Chapter 2. Probability

2.1 Introduction
2.2 Sample Spaces and the Algebra of Sets
2.3 The Probability Function
2.4 Conditional Probability
2.5 Independence
2.6 Combinatorics
2.7 Combinatorial Probability
2.8 Taking a Second Look at Statistics (Monte Carlo Techniques)

One of the most influential of seventeenth-century mathematicians, Fermat earned his living as a lawyer and administrator in Toulouse. He shares credit with Descartes for the invention of analytic geometry, but his most important work may have been in number theory. Fermat did not write for publication, preferring instead to send letters and papers to friends. His correspondence with Pascal was the starting point for the development of a mathematical theory of probability.

—Pierre de Fermat (1601–1665)

Pascal was the son of a nobleman. A prodigy of sorts, he had already published a treatise on conic sections by the age of sixteen. He also invented one of the early calculating machines to help his father with accounting work. Pascal’s contributions to probability were stimulated by his correspondence, in 1654, with Fermat. Later that year he retired to a life of religious meditation.

—Blaise Pascal (1623–1662)

2.1 Introduction

Experts have estimated that the likelihood of any given UFO sighting being genuine is on the order of one in one hundred thousand. Since the early 1950s, some ten thousand sightings have been reported to civil authorities. What is the probability that at least one of those objects was, in fact, an alien spacecraft? In 1978, Pete Rose of the Cincinnati Reds set a National League record by batting safely in forty-four consecutive games. How unlikely was that event, given that Rose was a lifetime .303 hitter? By definition, the mean free path is the average distance a molecule in a gas travels before colliding with another molecule. How likely is it that the distance a molecule travels between collisions will be at least twice its mean free path? Suppose a boy’s mother and father both have genetic markers for sickle cell anemia, but neither parent exhibits any of the disease’s symptoms. What are the chances that their son will also be asymptomatic? What are the odds that a poker player is dealt
a full house or that a craps-shooter makes his “point”? If a woman has lived to age seventy, how likely is it that she will die before her ninetieth birthday? In 1994, Tom Foley was Speaker of the House and running for re-election. The day after the election, his race had still not been “called” by any of the networks: he trailed his Republican challenger by 2174 votes, but 14,000 absentee ballots remained to be counted. Foley, however, conceded. Should he have waited for the absentee ballots to be counted, or was his defeat at that point a virtual certainty?

As the nature and variety of these questions would suggest, probability is a subject with an extraordinary range of real-world, everyday applications. What began as an exercise in understanding games of chance has proven to be useful everywhere. Maybe even more remarkable is the fact that the solutions to all of these diverse questions are rooted in just a handful of definitions and theorems. Those results, together with the problem-solving techniques they empower, are the sum and substance of Chapter 2. We begin, though, with a bit of history.

The Evolution of the Definition of Probability

Over the years, the definition of probability has undergone several revisions. There is nothing contradictory in the multiple definitions—the changes primarily reflected the need for greater generality and more mathematical rigor. The first formulation (often referred to as the classical definition of probability) is credited to Gerolamo Cardano (recall Section 1.3). It applies only to situations where (1) the number of possible outcomes is finite and (2) all outcomes are equally likely. Under those conditions, the probability of an event comprised of \( m \) outcomes is the ratio \( m/n \), where \( n \) is the total number of (equally likely) outcomes. Tossing a fair, six-sided die, for example, gives \( m/n = \frac{3}{6} \) as the probability of rolling an even number (that is, either 2, 4, or 6).

While Cardano’s model was well-suited to gambling scenarios (for which it was intended), it was obviously inadequate for more general problems, where outcomes are not equally likely and/or the number of outcomes is not finite. Richard von Mises, a twentieth-century German mathematician, is often credited with avoiding the weaknesses in Cardano’s model by defining “empirical” probabilities. In the von Mises approach, we imagine an experiment being repeated over and over again under presumably identical conditions. Theoretically, a running tally could be kept of the number of times \( m \) the outcome belongs to a given event divided by \( n \), the total number of times the experiment is performed. According to von Mises, the probability of the given event is the limit (as \( n \) goes to infinity) of the ratio \( m/n \). Figure 2.1.1 illustrates the empirical probability of getting a head by tossing a fair coin: as the number of tosses continues to increase, the ratio \( m/n \) converges to 1/2.
The von Mises approach definitely shores up some of the inadequacies seen in the Cardano model, but it is not without shortcomings of its own. There is some conceptual inconsistency, for example, in extolling the limit of $m/n$ as a way of defining a probability empirically, when the very act of repeating an experiment under identical conditions an infinite number of times is physically impossible. And left unanswered is the question of how large $n$ must be in order for $m/n$ to be a good approximation for $\lim m/n$.

Andrei Kolmogorov, the great Russian probabilist, took a different approach. Aware that many twentieth-century mathematicians were having success developing subjects axiomatically, Kolmogorov wondered whether probability might similarly be defined operationally, rather than as a ratio (like the Cardano model) or as a limit (like the von Mises model). His efforts culminated in a masterpiece of mathematical elegance when he published Grundbegriffe der Wahrscheinlichkeitsrechnung (Foundations of the Theory of Probability) in 1933. In essence, Kolmogorov was able to show that a maximum of four simple axioms is necessary and sufficient to define the way any and all probabilities must behave. (These will be our starting point in Section 2.3.)

We begin Chapter 2 with some basic (and, presumably, familiar) definitions from set theory. These are important because probability will eventually be defined as a set function—that is, a mapping from a set to a number. Then, with the help of Kolmogorov’s axioms in Section 2.3, we will learn how to calculate and manipulate probabilities. The chapter concludes with an introduction to combinatorics—the mathematics of systematic counting—and its application to probability.

