CONTENTS

CHAPTER 1

BASIC PROBABILITY THEORY

1.1 INTRODUCTION

Probability theory is foundational for economics and econometrics. Probability is the mathematical language used to handle uncertainty, which is central for modern economic theory. Probability theory is also the foundation of mathematical statistics, which is the foundation of econometric theory.

Probability is used to model uncertainty, variability, and randomness. When we say that something is "uncertain", we mean that the outcome is unknown. For example, how many students will there be in next year's Ph.D. entering class at your university? "Variability" means that the outcome is not the same across all occurrences. For example, the number of Ph.D. students fluctuates from year to year. "Randomness" means that the variability has some sort of pattern. For example, the number of Ph.D. students may fluctuate between 20 and 30, with 25 more likely than either 20 or 30. Probability gives us a mathematical language to describe uncertainty, variability, and randomness.

1.2 OUTCOMES AND EVENTS

Suppose you take a coin, flip it in the air, and let it land on the ground. What will happen? Will the result be "heads" (H) or "tails" (T)? We do not know the result in advance, so we describe the outcome as **random**.

Suppose you record the change in the value of a stock index over a period of time. Will the value increase or decrease? Again, we do not know the result in advance, so we describe the outcome as random.

Suppose you select an individual at random and survey them about their economic situation. What is their hourly wage? We do not know in advance. The lack of foreknowledge leads us to describe the outcome as random.

We will use the following terms.

An **outcome** is a specific result. For example, in a coin flip, an outcome is either H or T. If two coins are flipped in sequence, we can write an outcome as HT for a head and then a tails. A roll of a six-sided die has the six outcomes $\{1, 2, 3, 4, 5, 6\}$.

The **sample space** S is the set of all possible outcomes. In a coin flip, the sample space is $S = \{H, T\}$. If two coins are flipped, the sample space is $S = \{HH, HT, TH, TT\}$.

An **event** A is a subset of outcomes in S. An example event from the roll of a die is $A = \{1, 2\}$.

The one-coin and two-coin sample spaces are illustrated in Figure 1.1. The event $\{HH, HT\}$ is illustrated by the ellipse in Figure 1.1(b).

Set theoretic manipulations are helpful in describing events. We will use the following concepts.

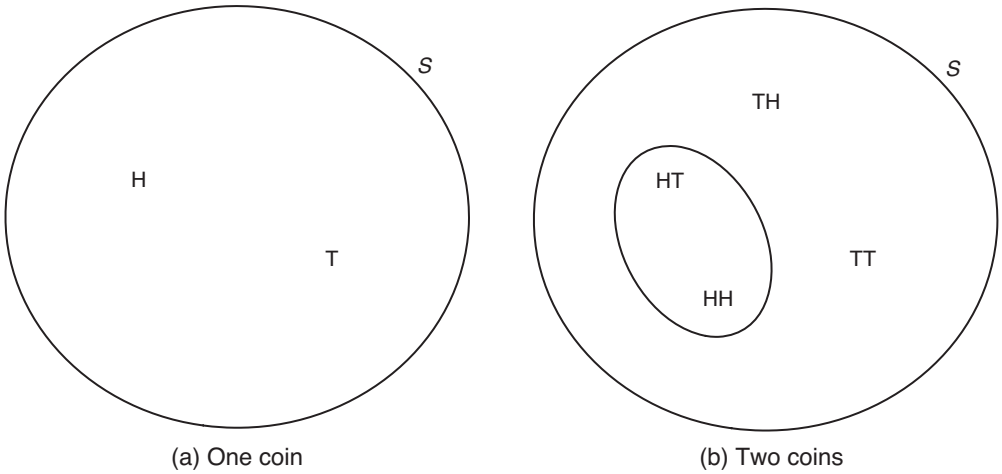

FIGURE 1.1 Sample space

Definition 1.1 For events A and B:

1. A is a **subset** of B, written $A \subset B$, if every element of A is an element of B.
2. The event with no outcomes $\varnothing = \{\ \}$ is called the **null** or **empty set**.
3. The **union** $A \cup B$ is the collection of all outcomes that are in either A **or** B (or both).
4. The **intersection** $A \cap B$ is the collection of elements that are in both A **and** B.
5. The **complement** A^c of A are all outcomes in S which are not in A.
6. The events A and B are **disjoint** if they have no outcomes in common: $A \cap B = \varnothing$.
7. The events $A_1, A_2, \ldots$ are a **partition** of S if they are mutually disjoint and their union is S.

Events satisfy the rules of set operations, including the commutative, associative, and distributive laws. The following theorem is useful.

Theorem 1.1 **Partitioning Theorem**. If $\{B_1, B_2, \cdots\}$ is a partition of S, then for any event A,

$$A = \bigcup_{i=1}^{\infty} (A \cap B_i)\,.$$

The sets $(A \cap B_i)$ are mutually disjoint.

A proof is provided in Section 1.15.

1.3 PROBABILITY FUNCTION

Definition 1.2 A function $\mathbb{P}$ which assigns a numerical value to events[1] is called a **probability function** if it satisfies the following **axioms of probability**:

1. $\mathbb{P}[A] \geq 0$.

[1] For events in a sigma field. See Section 1.14.

2. $\mathbb{P}[S] = 1$.
3. If $A_1, A_2, \ldots$ are disjoint, then $\mathbb{P}\left[\bigcup_{j=1}^{\infty} A_j\right] = \sum_{j=1}^{\infty} \mathbb{P}\left[A_j\right]$.

This textbook uses the notation $\mathbb{P}[A]$ for the probability of an event A. Other common notations include $P(A)$ and $\Pr(A)$.

