designated hitter. If the designated hitter must be chosen among all the field players, how many possible starting lineups are there?

1.8.5. ■ Suppose that in Problem 1.8.4, the designated hitter can be chosen from among all the players. How many possible starting lineups are there?

1.8.6. ■ A basketball team has three pure centers, four pure forwards, four pure guards and one swingman who can play either guard or forward. A "pure" position player can play only the designated position. If the coach must start a lineup with one center, two forwards and two guards, how many possible lineups can the coach choose?

1.8.7. ♦ An instant lottery ticket consists of a collection of boxes covered with gray wax. For a subset of the boxes, the gray wax hides a special mark. If a player scratches off the correct number of the marked boxes (and no boxes without the mark), then that ticket is a winner. Design an instant lottery game in which a player scratches five boxes and the probability that a ticket is a winner is approximately 0.01.

1.9.1. ● Consider a binary code with 5 bits (0 or 1) in each code word. An example of a code word is 01010. In each code word, a bit is a zero with probability 0.8, independent of any other bit.

(a) What is the probability of the code word 00111?

(b) What is the probability that a code word contains exactly three ones?

1.9.2. ● The Boston Celtics have won 16 NBA championships over approximately 50 years. Thus it may seem reasonable to assume that in a given year the Celtics win the title with probability $p = 0.32$, independent of any other year. Given such a model, what would be the probability of the Celtics winning eight straight championships beginning in 1959? Also, what would be the probability of the Celtics winning the title in 10 out of 11 years, starting in 1959? Given your answers, do you trust this simple probability model?

1.9.3. ● A better model for traffic lights than that given in Example 1.8 would include the effect of yellow lights. Suppose each day that you drive to work a traffic light that you encounter is either green with probability 7/16, red with probability 7/16, or yellow with probability 1/8, independent of the status of the light on any other day. If over the course of five days, G, Y, and R denote the number of times the light is found to be green, yellow, or red, respectively, what is the probability that $P[G=2,Y=1,R=2]$? Also, what is the probability $P[G=R]$?

1.9.4. ■ Suppose a 10-digit phone number is transmitted by a cellular phone using four binary symbols for each digit using the model of binary symbol errors and deletions given in Problem 1.9.6. If C denotes the number of bits sent correctly, D the number of deletions, and E the number of errors, what is $P[C=c,D=d,E=e]$? Your answer should be correct for any choice of c, d, and e.

1.9.5. ■ A particular operation has six components. Each component has a failure probability q, independent of any other component. The operation is successful if both

- Components 1, 2, and 3 all work or component 4 works.
- Either component 5 or component 6 works.

Sketch a block diagram for this operation similar to those of Figure 1.1 on page 32. What is the probability $P[W]$ that the operation is successful?

1.9.6. ■ We wish to modify the cellular telephone coding system in Example 1.36 in order to reduce the number of errors. In particular, if there are two or three zeroes in the received sequence of 5 bits, we will say that a deletion (event D) occurs. Otherwise, if at least 4 zeroes are received, then the receiver decides a zero was sent. Similarly, if at least 4 ones are received, then the receiver decides a one was sent. We say that an error occurs if either a one was sent and the receiver decides zero was sent or if a zero was sent and the receiver decides a one was sent. For this modified protocol, what is the probability $P[E]$ of an error? What is the probability $P[D]$ of a deletion?

1.9.7. ■ In a game between two equal teams, the home team wins any game with probability $p > 1/2$. In a best of three playoff series, a team with the home advantage has a game at home, followed by a game away, followed by a home game if necessary. The series is over as soon as one team wins two games. What is $P[H]$, the probability that the team with the home advantage wins the series? Is the home advantage increased by playing a three game series rather than one game playoff? That is, is it true that $P[H] \geq p$ for all $p \geq 1/2$?

1.9.8. ◆ Consider the device described in Problem 1.9.5. Suppose we can replace any one of the components by an ultrareliable component that has a failure probability of $q/2$. Which component should we replace?

1.9.9. ◆ There is a collection of field goal kickers, which can be divided into two groups 1 and 2. Group i has $3i$ shooters. On any kick, a kicker from group i will kick a field goal with probability $1/(i+1)$, independent of the outcome of any other kicks by that kicker or any other kicker.

(a) A kicker is selected at random from among all the kickers and attempts one field goal. Let K be the event that the a field goal is kicked. Find $P[K]$.

(b) Two kickers are selected at random. For $j=1,2$, let K_j be the event that kicker j kicks a field goal. Find $P[K_1 \cap K_2]$. Are K_1 and K_2 independent events?

(c) A kicker is selected at random and attempts 10 field goals. Let M be the number of misses. Find $P[M=5]$.