2.2 Sample Spaces and the Algebra of Sets

The starting point for studying probability is the definition of four key terms: experiment, sample outcome, sample space, and event. The latter three, all carryovers from classical set theory, give us a familiar mathematical framework within which to work; the former is what provides the conceptual mechanism for casting real-world phenomena into probabilistic terms.

By an experiment we will mean any procedure that (1) can be repeated, theoretically, an infinite number of times; and (2) has a well-defined set of possible outcomes. Thus, rolling a pair of dice qualifies as an experiment; so does measuring a hypertensive’s blood pressure or doing a spectrographic analysis to determine the carbon content of moon rocks. Asking a would-be psychic to draw a picture of an image presumably transmitted by another would-be psychic does not qualify as an experiment, because the set of possible outcomes cannot be listed, characterized, or otherwise defined.

Each of the potential eventualities of an experiment is referred to as a sample outcome, $s$, and their totality is called the sample space, $S$. To signify the membership of $s$ in $S$, we write $s \in S$. Any designated collection of sample outcomes, including individual outcomes, the entire sample space, and the null set, constitutes an event. The latter is said to occur if the outcome of the experiment is one of the members of the event.

Consider the experiment of flipping a coin three times. What is the sample space? Which sample outcomes make up the event $A$: Majority of coins show heads? Think of each sample outcome here as an ordered triple, its components representing the outcomes of the first, second, and third tosses, respectively. Altogether,
there are eight different triples, so those eight comprise the sample space:

\[ S = \{ \text{HHH}, \text{HHT}, \text{HTH}, \text{THH}, \text{HTT}, \text{THT}, \text{TTT} \} \]

By inspection, we see that four of the sample outcomes in \( S \) constitute the event \( A \):

\[ A = \{ \text{HHH}, \text{HHT}, \text{HTH}, \text{THH} \} \]

Imagine rolling two dice, the first one red, the second one green. Each sample outcome is an ordered pair (face showing on red die, face showing on green die), and the entire sample space can be represented as a 6 \( \times \) 6 matrix (see Figure 2.2.1).

**Example 2.2.2**

![Figure 2.2.1](image)

Gamblers are often interested in the event \( A \) that the sum of the faces showing is a 7. Notice in Figure 2.2.1 that the sample outcomes contained in \( A \) are the six diagonal entries, \((1, 6), (2, 5), (3, 4), (4, 3), (5, 2), \) and \((6, 1)\).

**Example 2.2.3**

A local TV station advertises two newscasting positions. If three women (\( W_1, W_2, W_3 \)) and two men (\( M_1, M_2 \)) apply, the “experiment” of hiring two coanchors generates a sample space of ten outcomes:

\[ S = \{ (W_1, W_2), (W_1, W_3), (W_2, W_3), (W_1, M_1), (W_1, M_2), (W_2, M_1), (W_2, M_2), (W_3, M_1), (W_3, M_2), (M_1, M_2) \} \]

Does it matter here that the two positions being filled are equivalent? Yes. If the station were seeking to hire, say, a sports announcer and a weather forecaster, the number of possible outcomes would be twenty: \((W_2, M_1)\), for example, would represent a different staffing assignment than \((M_1, W_2)\).

**Example 2.2.4**

The number of sample outcomes associated with an experiment need not be finite. Suppose that a coin is tossed until the first tail appears. If the first toss is itself a tail, the outcome of the experiment is \( T \); if the first tail occurs on the second toss, the outcome is \( HT \); and so on. Theoretically, of course, the first tail may never occur, and the infinite nature of \( S \) is readily apparent:

\[ S = \{ T, HT, HHT, HHHT, \ldots \} \]

**Example 2.2.5**

There are three ways to indicate an experiment’s sample space. If the number of possible outcomes is small, we can simply list them, as we did in Examples 2.2.1 through 2.2.3. In some cases it may be possible to characterize a sample space by showing the structure its outcomes necessarily possess. This is what we did in Example 2.2.4.
A third option is to state a mathematical formula that the sample outcomes must satisfy.

A computer programmer is running a subroutine that solves a general quadratic equation, \( ax^2 + bx + c = 0 \). Her “experiment” consists of choosing values for the three coefficients \( a, b, \) and \( c \). Define (1) \( S \) and (2) the event \( A \): Equation has two equal roots.

First, we must determine the sample space. Since presumably no combinations of finite \( a, b, \) and \( c \) are inadmissible, we can characterize \( S \) by writing a series of inequalities:

\[
S = \{(a, b, c) : -\infty < a < \infty, -\infty < b < \infty, -\infty < c < \infty\}
\]

Defining \( A \) requires the well-known result from algebra that a quadratic equation has equal roots if and only if its discriminant, \( b^2 - 4ac \), vanishes. Membership in \( A \), then, is contingent on \( a, b, \) and \( c \) satisfying an equation:

\[
A = \{(a, b, c) : b^2 - 4ac = 0\}
\]

Questions

2.2.1. A graduating engineer has signed up for three job interviews. She intends to categorize each one as being either a “success” or a “failure” depending on whether it leads to a plant trip. Write out the appropriate sample space. What outcomes are in the event \( A \): Second success occurs on third interview? In \( B \): First success never occurs? (Hint: Notice the similarity between this situation and the coin-tossing experiment described in Example 2.2.1.)

2.2.2. Three dice are tossed, one red, one blue, and one green. What outcomes make up the event \( A \) that the sum of the three faces showing equals 5?