Let us examine this definition. The phrase "a function $\mathbb{P}$ which assigns a numerical value to events" means that $\mathbb{P}$ is a function from the space of events to the real line. Thus probabilities are numbers. Now consider the axioms. The first axiom states that probabilities are nonnegative. The second axiom is essentially a normalization: the probability that "something happens" is 1.

The third axiom imposes considerable structure. It states that probabilities are additive on disjoint events. That is, if A and B are disjoint, then

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B].$$

Take, for example, the roll of a six-sided die which has the possible outcomes $\{1, 2, 3, 4, 5, 6\}$. Since the outcomes are mutually disjoint, the third axiom states that $\mathbb{P}[1 \text{ or } 2] = \mathbb{P}[1] + \mathbb{P}[2]$.

When using the third axiom, it is important to be careful that it is applied only to disjoint events. Take, for example, the roll of a pair of dice. Let A be the event "1 on the first roll" and B the event "1 on the second roll". It is tempting to write $\mathbb{P}[\text{"1 on either roll"}] = \mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B]$, but the second equality is incorrect, since A and B are not disjoint. The outcome "1 on both rolls" is an element of both A and B.

Any function $\mathbb{P}$ which satisfies the axioms is a valid probability function. Take the coin flip example. One valid probability function sets $\mathbb{P}[H] = 0.5$ and $\mathbb{P}[T] = 0.5$. (This is typically called a **fair coin**.) A second valid probability function sets $\mathbb{P}[H] = 0.6$ and $\mathbb{P}[T] = 0.4$. However, a function which sets $\mathbb{P}[H] = -0.6$ is not valid (it violates the first axiom), and a function which sets $\mathbb{P}[H] = 0.6$ and $\mathbb{P}[T] = 0.6$ is not valid (it violates the second axiom).

While the definition states that a probability function must satisfy certain rules, it does not describe the *meaning* of probability. The reason is because there are multiple interpretations. One view is that probabilities are the relative frequency of outcomes, as in a controlled experiment. The probability that the stock market will increase is the frequency of increases. The probability that an unemployment duration will exceed one month is the frequency of unemployment durations exceeding one month. The probability that a basketball player will make a free throw shot is the frequency with which the player makes free throw shots. The probability that a recession will occur is the relative frequency of recessions. In some examples, this definition is conceptually straightforward, as the experiment repeats or has multiple occurances. In other cases, a situation occurs exactly once and will never be repeated. As I write this paragraph, questions of uncertainty of general interest include "Will global warming exceed 2 degrees?" and "When will the COVID-19 epidemic end?" In these cases, it is difficult to interpret a probability as a relative frequency, as the outcome can only occur once. The interpretation can be salvaged by viewing "relative frequency" abstractly by imagining many alternative universes which start from the same initial conditions but evolve randomly. While this solution works (technically), it is not completely satisfactory.

Another view is that probability is subjective. This view holds that probabilities can be interpreted as degrees of belief. If I say "The probability of rain tomorrow is 80%", I mean that this is my personal subjective assessment of the likelihood based on the information available to me. This view may seem too broad, as it allows for arbitrary beliefs, but the subjective interpretation requires subjective probability to follow the axioms and rules of probability. A major disadvantage associated with this approach is that it is not necessarily appropriate for scientific discourse.

What is common between the two definitions is that the probability function follows the same axioms—otherwise, the label "probability" should not be used.

This concept can be illustrated with two real-world examples. The first is from finance. Let U be the event that the S&P stock index increases in a given week, and let D be the event that the index decreases. This is similar to a coin flip. The sample space is $\{U, D\}$. We compute[2] that $\mathbb{P}[U] = 0.57$ and $\mathbb{P}[D] = 0.43$. The probability 57% of an increase is somewhat higher than a fair coin. The probability interpretation is that the index will increase in value in 57% of randomly selected weeks.

The second example concerns wage rates in the United States. Take a randomly selected wage earner. Let H be the event that their wage rate exceeds \$25/hour, and L be the event that their wage rate is less than \$25/hour. Again the structure is similar to a coin flip. We calculate[3] that $\mathbb{P}[H] = 0.31$ and $\mathbb{P}[L] = 0.69$. To interpret this as a probability, we can imagine surveying a random individual. Before the survey, we know nothing about the individual. Their wage rate is uncertain and random.

1.4 PROPERTIES OF THE PROBABILITY FUNCTION

The following properties of probability functions can be derived from the axioms of probability.

Theorem 1.2 For events A and B, the following properties hold:

1. $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$.
2. $\mathbb{P}[\varnothing] = 0$.
3. $\mathbb{P}[A] \leq 1$.
4. **Monotone Probability Inequality**: If $A \subset B$, then $\mathbb{P}[A] \leq \mathbb{P}[B]$.
5. **Inclusion-Exclusion Principle**: $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$.
6. **Boole's Inequality**: $\mathbb{P}[A \cup B] \leq \mathbb{P}[A] + \mathbb{P}[B]$.
7. **Bonferroni's Inequality**: $\mathbb{P}[A \cap B] \geq \mathbb{P}[A] + \mathbb{P}[B] - 1$.

Proofs are provided in Section 1.15.

Property 1 states that the probability that an event does not occur equals 1 minus the probability that the event occurs.

Property 2 states that "nothing happens" occurs with 0 probability. (Remember this when asked "What happened today in class?")

Property 3 states that probabilities cannot exceed 1.

Property 4 shows that larger sets necessarily have larger probability.

Property 5 is a useful decomposition of the probability of the union of two events.

Properties 6 and 7 are implications of the inclusion-exclusion principle and are frequently used in probability calculations. Boole's inequality shows that the probability of a union is bounded by the sum of the individual probabilities. Bonferroni's inequality shows that the probability of an intersection is bounded below by an expression involving the individual probabilities. A useful feature of these inequalities is that the right-hand sides only depend on the individual probabilities.

[2] Calculated from a sample of 3,584 weekly prices of the S&P Index between 1950 and 2017.

[3] Calculated from a sample of 50,742 U.S. wage earners in 2009.

A further comment related to property 2 is that any event which occurs with probability 0 or 1 is called **trivial**. Such events are essentially nonrandom. In the coin flip example, we could define the sample space as $S = \{H, T, \text{Edge}, \text{Disappear}\}$, where "Edge" means the coin lands on its edge and "Disappear" means the coin disappears into the air. If $\mathbb{P}\left[\text{Edge}\right] = 0$ and $\mathbb{P}\left[\text{Disappear}\right] = 0$, then these events are trivial.

1.5 EQUALLY LIKELY OUTCOMES

When we build probability calculations from foundations, it is often useful to consider settings where symmetry implies that a set of outcomes is equally likely. Standard examples are a coin flip and the toss of a die. We describe a coin as **fair** if the event of a head is as equally likely as the event of a tail. We describe a die as **fair** if the event of each face is equally likely. Applying the axioms, we deduce the following.

Theorem 1.3 Principle of Equally Likely Outcomes: If an experiment has N outcomes $a_1, \ldots, a_N$ which are symmetric in the sense that each outcome is equally likely, then $\mathbb{P}[a_i] = \frac{1}{N}$.

For example, a fair coin satisfies $\mathbb{P}[H] = \mathbb{P}[T] = 1/2$, and a fair die satisfies $\mathbb{P}[1] = \cdots = \mathbb{P}[6] = 1/6$.

In some contexts, deciding which outcomes are symmetric and equally likely can be confusing. Take the two-coin example. We could define the sample space as {HH,TT,HT}, where HT means "one head and one tail". If we guess that all outcomes are equally likely, we would set $\mathbb{P}[HH] = 1/3$, etc. However, if we define the sample space as {HH,TT,HT,TH} and guess that all outcomes are equally likely, we would find $\mathbb{P}[HH] = 1/4$. Both answers (1/3 and 1/4) cannot be correct. The implication is that we should not apply the principle of equally likely outcomes simply because there is a list of outcomes. Instead, there should be a justifiable reason for the outcomes to be equally likely. In this two-coin example, there is no principled reason for symmetry without further analysis, so the property should not be applied. We return to this issue in Section 1.8.

1.6 JOINT EVENTS

Take two events H and C. For concreteness, let H be the event that an individual's wage exceeds \$25/hour, and let C be the event that the individual has a college degree. We are interested in the probability of the joint event $H \cap C$. This is the event "H and C", or in words, that the individual's wage exceeds \$25/hour and they have a college degree. Previously it was noted that $\mathbb{P}[H] = 0.31$. We can similarly calculate that $\mathbb{P}[C] = 0.36$. What about the joint event $H \cap C$?

From Theorem 1.2, we can deduce that $0 \leq \mathbb{P}[H \cap C] \leq 0.31$. (The upper bound is Bonferroni's inequality.) Thus from the knowledge of $\mathbb{P}[H]$ and $\mathbb{P}[C]$ alone, we can bound the joint probability but not determine its value. It turns out that the actual[4] probability is $\mathbb{P}[H \cap C] = 0.19$.

From the three known probabilities and the properties of Theorem 1.2, we can calculate the probabilities of the various intersections. The results are displayed in the following chart. The four numbers in the central box are the probabilities of the joint events; for example, 0.19 is the probability of both a high wage and a college degree. The largest of the four probabilities is 0.52: the joint event of a low wage and no college degree. The four probabilities sum to 1, because the events are a partition of the sample space. The sums of the probabilities in

[4] Calculated from the same sample of 50,742 U.S. wage earners in 2009.

each column are reported in the bottom row: the probabilities of a college degree and no degree, respectively. The sums by row are reported in the rightmost column: the probabilities of a high and low wage, respectively.

Joint Probabilities: Wages and Education

	C	N	Any Education
H	0.19	0.12	0.31
L	0.17	0.52	0.69
Any Wage	0.36	0.64	1.00

As another illustration, let us examine stock price changes. We reported before that the probability of an increase in the S&P stock index in a given week is 57%. Now consider the change in the stock index over 2 sequential weeks. What is the joint probability? The results are displayed in the following chart. U_t means that the index increases, D_t means that the index decreases, U_{t-1} means that the index increases in the previous week, and D_{t-1} means that the index decreases in the previous week.

Joint Probabilities: Stock Returns

	U_{t-1}	D_{t-1}	Any Past Return
U_t	0.322	0.245	0.567
D_t	0.245	0.188	0.433
Any Return	0.567	0.433	1.000

The four numbers in the central box sum to 1, since they are a partition of the sample space. We can see that the probability that the stock price increases for 2 weeks in a row is 32% and that it decreases for 2 weeks in a row is 19%. The probability is 25% for an increase followed by a decrease, and also 25% for a decrease followed by an increase.

1.7 CONDITIONAL PROBABILITY

Take two events A and B. For example, let A be the event "Receive a grade of A on the econometrics exam", and let B be the event "Study econometrics 12 hours a day". We might be interested in the question: Does B affect the likelihood of A? Alternatively, we may be interested in questions such as: Does attending college affect the likelihood of obtaining a high wage? Or: Do tariffs affect the likelihood of price increases? These are questions of **conditional probability**.