2.2.3. An urn contains six chips numbered 1 through 6. Three are drawn out. What outcomes are in the event "Second smallest chip is a 3"? Assume that the order of the chips is irrelevant.

2.2.4. Suppose that two cards are dealt from a standard 52-card poker deck. Let \( A \) be the event that the sum of the two cards is 8 (assume that aces have a numerical value of 1). How many outcomes are in \( A \)?

2.2.5. In the lingo of craps-shooters (where two dice are tossed and the underlying sample space is the matrix pictured in Figure 2.2.1) is the phrase "making a hard eight." What might that mean?

2.2.6. A poker deck consists of fifty-two cards, representing thirteen denominations (2 through ace) and four suits (diamonds, hearts, clubs, and spades). A five-card hand is called a flush if all five cards are in the same suit but not all five denominations are consecutive. Pictured in the next column is a flush in hearts. Let \( N \) be the set of five cards in hearts that are not flushes. How many outcomes are in \( N \)? [Note: In poker, the denominations (A, 2, 3, 4, 5) are considered to be consecutive (in addition to sequences such as (8, 9, 10, J, Q)).]

<table>
<thead>
<tr>
<th>Denominations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 3 4 5 6 7 8 9 10 J Q K A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Suits</th>
<th>H X X</th>
<th>X</th>
<th>X X</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>C</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.2.7. Let \( P \) be the set of right triangles with a 5' hypotenuse and whose height and length are \( a \) and \( b \), respectively. Characterize the outcomes in \( P \).

2.2.8. Suppose a baseball player steps to the plate with the intention of trying to "coax" a base on balls by never swinging at a pitch. The umpire, of course, will necessarily call each pitch either a ball (\( B \)) or a strike (\( S \)). What outcomes make up the event \( A \), that a batter walks on the sixth pitch? (Note: A batter "walks" if the fourth ball is called before the third strike.)

2.2.9. A telemarketer is planning to set up a phone bank to bilk widows with a Ponzi scheme. His past experience (prior to his most recent incarceration) suggests that each phone will be in use half the time. For a given phone at a given time, let 0 indicate that the phone is available and let 1 indicate that a caller is on the line. Suppose that the telemarketer’s "bank" is comprised of four telephones.
Write out the outcomes in the sample space. What outcomes would make up the event that exactly two phones are being used? Suppose the telemarketer had \( k \) phones. How many outcomes would allow for the possibility that at most one more call could be received? (Hint: How many lines would have to be busy?)

2.2.10. Two darts are thrown at the following target:

Let \((u, v)\) denote the outcome that the first dart lands in region \( u \) and the second dart, in region \( v \). List the sample space of \((u, v)\)'s. List the outcomes in the sample space of \( u + v \).

2.2.11. A woman has her purse snatched by two teenagers. She is subsequently shown a police lineup consisting of five suspects, including the two perpetrators. What is the sample space associated with the experiment "Woman picks two suspects out of lineup"? Which outcomes are in the event \( A \): She makes at least one incorrect identification?

2.2.12. Consider the experiment of choosing coefficients for the quadratic equation \( ax^2 + bx + c = 0 \). Characterize the values of \( a, b, \) and \( c \) associated with the event \( A \): Equation has complex roots.

2.2.13. In the game of craps, the person rolling the dice (the shooter) wins outright if his first toss is a 7 or an 11. If his first toss is a 2, 3, or 12, he loses outright. If his first roll is something else, say, a 9, that number becomes his "point" and he keeps rolling the dice until he either rolls another 9, in which case he wins, or a 7, in which case he loses. Characterize the sample outcomes contained in the event "Shooter wins with a point of 9."

2.2.14. A probability-minded despot offers a convicted murderer a final chance to gain his release. The prisoner is given twenty chips, ten white and ten black. All twenty are to be placed into two urns, according to any allocation scheme the prisoner wishes, with the one proviso being that each urn contain at least one chip. The executioner will then pick one of the two urns at random and from that urn, one chip at random. If the chip selected is white, the prisoner will be set free; if it is black, he "buys the farm." Characterize the sample space describing the prisoner's possible allocation options. (Intuitively, which allocation affords the prisoner the greatest chance of survival?)

2.2.15. Suppose that ten chips, numbered 1 through 10, are put into an urn at one minute to midnight, and chip number 1 is quickly removed. At one-half minute to midnight, chips numbered 11 through 20 are added to the urn, and chip number 2 is quickly removed. Then at one-fourth minute to midnight, chips numbered 21 to 30 are added to the urn, and chip number 3 is quickly removed. If that procedure for adding chips to the urn continues, how many chips will be in the urn at midnight (148)?

Unions, Intersections, and Complements

Associated with events defined on a sample space are several operations collectively referred to as the algebra of sets. These are the rules that govern the ways in which one event can be combined with another. Consider, for example, the game of craps described in Question 2.2.13. The shooter wins on his initial roll if he throws either a 7 or an 11. In the language of the algebra of sets, the event "Shooter rolls a 7 or an 11" is the union of two simpler events, "Shooter rolls a 7" and "Shooter rolls an 11." If \( E \) denotes the union and if \( A \) and \( B \) denote the two events making up the union, we write \( E = A \cup B \). The next several definitions and examples illustrate those portions of the algebra of sets that we will find particularly useful in the chapters ahead.

**Definition 2.2.1.** Let \( A \) and \( B \) be any two events defined over the same sample space \( S \). Then

The **intersection** of \( A \) and \( B \), written \( A \cap B \), is the event whose outcomes belong to both \( A \) and \( B \).