Abstractly, consider two events A and B. Suppose that we know that B has occurred. Then the only way for A to occur is if the outcome is in the intersection $A \cap B$. So we are asking: "What is the probability that $A \cap B$ occurs, given that B occurs?" The answer is not simply $\mathbb{P}[A \cap B]$. Instead, we can think of the "new" sample space as B. To do so, we normalize all probabilities by $\mathbb{P}[B]$. We arrive at the following definition.

Definition 1.3 If $\mathbb{P}[B] > 0$, the **conditional probability** of A given B is

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

The notation "$A \mid B$" means "A given B" or "A assuming that B is true". To add clarity, we will sometimes refer to $\mathbb{P}[A]$ as the **unconditional probability** to distinguish it from $\mathbb{P}[A \mid B]$.

For example, take the roll of a fair die. Let $A = \{1, 2, 3, 4\}$ and $B = \{4, 5, 6\}$. The intersection is $A \cap B = \{4\}$, which has probability $\mathbb{P}[A \cap B] = 1/6$. The probability of B is $\mathbb{P}[B] = 1/2$. Thus $\mathbb{P}[A \mid B] = (1/6)/(1/2) = 1/3$. This can also be calculated by observing that conditional on B, the events $\{4\}$, $\{5\}$, and $\{6\}$ each have probability 1/3. Event A only occurs given B if $\{4\}$ occurs. Thus $\mathbb{P}[A \mid B] = \mathbb{P}[4 \mid B] = 1/3$.

Consider our example of wages and college education. From the probabilities reported in Section 1.6, we can calculate that

$$\mathbb{P}[H \mid C] = \frac{\mathbb{P}[H \cap C]}{\mathbb{P}[C]} = \frac{0.19}{0.36} = 0.53$$

and

$$\mathbb{P}[H \mid N] = \frac{\mathbb{P}[H \cap N]}{\mathbb{P}[N]} = \frac{0.12}{0.64} = 0.19.$$

There is a considerable difference in the conditional probability of receiving a high wage conditional on a college degree: 53% versus 19%.

As another illustration, let us examine stock price changes. We calculate that

$$\mathbb{P}[U_t \mid U_{t-1}] = \frac{\mathbb{P}[U_t \cap U_{t-1}]}{\mathbb{P}[U_{t-1}]} = \frac{0.322}{0.567} = 0.568$$

and

$$\mathbb{P}[U_t \mid D_{t-1}] = \frac{\mathbb{P}[U_t \cap D_{t-1}]}{\mathbb{P}[D_{t-1}]} = \frac{0.245}{0.433} = 0.566.$$

In this case, the two conditional probabilities are essentially identical. Thus the probability of a price increase in a given week is unaffected by the previous week's result. This is an important special case and is explored further in the next section.

1.8 INDEPENDENCE

We say that events are **independent** if their occurrence is unrelated, or equivalently, that the knowledge of one event does not affect the conditional probability of the other event. Take two coin flips. If there is no mechanism connecting the two flips, we would typically expect that neither flip is affected by the outcome of the other. Similarly, if we take two die throws, we typically expect there is no mechanism connecting the throws and thus no reason to expect that one is affected by the outcome of the other. As a third example, consider the crime rate in London and the price of tea in Shanghai. There is no reason to expect one of these two events to affect the other event.[5] In each of these cases, we describe the events as independent.

This discussion implies that two unrelated (independent) events A and B will satisfy the properties $\mathbb{P}[A \mid B] = \mathbb{P}[A]$ and $\mathbb{P}[B \mid A] = \mathbb{P}[B]$. In words, the probability that a coin is H is unaffected by the outcome (H or T) of another coin. From the definition of conditional probability, this implies $\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$. Let us use this as the formal definition.

Definition 1.4 The events A and B are **statistically independent** if $\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$.

We typically use the simpler label **independent** for brevity. As an immediate consequence of the derivation, we obtain the following equivalence.

[5] Except in a James Bond movie.

Theorem 1.4 If A and B are independent with $\mathbb{P}[A] > 0$ and $\mathbb{P}[B] > 0$, then

$$\mathbb{P}[A \mid B] = \mathbb{P}[A]$$

$$\mathbb{P}[B \mid A] = \mathbb{P}[B].$$

Consider the stock index illustration in Section 1.6. We found that $\mathbb{P}[U_t \mid U_{t-1}] = 0.57$ and $\mathbb{P}[U_t \mid D_{t-1}] = 0.57$. This means that the probability of an increase is unaffected by the outcome from the previous week, which satisfies the definition of independence. It follows that the events U_t and U_{t-1} are independent.

When events are independent, then joint probabilities can be calculated by multiplying individual probabilities. Take two independent coin flips. Write the possible results of the first coin as $\{H_1, T_1\}$ and the possible results of the second coin as $\{H_2, T_2\}$. Let $p = \mathbb{P}[H_1]$ and $q = \mathbb{P}[H_2]$. We obtain the following chart for the joint probabilities.

Joint Probabilities: Independent Events

	H_1	T_1	
H_2	pq	$(1-p)q$	q
T_2	$p(1-q)$	$(1-p)(1-q)$	$1-q$
	p	$1-p$	1

The chart shows that the four joint probabilities are determined by p and q, the probabilities of the individual coins. The entries in each column sum to p and $1-p$, and the entries in each row sum to q and $1-q$.

If two events are not independent, we say that they are **dependent**. In this case, the joint event $A \cap B$ occurs at a different rate than predicted if the events were independent.

For example, consider wage rates and college degrees. We have already shown that the conditional probability of a high wage is affected by a college degree, which demonstrates that the two events are dependent. What we now do is see what happens when we calculate the joint probabilities from the individual probabilities under the (false) assumption of independence. The results are shown in the following chart.

Joint Probabilities: Wages and Education

	C	N	Any Education
H	0.11	0.20	0.31
L	0.25	0.44	0.69
Any Wage	0.36	0.64	1.00

The entries in the central box are obtained by multiplication of the individual probabilities (e.g., $\mathbb{P}[H \cap C] = 0.31 \times 0.36 = 0.11$). What we see is that the diagonal entries are much smaller, and the off-diagonal entries are much larger, than the corresponding correct joint probabilities. In this example, the joint events $H \cap C$ and $L \cap N$ occur more frequently than that predicted if wages and education were independent.

We can use independence to make probability calculations. Take the two-coin example. If two sequential fair coin flips are independent, then the probability that both are heads is

$$\mathbb{P}[H_1 \cap H_2] = \mathbb{P}[H_1] \times \mathbb{P}[H_2] = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

This addresses the issue raised in Section 1.5. The probability of HH is 1/4, not 1/3. The key is the assumption of independence, not how the outcomes are listed.

As another example, consider throwing a pair of fair dice. If the two dice are independent, then the probability of two 1's is $\mathbb{P}[1] \times \mathbb{P}[1] = 1/36$.

Naïvely, one might think that independence relates to disjoint events, but the converse is true. If A and B are disjoint, then they cannot be independent. That is, disjointness means $A \cap B = \varnothing$, and by property 2 of Theorem 1.2,

$$\mathbb{P}[A \cap B] = \mathbb{P}[\varnothing] = 0 \neq \mathbb{P}[A]\,\mathbb{P}[B]$$

and the right side is nonzero by the definition of independence.

Independence lies at the core of many probability calculations. If you can break an event into the joint occurance of several independent events, then the probability of the event is the product of the individual probabilities.

Take, for example, the two-coin example and the event $\{HH, HT\}$. This equals {First coin is H, Second coin is either H or T}. If the two coins are independent, this has probability

$$\mathbb{P}[H] \times \mathbb{P}[H \text{ or } T] = \frac{1}{2} \times 1 = \frac{1}{2}.$$

As a bit more complicated example, what is the probability of "rolling a seven" from a pair of dice, meaning that the two faces add to seven? We can calculate this as follows. Let (x, y) denote the outcomes from the two (ordered) dice. The following outcomes yield a seven: $\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$. The outcomes are disjoint. Thus by the third axiom, the probability of a seven is the sum

$$\mathbb{P}[7] = \mathbb{P}[1,6] + \mathbb{P}[2,5] + \mathbb{P}[3,4] + \mathbb{P}[4,3] + \mathbb{P}[5,2] + \mathbb{P}[6,1]\,.$$

Assume that the two dice are independent of one another, so the probabilities are products. For fair dice, the above expression equals

$$\begin{aligned}
&\mathbb{P}[1] \times \mathbb{P}[6] + \mathbb{P}[2] \times \mathbb{P}[5] + \mathbb{P}[3] \times \mathbb{P}[4] + \mathbb{P}[4] \times \mathbb{P}[3] + \mathbb{P}[5] \times \mathbb{P}[2] + \mathbb{P}[6] \times \mathbb{P}[1] \\
&= \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} \\
&= 6 \times \frac{1}{6^2} \\
&= \frac{1}{6}.
\end{aligned}$$

Now suppose that the dice are not fair. Suppose they are independent, but each is weighted so that the probability of a "1" is 2/6 and the probability of a "6" is 0. We revise the calculation to find

$$\begin{aligned}
&\mathbb{P}[1] \times \mathbb{P}[6] + \mathbb{P}[2] \times \mathbb{P}[5] + \mathbb{P}[3] \times \mathbb{P}[4] + \mathbb{P}[4] \times \mathbb{P}[3] + \mathbb{P}[5] \times \mathbb{P}[2] + \mathbb{P}[6] \times \mathbb{P}[1] \\
&= \frac{2}{6} \times \frac{0}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{0}{6} \times \frac{2}{6} \\
&= \frac{1}{9}.
\end{aligned}$$

1.9 LAW OF TOTAL PROBABILITY

An important relationship can be derived from the partitioning theorem (Theorem 1.1) which states that if $\{B_i\}$ is a partition of the sample space S, then

$$A = \bigcup_{i=1}^{\infty} (A \cap B_i) .$$

Since the events $(A \cap B_i)$ are disjoint, an application of the third axiom and the definition of conditional probability implies

$$\mathbb{P}[A] = \sum_{i=1}^{\infty} \mathbb{P}[A \cap B_i] = \sum_{i=1}^{\infty} \mathbb{P}[A \mid B_i] \mathbb{P}[B_i] .$$

This is called the Law of Total Probability.

Theorem 1.5 Law of Total Probability. If $\{B_1, B_2, \ldots\}$ is a partition of S, and $\mathbb{P}[B_i] > 0$ for all i, then

$$\mathbb{P}[A] = \sum_{i=1}^{\infty} \mathbb{P}[A \mid B_i] \mathbb{P}[B_i] .$$

For example, take the roll of a fair die and the events $A = \{1, 3, 5\}$ and $B_j = \{j\}$. We calculate that

$$\sum_{i=1}^{6} \mathbb{P}[A \mid B_i] \mathbb{P}[B_i] = 1 \times \frac{1}{6} + 0 \times \frac{1}{6} + 1 \times \frac{1}{6} + 0 \times \frac{1}{6} + 1 \times \frac{1}{6} + 0 \times \frac{1}{6} = \frac{1}{2},$$

which equals $\mathbb{P}[A] = 1/2$, as claimed.