The **union** of \( A \) and \( B \), written \( A \cup B \), is the event whose outcomes belong to either \( A \) or \( B \) or both.
A single card is drawn from a poker deck. Let $A$ be the event that an ace is selected:

$$A = \{\text{ace of hearts, ace of diamonds, ace of clubs, ace of spades}\}$$

Let $B$ be the event “Heart is drawn”:

$$B = \{2 \text{ of hearts, 3 of hearts, \ldots, ace of hearts}\}$$

Then

$$A \cap B = \{\text{ace of hearts}\}$$

and

$$A \cup B = \{2 \text{ of hearts, 3 of hearts, \ldots, ace of hearts, ace of diamonds, ace of clubs, ace of spades}\}$$

(Let $C$ be the event “Club is drawn.” Which cards are in $B \cup C$? In $B \cap C$?)

Let $A$ be the set of $x$’s for which $x^2 + 2x = 8$; let $B$ be the set for which $x^2 + x = 6$. Find $A \cap B$ and $A \cup B$.

Since the first equation factors into $(x + 4)(x - 2) = 0$, its solution set is $A = \{-4, 2\}$. Similarly, the second equation can be written $(x + 3)(x - 2) = 0$, making $B = \{-3, 2\}$. Therefore,

$$A \cap B = \{2\}$$

and

$$A \cup B = \{-4, -3, 2\}$$

Consider the electrical circuit pictured in Figure 2.2.2. Let $A_i$ denote the event that switch $i$ fails to close, $i = 1, 2, 3, 4$. Let $A$ be the event “Circuit is not completed.” Express $A$ in terms of the $A_i$’s.

Call the ① and ② switches line $a$; call the ③ and ④ switches line $b$. By inspection, the circuit fails only if both line $a$ and line $b$ fail. But line $a$ fails only if either ① or ② (or both) fail. That is, the event that line $a$ fails is the union $A_1 \cup A_2$. Similarly, the failure of line $b$ is the union $A_3 \cup A_4$. The event that the circuit fails, then, is an intersection:

$$A = (A_1 \cup A_2) \cap (A_3 \cup A_4)$$

Definition 2.2.2. Events $A$ and $B$ defined over the same sample space are said to be mutually exclusive if they have no outcomes in common—that is, if $A \cap B = \emptyset$, where $\emptyset$ is the null set.
**Example 2.2.9**

Consider a single throw of two dice. Define $A$ to be the event that the sum of the faces showing is odd. Let $B$ be the event that the two faces themselves are odd. Then clearly, the intersection is empty, the sum of two odd numbers necessarily being even. In symbols, $A \cap B = \emptyset$. (Recall the event $B \cap C$ asked for in Example 2.2.6.)

**Definition 2.2.3.** Let $A$ be any event defined on a sample space $S$. The **complement** of $A$, written $A^C$, is the event consisting of all the outcomes in $S$ other than those contained in $A$.

**Example 2.2.10**

Let $A$ be the set of $(x, y)$'s for which $x^2 + y^2 < 1$. Sketch the region in the $x$-$y$-plane corresponding to $A^C$.

From analytic geometry, we recognize that $x^2 + y^2 < 1$ describes the interior of a circle of radius 1 centered at the origin. Figure 2.2.3 shows the complement—the points on the circumference of the circle and the points outside the circle.

![Figure 2.2.3](image)

The notions of union and intersection can easily be extended to more than two events. For example, the expression $A_1 \cup A_2 \cup \ldots \cup A_k$ defines the set of outcomes belonging to any of the $A_i$'s (or to any combination of the $A_i$'s). Similarly, $A_1 \cap A_2 \cap \ldots \cap A_k$ is the set of outcomes belonging to all of the $A_i$'s.

**Example 2.2.11**

Suppose the events $A_1, A_2, \ldots, A_k$ are intervals of real numbers such that

$A_i = \{x : 0 \leq x < 1/2 \}$, \quad $i = 1, 2, \ldots, k$

Describe the sets $A_1 \cup A_2 \cup \ldots \cup A_k = \bigcup_{i=1}^{k} A_i$ and $A_1 \cap A_2 \cap \ldots \cap A_k = \bigcap_{i=1}^{k} A_i$.

Notice that the $A_i$'s are telescoping sets. That is, $A_1$ is the interval $0 \leq x < 1$, $A_2$ is the interval $0 \leq x < 1/2$, and so on. It follows, then, that the union of the $k A_i$'s is simply $A_1$ while the intersection of the $A_i$'s (that is, their overlap) is $A_k$.

**Questions**

**2.2.16.** Sketch the regions in the $x$-$y$-plane corresponding to $A \cup B$ and $A \cap B$ if

$A = \{(x, y) : 0 < x < 3, 0 < y < 3\}$

and

$B = \{(x, y) : 2 < x < 4, 2 < y < 4\}$

**2.2.17.** Referring to Example 2.2.7, find $A \cap B$ and $A \cup B$ if the two equations were replaced by inequalities: $x^2 + 2x \leq 0$ and $x^2 + x \leq 6$.

**2.2.18.** Find $A \cap B \cap C$ if $A = \{x: 0 \leq x \leq 4\}$, $B = \{x: 2 \leq x \leq 6\}$, and $C = \{x: x = 0, 1, 2, \ldots\}$.
2.2.19. An electronic system has four components divided into two pairs. The two components of each pair are wired in parallel; the two pairs are wired in series. Let $A_{ij}$ denote the event “$i$th component in $j$th pair fails,” $i = 1, 2; j = 1, 2$. Let $A$ be the event “System fails.” Write in terms of the $A_{ij}$’s.

2.2.20. Define $A = \{x : 0 \leq x \leq 1\}$, $B = \{x : 0 \leq x \leq 3\}$, and $C = \{x : -1 \leq x \leq 2\}$. Draw diagrams showing each of the following sets of points:

- $A^C \cap B \cap C$
- $A^C \cup (B \cap C)$
- $A \cap B \cap C^C$
- $[(A \cup B) \cap B] \cap C^C$

2.2.21. Let $A$ be the set of five-card hands dealt from a 52-card poker deck, where the denominations of the five cards are all consecutive—for example, (7 of hearts, 8 of spades, 9 of spades, 10 of hearts, jack of diamonds). Let $B$ be the set of five-card hands where the suits of the five cards are all the same. How many outcomes are in the event $A \cap B$?

2.2.22. Suppose that each of the twelve letters in the word

<table>
<thead>
<tr>
<th>T</th>
<th>E</th>
<th>S</th>
<th>S</th>
<th>E</th>
<th>L</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>T</td>
<td>I</td>
<td>O</td>
<td>N</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

is written on a chip. Define the events $F$, $R$, and $C$ as follows:

- $F$: letters in first half of alphabet
- $R$: letters that are repeated
- $V$: letters that are vowels

Which chips make up the following events?

- $F \cap R \cap V$
- $F \cap R \cap V^C$
- $F \cap R^C \cap V$

2.2.23. Let $A$, $B$, and $C$ be any three events defined on a sample space $S$. Show that

- the outcomes in $A \cup (B \cap C)$ are the same as the outcomes in $(A \cup B) \cap (A \cup C)$.
- the outcomes in $A \cap (B \cup C)$ are the same as the outcomes in $(A \cap B) \cup (A \cap C)$.

2.2.24. Let $A_1, A_2, \ldots, A_k$ be any set of events defined on a sample space $S$. What outcomes belong to the event

$$(A_1 \cup A_2 \cup \cdots \cup A_k) \cup A_1^c \cap A_2^c \cap \cdots \cap A_k^c$$

2.2.25. Let $A$, $B$, and $C$ be any three events defined on a sample space $S$. Show that the operations of union and intersection are associative by proving that

- $A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$
- $A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$

2.2.26. Suppose that three events—$A$, $B$, and $C$—are defined on a sample space $S$. Use the union, intersection, and complement operations to represent each of the following events:

- none of the three events occurs
- all three of the events occur
- only event $A$ occurs
- exactly one event occurs
- exactly two events occur

2.2.27. What must be true of events $A$ and $B$ if

- $A \cup B = B$
- $A \cap B = A$

2.2.28. Let events $A$ and $B$ and sample space $S$ be defined as the following intervals:

$S = \{x : 0 \leq x \leq 10\}$
$A = \{x : 0 < x < 5\}$
$B = \{x : 3 \leq x \leq 7\}$

Characterize the following events:

- $A^C$
- $A \cap B$
- $A \cup B$
- $A \cap B^C$
- $A^C \cup B$
- $A^C \cap B^C$

2.2.29. A coin is tossed four times and the resulting sequence of heads and/or tails is recorded. Define the events $A$, $B$, and $C$ as follows:

- $A$: exactly two heads appear
- $B$: heads and tails alternate
- $C$: first two tosses are heads

Which events, if any, are mutually exclusive? Which events, if any, are subsets of other sets?

2.2.30. Pictured on the next page are two organizational charts describing the way upper management vets new proposals. For both models, three vice presidents—1, 2, and 3—each voice an opinion.
For (a), all three must concur if the proposal is to pass; if any one of the three favors the proposal in (b), it passes. Let \( A_i \) denote the event that vice president \( i \) favors the proposal, \( i = 1, 2, 3 \), and let \( A \) denote the event that the proposal passes. Express \( A \) in terms of the \( A_i \)'s for the two office protocols. Under what sorts of situations might one system be preferable to the other?

Expressing Events Graphically: Venn Diagrams

Relationships based on two or more events can sometimes be difficult to express using only equations or verbal descriptions. An alternative approach that can be highly effective is to represent the underlying events graphically in a format known as a Venn diagram. Figure 2.2.4 shows Venn diagrams for an intersection, a union, a complement, and two events that are mutually exclusive. In each case, the shaded interior of a region corresponds to the desired event.

**Example 2.2.12**

When two events \( A \) and \( B \) are defined on a sample space, we will frequently need to consider

- the event that *exactly one* (of the two) occurs.
- the event that *at most one* (of the two) occurs.

Getting expressions for each of these is easy if we visualize the corresponding Venn diagrams.

The shaded area in Figure 2.2.5 represents the event \( E \) that either \( A \) or \( B \), but not both, occurs (that is, *exactly one* occurs).
Chapter 2 Probability

Just by looking at the diagram we can formulate an expression for \( E \). The portion of \( A \), for example, included in \( E \) is \( A \cap B^C \). Similarly, the portion of \( B \) included in \( E \) is \( B \cap A^C \). It follows that \( E \) can be written as a union:

\[
E = (A \cap B^C) \cup (B \cap A^C)
\]

(Convince yourself that an equivalent expression for \( E \) is \((A \cap B)^C \cap (A \cup B)\).) Figure 2.2.6 shows the event \( F \) that at most one (of the two events) occurs. Since the latter includes every outcome except those belonging to both \( A \) and \( B \), we can write

\[
F = (A \cap B)^C
\]

![Figure 2.2.6](image)

Questions

2.2.31. During orientation week, the latest Spiderman movie was shown twice at State University. Among the entering class of 6000 freshmen, 850 went to see it the first time, 690 the second time, while 4700 failed to see it either time. How many saw it twice?