1.10 BAYES RULE

A famous result is credited to Reverend Thomas Bayes.

Theorem 1.6 Bayes Rule. If $\mathbb{P}[A] > 0$ and $\mathbb{P}[B] > 0$, then

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[B \mid A] \mathbb{P}[A]}{\mathbb{P}[B \mid A] \mathbb{P}[A] + \mathbb{P}[B \mid A^c] \mathbb{P}[A^c]}.$$

Proof. The definition of conditional probability (applied twice) implies

$$\mathbb{P}[A \cap B] = \mathbb{P}[A \mid B] \mathbb{P}[B] = \mathbb{P}[B \mid A] \mathbb{P}[A] .$$

Solving, we find

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

Applying the law of total probability to $\mathbb{P}[B]$ using the partition $\{A, A_c\}$, we obtain the stated result. ■

Bayes Rule is terrifically useful in many contexts.

As one example, suppose you walk by a sports bar where you see a group of people watching a sports match which involves a popular local team. Suppose you suddenly hear a roar of excitement from the bar. Did

the local team just score? To investigate this by Bayes Rule, let $A = \{\text{score}\}$ and $B = \{\text{crowd roars}\}$. Assume that $\mathbb{P}[A] = 1/10$, $\mathbb{P}[B \mid A] = 1$, and $\mathbb{P}[B \mid A^c] = 1/10$ (there are other events which can cause a roar). Then

$$\mathbb{P}[A \mid B] = \frac{1 \times \frac{1}{10}}{1 \times \frac{1}{10} + \frac{1}{10} \times \frac{9}{10}} = \frac{10}{19} \simeq 53\%.$$

This is slightly over one-half. Under these assumptions, the roar of the crowd is informative though not definitive.[6]

As another example, suppose there are two types of workers: hard workers (H) and lazy workers (L). Suppose that we know from previous experience that $\mathbb{P}[H] = 1/4$ and $\mathbb{P}[L] = 3/4$. Suppose we can administer a screening test to determine whether an applicant is a hard worker. Let T be the event that an applicant has a high score on the test. Suppose that $\mathbb{P}[T \mid H] = 3/4$ and $\mathbb{P}[T \mid L] = 1/4$. That is, the test has some signal but is not perfect. We are interested in calculating $\mathbb{P}[H \mid T]$, the conditional probability that an applicant is a hard worker, given that they have a high test score. Bayes Rule tells us

$$\mathbb{P}[H \mid T] = \frac{\mathbb{P}[T \mid H]\,\mathbb{P}[H]}{\mathbb{P}[T \mid H]\,\mathbb{P}[H] + \mathbb{P}[T \mid L]\,\mathbb{P}[L]} = \frac{\frac{3}{4} \times \frac{1}{4}}{\frac{3}{4} \times \frac{1}{4} + \frac{1}{4} \times \frac{3}{4}} = \frac{1}{2}.$$

The probability the applicant is a hard worker is only 50%! Does this mean the test is useless? Consider the question: What is the probability an applicant is a hard worker, given that they had a poor (P) test score? We find

$$\mathbb{P}[H \mid P] = \frac{\mathbb{P}[P \mid H]\,\mathbb{P}[H]}{\mathbb{P}[P \mid H]\,\mathbb{P}[H] + \mathbb{P}[P \mid L]\,\mathbb{P}[L]} = \frac{\frac{1}{4} \times \frac{1}{4}}{\frac{1}{4} \times \frac{1}{4} + \frac{3}{4} \times \frac{3}{4}} = \frac{1}{10}.$$

This is only 10%. Thus what the test tells us is that if an applicant scores high, we are uncertain about that applicant's work habits; but if an applicant scores low, it is unlikely that they are a hard worker.

To revisit our real-world example of education and wages, recall that we calculated that the probability of a high wage (H) given a college degree (C) is $\mathbb{P}[H \mid C] = 0.53$. Applying Bayes Rule, we can find the probability that an individual has a college degree given that they have a high wage is

$$\mathbb{P}[C \mid H] = \frac{\mathbb{P}[H \mid C]\,\mathbb{P}[C]}{\mathbb{P}[H]} = \frac{0.53 \times 0.36}{0.31} = 0.62.$$

The probability of a college degree given that they have a low wage (L) is

$$\mathbb{P}[C \mid L] = \frac{\mathbb{P}[L \mid C]\,\mathbb{P}[C]}{\mathbb{P}[L]} = \frac{0.47 \times 0.36}{0.69} = 0.25.$$

Thus given this one piece of information (if the wage is above or below $25), we have probabilistic information about whether the individual has a college degree.

1.11 PERMUTATIONS AND COMBINATIONS

For some calculations, it is useful to count the number of individual outcomes. For some of these calculations, the concepts of counting rules, permutations, and combinations are useful.

The first definition we explore is the counting rule, which shows how to count options when we combine tasks. For example, suppose you own ten shirts, three pairs of jeans, five pairs of socks, four coats and two

[6] Consequently, it is reasonable to enter the sports bar to learn the truth!

hats. How many clothing outfits can you create, assuming you use one of each category? The answer is $10 \times 3 \times 5 \times 4 \times 2 = 1200$ distinct outfits.[7]

Theorem 1.7 **Counting Rule**. If a job consists of K separate tasks, the k^{th} of which can be done in n_k ways, then the entire job can be done in $n_1 n_2 \cdots n_K$ ways.

The counting rule is intuitively simple but is useful in a variety of modeling situations.