2.2.32. Let \( A \) and \( B \) be any two events. Use Venn diagrams to show that

the complement of their intersection is the union of their complements:

\[
(A \cap B)^C = A^C \cup B^C
\]

the complement of their union is the intersection of their complements:

\[
(A \cup B)^C = A^C \cap B^C
\]

(These two results are known as De Morgan’s laws.)

2.2.33. Let \( A \), \( B \), and \( C \) be any three events. Use Venn diagrams to show that

\[
A \cap (B \cup C) = (A \cap B) \cup (A \cap C)
\]

\[
A \cup (B \cap C) = (A \cup B) \cap (A \cup C)
\]

2.2.34. Let \( A \), \( B \), and \( C \) be any three events. Use Venn diagrams to show that

\[
A \cup (B \cup C) = (A \cup B) \cup C
\]

\[
A \cap (B \cap C) = (A \cap B) \cap C
\]

2.2.35. Let \( A \) and \( B \) be any two events defined on a sample space \( S \). Which of the following sets are necessarily subsets of which other sets?

\[
A \quad B \quad A \cup B \quad A \cap B \quad A \cap B^C \quad A^C \cap B \quad A \cap B^C \\
A \cap (B \cup C) \quad (A \cap B)^C \cap (A \cup B)^C
\]

2.2.36. Use Venn diagrams to suggest an equivalent way of representing the following events:

\[
(A \cap B)^C \quad B \cup (A \cup B)^C \quad A \cap (A \cup B)^C
\]

2.2.37. A total of twelve hundred graduates of State Tech have gotten into medical school in the past several years. Of that number, one thousand earned scores of twenty-seven or higher on the MCAT and four hundred had GPAs that were 3.5 or higher. Moreover, three hundred had MCATs that were twenty-seven or higher and GPAs that were 3.5 or higher. What proportion of those twelve hundred graduates got into medical school with an MCAT lower than twenty-seven and a GPA below 3.5?

2.2.38. Let \( A \), \( B \), and \( C \) be any three events defined on a sample space \( S \). Let \( N(A) \), \( N(B) \), \( N(C) \), \( N(A \cap B) \), \( N(A \cap C) \), \( N(B \cap C) \), and \( N(A \cap B \cap C) \) denote the numbers of outcomes in all the different intersections in which \( A \), \( B \), and \( C \) are involved. Use a Venn diagram to suggest a formula for \( N(A \cup B \cup C) \). [Hint: Start with the
sum $N(A) + N(B) + N(C)$ and use the Venn diagram to identify the “adjustments” that need to be made to that sum before it can equal $N(A \cup B \cup C)$. As a precedent, note that $N(A \cup B) = N(A) + N(B) - N(A \cap B)$. There, in the case of two events, subtracting $N(A \cap B)$ is the “adjustment.”

2.2.39. A poll conducted by a potential presidential candidate asked two questions: (1) Do you support the candidate’s position on taxes? and (2) Do you support the candidate’s position on homeland security? A total of twelve hundred responses were received; six hundred said “yes” to the first question and four hundred said “yes” to the second. If three hundred respondents said “no” to the taxes question and “yes” to the homeland security question, how many said “yes” to the taxes question but “no” to the homeland security question?

2.2.40. For two events $A$ and $B$ defined on a sample space $S$, $N(A \cap B^C) = 15$, $N(A^C \cap B) = 50$, and $N(A \cap B) = 2$. Given that $N(S) = 120$, how many outcomes belong to neither $A$ nor $B$?

2.3 The Probability Function

Having introduced in Section 2.2 the twin concepts of “experiment” and “sample space,” we are now ready to pursue in a formal way the all-important problem of assigning a probability to an experiment’s outcome—and, more generally, to an event. Specifically, if $A$ is any event defined on a sample space $S$, the symbol $P(A)$ will denote the probability of $A$, and we will refer to $P$ as the probability function. It is, in effect, a mapping from a set (i.e., an event) to a number. The backdrop for our discussion will be the unions, intersections, and complements of set theory; the starting point will be the axioms referred to in Section 2.1 that were originally set forth by Kolmogorov.

If $S$ has a finite number of members, Kolmogorov showed that as few as three axioms are necessary and sufficient for characterizing the probability function $P$:

**Axiom 1.** Let $A$ be any event defined over $S$. Then $P(A) \geq 0$.

**Axiom 2.** $P(S) = 1$.

**Axiom 3.** Let $A$ and $B$ be any two mutually exclusive events defined over $S$. Then

$$P(A \cup B) = P(A) + P(B)$$

When $S$ has an infinite number of members, a fourth axiom is needed:

**Axiom 4.** Let $A_1, A_2, \ldots$, be events defined over $S$. If $A_i \cap A_j = \emptyset$ for each $i \neq j$, then

$$P(A_i) = \sum_{i=1}^{\infty} P(A_i)$$

From these simple statements come the general rules for manipulating the probability function that apply no matter what specific mathematical form the function may take in a particular context.

**Some Basic Properties of $P$**

Some of the immediate consequences of Kolmogorov’s axioms are the results given in Theorems 2.3.1 through 2.3.6. Despite their simplicity, several of these properties—as we will soon see—prove to be immensely useful in solving all sorts of problems.
Chapter 2 Probability

Theorem 2.3.1
\[ P( A^C ) = 1 - P( A). \]

Proof By Axiom 2 and Definition 2.2.3,
\[ P(S) = 1 = P( A \cup A^C ) \]
But A and \( A^C \) are mutually exclusive, so
\[ P( A \cup A^C ) = P( A) + P( A^C ) \]
and the result follows.

Theorem 2.3.2
\[ P(\emptyset) = 0. \]

Proof Since \( \emptyset = S^C \), \( P(\emptyset) = P(S^C) = 1 - P(S) = 0 \).

Theorem 2.3.3
If \( A \subseteq B \), then \( P( A) \leq P( B) \).

Proof Note that the event \( B \) may be written in the form
\[ B = A \cup ( B \cap A^C ) \]
where A and \( B \cap A^C \) are mutually exclusive. Therefore,
\[ P( B) = P( A) + P( B \cap A^C ) \]
which implies that \( P( B) \geq P( A) \) since \( P( B \cap A^C ) \geq 0 \).

Theorem 2.3.4
For any event \( A \), \( P( A) \leq 1 \).

Proof The proof follows immediately from Theorem 2.3.3 because \( A \subseteq S \) and \( P(S) = 1 \).

Theorem 2.3.5
Let \( A_1, A_2, \ldots, A_n \) be events defined over \( S \). If \( A_i \cap A_j = \emptyset \) for \( i \neq j \), then
\[ P \sum_{i=1}^{n} A_i = \sum_{i=1}^{n} P( A_i) \]

Proof The proof is a straightforward induction argument with Axiom 3 being the starting point.

Theorem 2.3.6
\[ P( A \cup B) = P( A) + P( B) - P( A \cap B). \]

Proof The Venn diagram for \( A \cup B \) certainly suggests that the statement of the theorem is true (recall Figure 2.2.4). More formally, we have from Axiom 3 that
\[ P( A) = P( A \cap B^C) + P( A \cap B) \]
and
\[ P( B) = P( B \cap A^C) + P( A \cap B) \]
Adding these two equations gives
\[ P( A) + P( B) = [P( A \cap B^C) + P( B \cap A^C) + P( A \cap B)] + P( A \cap B) \]
By Theorem 2.3.5, the sum in the brackets is \( P( A \cup B) \). If we subtract \( P( A \cap B) \) from both sides of the equation, the result follows.
Example 2.3.1

Let $A$ and $B$ be two events defined on a sample space $S$ such that $P( A ) = 0.3$, $P( B ) = 0.5$, and $P( A \cup B ) = 0.7$. Find (a) $P( A \cap B )$, (b) $P( A^C \cup B^C )$, and (c) $P( A^C \cap B )$.

a. Transposing the terms in Theorem 2.3.6 yields a general formula for the probability of an intersection:

$$P( A \cap B ) = P( A ) + P( B ) - P( A \cup B )$$

Here

$$P( A \cap B ) = 0.3 + 0.5 - 0.7$$

$$= 0.1$$

The two cross-hatched regions in Figure 2.3.1 correspond to $A^C$ and $B^C$. The union of $A^C$ and $B^C$ consists of those regions that have cross-hatching in either or both directions. By inspection, the only portion of $S$ not included in $A^C \cup B^C$ is the intersection, $A \cap B$. By Theorem 2.3.1, then,

$$P( A^C \cup B^C ) = 1 - P( A \cap B )$$

$$= 1 - 0.1$$

$$= 0.9$$

![Figure 2.3.1](image1)

![Figure 2.3.2](image2)

c. The event $A^C \cap B$ corresponds to the region in Figure 2.3.2 where the cross-hatching extends in both directions—that is, everywhere in $B$ except the intersection with $A$. Therefore,

$$P( A^C \cap B ) = P( B ) - P( A \cap B )$$

$$= 0.5 - 0.1$$

$$= 0.4$$
Example 2.3.2

Show that

\[ P(A \cap B) \geq 1 - P(A^C) - P(B^C) \]

for any two events \( A \) and \( B \) defined on a sample space \( S \).

From Example 2.3.1a and Theorem 2.3.1,

\[
P(A \cap B) = P(A) + P(B) - P(A \cup B) = 1 - P(A^C) + 1 - P(B^C) - P(A \cup B)
\]

But \( P(A \cup B) \leq 1 \) from Theorem 2.3.4, so

\[ P(A \cap B) \geq 1 - P(A^C) - P(B^C) \]

Example 2.3.3

Two cards are drawn from a poker deck without replacement. What is the probability that the second is higher in rank than the first?

Let \( A_1, A_2, \) and \( A_3 \) be the events “First card is lower in rank,” “First card is higher in rank,” and “Both cards have same rank,” respectively. Clearly, the three \( A_i \)'s are mutually exclusive and they account for all possible outcomes, so from Theorem 2.3.5,

\[
P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) = P(S) = 1
\]

Once the first card is drawn, there are three choices for the second that would have the same rank—that is, \( P(A_3) = \frac{51}{52} \). Moreover, symmetry demands that \( P(A_1) = P(A_2) \), so

\[
2P(A_2) + \frac{3}{51} = 1
\]

implying that \( P(A_2) = \frac{17}{52} \).

Example 2.3.4

In a newly released martial arts film, the actress playing the lead role has a stunt double who handles all of the physically dangerous action scenes. According to the script, the actress appears in 40% of the film’s scenes, her double appears in 30%, and the two of them are together 5% of the time. What is the probability that in a given scene, (a) only the stunt double appears and (b) neither the lead actress nor the double appears?