The second definition we explore is that of a permutation. A **permutation** is a rearrangement of the order. Suppose you take a classroom of 30 students. How many ways can you arrange their order? Each arrangement is called a "permutation." To calculate the number of permutations, observe that there are 30 students who can be placed first. Given this choice, there are 29 students who can be placed second. Given these two choices, there are 28 students for the third position, and so on. The total number of permutations is

$$30 \times 29 \times \cdots \times 1 = 30!$$

Here, the symbol ! denotes the factorial. (See Section A.3.)

The general solution is as follows.

Theorem 1.8 The number of **permutations** of a group of N objects is $N!$

Suppose we are trying to select an ordered five-student team from a 30-student class for a competition. How many ordered groups of five are there? The calculation is much the same as above, but we stop once the fifth position is filled. Thus the number is

$$30 \times 29 \times 28 \times 27 \times 26 = \frac{30!}{25!}.$$

The general solution is as follows.

Theorem 1.9 The number of **permutations** of a group of N objects taken K at a time is

$$P(N, K) = \frac{N!}{(N-K)!}.$$

The third definition we explore is that of a combination. A **combination** is an unordered group of objects. For example, revisit the idea of selecting a five-student team for a competition, but now assume that the team is unordered. Then the question is: How many five-member teams can we construct from a class of 30 students? In general, how many groups of K objects can be extracted from a group of N objects? We call this the "number of combinations".

The extreme cases are easy. If $K = 1$, then there are N combinations (each individual student). If $K = N$, then there is one combination (the entire class). The general answer can be found by noting that the number of ordered groups is the number of permutations $P(N, K)$. Each group of K can be ordered $K!$ ways (since this is the number of permutations of a group of K). Thus the number of unordered groups is $P(N, K)/K!$. We have found the following theorem.

Theorem 1.10 The number of **combinations** of a group of N objects taken K at a time is

$$\binom{N}{K} = \frac{N!}{K!\,(N-K)!}.$$

[7] Remember this when you (or a friend) asserts "I have nothing to wear!"

The symbol $\binom{N}{K}$, in words "N choose K", is a commonly used notation for combinations. They are also known as the **binomial coefficients**. The latter name is used because they are the coefficients from the binomial expansion.

Theorem 1.11 Binomial Theorem. For any integer $N \geq 0$,

$$(a+b)^N = \sum_{K=0}^{N} \binom{N}{K} a^K b^{N-K}.$$

The proof of the binomial theorem is given in Section 1.15.

The permutation and combination rules introduced in this section are useful in certain counting applications but may not be necessary for a general understanding of probability. My view is that the tools should be understood but not memorized. Instead, these tools can be looked up when needed.

1.12 SAMPLING WITH AND WITHOUT REPLACEMENT

Consider the problem of sampling from a finite set. For example, consider a $2 Powerball lottery ticket which consists of five integers each between 1 and 69. If all five numbers match the winning numbers, the player wins[8] $1 million!

To calculate the probability of winning the lottery, we need to count the number of potential tickets. The answer depends on two factors: (1) Can the numbers repeat? (2) Does the order matter? The number of tickets could have four distinct values, depending on the two choices just described.

The first question, of whether a number can repeat or not, is called "sampling with replacement" versus "sampling without replacement". In the actual Powerball game, 69 ping-pong balls are numbered and put in a rotating air machine with a small exit. As the balls bounce around, some of them find the exit. The first five to exit are the winning numbers. In this setting, we have "sampling without replacement", as once a ball exits, it is no longer among the remaining balls. A consequence for the lottery is that a winning ticket cannot have duplicate numbers. However, an alternative way to play the game would be to extract the first ball, replace it in the chamber, and repeat. This would be "sampling with replacement". In this game, a winning ticket could have repeated numbers.

The second question, of whether the order matters, is the same as the distinction between permutations and combinations as discussed in the previous section. In the case of the Powerball game, the balls emerge in a specific order. However, this order is ignored for the purpose of determining a winning ticket. This is the case of unordered sets. If the rules of the game were different, the order could matter. If so, we would use the tools of ordered sets.

We now describe the four sampling problems. We want to find the number of groups of size K which can be taken from N items, for example, the number of five integers taken from the set $\{1, \ldots, 69\}$.

Ordered, with replacement. Consider selecting the items in sequence. The first item can be any of the N, the second can be any of the N, the third can be any of the N, etc. So by the counting rule, the total number of possible groups is

$$N \times N \times \cdots \times N = N^K.$$

[8] There are also other prizes for other combinations.

In the Powerball example, this is

$$69^5 = 1{,}564{,}031{,}359.$$

This is a very large number of potential tickets!

Ordered, without replacement. This is the number of permutations $P(N, K) = N!/(N-K)!$ In the powerball example, this number is

$$\frac{69!}{(69-5)!} = \frac{69!}{64!} = 69 \times 68 \times 67 \times 66 \times 65 = 1{,}348{,}621{,}560.$$

This is nearly as large as the case with replacement.

Unordered, without replacement. This is the number of combinations $N!/(K!(N-K)!)$. In the powerball example, this number is

$$\frac{69!}{5!\,(69-5)!} = 11{,}238{,}513.$$

This is a large number but considerably smaller than the cases of ordered sampling.

Unordered, with replacement. This computation is tricky. It is not N^K (ordered with replacement) divided by $K!$, because the number of orderings per group depends on whether there are repeats. The trick is to recast the question as a different problem. It turns out that the number we are looking for is the same as the number of N-tuples of nonnegative integers $\{x_1, \ldots, x_N\}$ whose sum is K. To see this, a lottery ticket (unordered with replacement) can be represented by the number of "1's" x_1, the number of "2's" x_2, the number of "3's" x_3, and so forth, and we know that the sum of these numbers $(x_1 + \cdots + x_N)$ must equal K. The solution has a clever name based on the original proof notation.