If \( L \) is the event “Lead actress appears in scene” and \( D \) is the event “Double appears in scene,” we are given that \( P(L) = 0.40, P(D) = 0.30, \) and \( P(L \cap D) = 0.05 \). It follows that

\[
P(\text{Only double appears}) = P(D) - P(L \cap D)
\]

\[
= 0.30 - 0.05
\]

\[
= 0.25
\]

(recall Example 2.3.1c).
b. The event “Neither appears” is the complement of the event “At least one appears.” But \( P(\text{At least one appears}) = P(L \cup D) \). From Theorems 2.3.1 and 2.3.6, then,

\[
P(\text{Neither appears}) = 1 - P(L \cup D) \\
= 1 - [P(L) + P(D) - P(L \cap D)] \\
= 1 - [0.40 + 0.30 - 0.05] \\
= 0.35
\]

**Example 2.3.5**

Having endured (and survived) the mental trauma that comes from taking two years of chemistry, a year of physics, and a year of biology, Biff decides to test the medical school waters and sends his MCATs to two colleges, \( X \) and \( Y \). Based on how his friends have fared, he estimates that his probability of being accepted at \( X \) is 0.7, and at \( Y \) is 0.4. He also suspects there is a 75% chance that at least one of his applications will be rejected. What is the probability that he gets at least one acceptance?

Let \( A \) be the event “School \( X \) accepts him” and \( B \) the event “School \( Y \) accepts him.” We are given that \( P(\ A) = 0.7, P(\ B) = 0.4, \) and \( P(\ A^C \cup B^C) = 0.75. \) The question is asking for \( P(\ A \cup B) \).

From Theorem 2.3.6,

\[
P(\ A \cup B) = P(\ A) + P(\ B) - P(\ A \cap B)
\]

Recall from Question 2.2.32 that \( A^C \cup B^C = (A \cap B)^C \), so

\[
P(\ A \cap B) = 1 - P(\ A \cap B)^C = 1 - 0.75 = 0.25
\]

It follows that Biff’s prospects are not all that bleak—he has an 85% chance of getting in somewhere:

\[
P(\ A \cup B) = 0.7 + 0.4 - 0.25
\]

\[
= 0.85
\]

**Comment** Notice that \( P(\ A \cup B) \) varies directly with \( P(\ A^C \cup B^C) \):

\[
P(\ A \cup B) = P(\ A) + P(\ B) - [1 - P(\ A^C \cup B^C)] = P(\ A) + P(\ B) - 1 + P(\ A^C \cup B^C)
\]

If \( P(\ A) \) and \( P(\ B) \), then, are fixed, we get the curious result that Biff’s chances of getting at least one acceptance increase if his chances of at least one rejection increase.

**Questions**

2.3.1. According to a family-oriented lobbying group, there is too much crude language and violence on tele-vision. Forty-two percent of the programs they screened had language they found offensive, 27% were too violent, and 10% were considered excessive in both language and violence. What percentage of programs did comply with the group’s standards?

2.3.2. Let \( A \) and \( B \) be any two events defined on \( S \). Suppose that \( P(\ A) = 0.4, P(\ B) = 0.5, \) and \( P(\ A \cap B) = 0.1. \) What is the probability that \( A \) or \( B \) but not both occur?

2.3.3. Express the following probabilities in terms of \( P(\ A), P(\ B), \) and \( P(\ A \cap B). \)
Chapter 2 Probability

\[ P( A^C \cup B^C ) \]
\[ P( A^C \cap ( A \cup B )) \]

2.3.4. Let \( A \) and \( B \) be two events defined on \( S \). If the probability that at least one of them occurs is 0.3 and the probability that \( A \) occurs but \( B \) does not occur is 0.1, what is \( P( B ) \)?

2.3.5. Suppose that three fair dice are tossed. Let \( A_i \) be the event that a 6 shows on the \( i \)th die, \( i = 1, 2, 3 \). Does \( P( A_1 \cup A_2 \cup A_3 ) = \frac{1}{2} \) ? Explain.

2.3.6. Events \( A \) and \( B \) are defined on a sample space \( S \) such that \( P( A \cup B )^C = 0.5 \) and \( P( A \cap B ) = 0.2 \). What is the probability that either \( A \) or \( B \) but not both will occur?

2.3.7. Let \( A_1, A_2, \ldots, A_n \) be a series of events for which \( A_i \cap A_j = \emptyset \) if \( i \neq j \) and \( A_1 \cup A_2 \cup \ldots \cup A_n = S \). Let \( B \) be any event defined on \( S \). Express \( B \) as a union of intersections.

2.3.8. Draw the Venn diagrams that would correspond to the equations (a) \( P( A \cap B ) = P( B ) \) and (b) \( P( A \cup B ) = P( B ) \).

2.3.9. In the game of “odd man out” each player tosses a fair coin. If all the coins turn up the same except for one, the player tossing the different coin is declared the odd man out and is eliminated from the contest. Suppose that three people are playing. What is the probability that someone will be eliminated on the first toss? (Hint: Use Theorem 2.3.1.)

2.3.10. An urn contains twenty-four chips, numbered 1 through 24. One is drawn at random. Let \( A \) be the event that the number is divisible by 2 and let \( B \) be the event that the number is divisible by 3. Find \( P( A \cup B ) \).

2.3.11. If State’s football team has a 10% chance of winning Saturday’s game, a 30% chance of winning two weeks from now, and a 65% chance of losing both games, what are their chances of winning exactly once?

2.3.12. Events \( A_1 \) and \( A_2 \) are such that \( A_1 \cup A_2 = S \) and \( A_1 \cap A_2 = \emptyset \). Find \( p_2 \) if \( P( A_1 ) = p_1, P( A_2 ) = p_2, \) and \( 3p_1 - p_2 = \frac{1}{2} \).

2.3.13. Consolidated Industries has come under considerable pressure to eliminate its seemingly discriminatory