Theorem 1.12 Stars and Bars Theorem. The number of N-tuples of nonnegative integers whose sum is K is equal to $\binom{N+K-1}{K}$.

The proof of the stars and bars theorem is omitted, as it is rather tedious. It does give us the answer to the question we started to address, namely, the number of unordered sets taken with replacement. In the Powerball example, this is

$$\binom{69+5-1}{5} = \frac{73!}{5!68!} = 15{,}020{,}334.$$

Table 1.1 summarizes the four sampling results.

Table 1.1
Number of possible arrangments of size K from N items

	Without Replacement	With Replacement
Ordered	$\frac{N!}{(N-K)!}$	N^K
Unordered	$\binom{N}{K}$	$\binom{N+K-1}{K}$

The actual Powerball game uses sampling that is unordered without replacement. Thus there are about 11 million potential tickets. As each ticket has an equal chance of occurring (if the random process is fair), this means the probability of winning is about 1/11,000,000. Since a player wins $1 million once for every 11 million tickets sold, the expected payout (ignoring the other payouts) is about $0.09. This is a low payout (considerably below a "fair" bet, given that a ticket costs $2) but is sufficiently high to attract meaningful interest from players.

1.13 POKER HANDS

A fun application of probability theory is to the game of poker. Similar types of calculations can be useful in economic examples involving multiple choices.

A standard game of poker is played with a 52-card deck containing 13 denominations {2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King, Ace} in each of four suits {club, diamond, heart, spade}. The deck is shuffled (so the order is random) and a player is dealt[9] five cards called a "hand". Hands are ranked based on whether there are multiple cards (pair, two pair, three-of-a-kind, full house, or four-of-a-kind), all five cards in sequence (called a "straight"), or all five cards of the same suit (called a "flush"). Players win if they have the best hand.

We are interested in calculating the probability of receiving a winning hand.

The structure is unordered sampling without replacement. The number of possible poker hands is

$$\binom{52}{5} = \frac{52!}{47!5!} = \frac{48 \times 49 \times 50 \times 51 \times 52}{2 \times 3 \times 4 \times 5} = 48 \times 49 \times 5 \times 17 \times 13 = 2,598,560.$$

Since the draws are symmetric and random, all hands have the same probability of receipt, implying that the probability of receiving any specific hand is 1/2, 598, 560, an infinitesimally small number.

Another way of calculating this probability is as follows. Imagine picking a specific five-card hand. The probability of receiving one of the five cards on the first draw is 5/52, the probability of receiving one of the remaining four on the second draw is 4/51, the probability of receiving one of the remaining three on the third draw is 3/50, etc., so the probability of receiving the five-card hand is

$$\frac{5 \times 4 \times 3 \times 2 \times 1}{52 \times 51 \times 50 \times 49 \times 48} = \frac{1}{13 \times 17 \times 5 \times 49 \times 48} = \frac{1}{2,598,960}.$$

One way to calculate the probability of a winning hand is to enumerate and count the number of winning hands in each category and then divide by the total number of hands, 2, 598, 560. Let us consider a few examples.

Four of a kind. Consider the number of hands with four of a specific denomination (such as Kings). The hand contains all four Kings plus an additional card, which can be any of the remaining 48. Thus there are exactly 48 five-card hands with all four Kings. There are 13 denominations, so there are $13 \times 48 = 624$ hands with four-of-a-kind. Thus the probability of drawing a four-of-a-kind is

$$\frac{13 \times 48}{13 \times 17 \times 5 \times 49 \times 48} = \frac{1}{17 \times 5 \times 49} = \frac{1}{4165} \simeq 0.0\%.$$

[9] A typical game involves additional complications, which we ignore.

Three of a kind. Consider the number of hands with three of a specific denomination (such as Aces). There are $\binom{4}{3} = 4$ groups of three Aces. There are 48 cards from which to choose the remaining two. The number of such arrangements is $\binom{48}{2} = \frac{48!}{46!2!} = 47 \times 24$. However, this includes pairs. There are twelve denominations each of which has $\binom{4}{2} = 6$ pairs, so there are $12 \times 6 = 72$ pairs. Thus the number of two-card arrangements excluding pairs is $47 \times 24 - 72 = 44 \times 24$. Hence the number of hands with three Aces and no pair is $4 \times 44 \times 24$. As there are 13 possible denominations, the number of hands with a three of a kind is $13 \times 4 \times 44 \times 24$. Thus the probability of drawing a three-of-a-kind is

$$\frac{13 \times 4 \times 44 \times 24}{13 \times 17 \times 5 \times 49 \times 48} = \frac{88}{17 \times 5 \times 49} \simeq 2.1\%.$$

One pair. Consider the number of hands with two of a specific denomination (such as a "7"). There are $\binom{4}{2} = 6$ pairs of 7's. From the 48 remaining cards, the number of three-card arrangements is $\binom{48}{3} = \frac{48!}{45!3!} = 23 \times 47 \times 16$. However, this includes three-card groups and two-card pairs. There are twelve denominations. Each has $\binom{4}{3} = 4$ three-card groups. Each also has $\binom{4}{2} = 6$ pairs and 44 remaining cards from which to select the third card. Thus there are $12 \times (4 + 6 \times 44)$ three-card arrangements with either a three-card group or a pair. Subtracting, we find that the number of hands with two 7's and no other pairs is

$$6 \times (23 \times 47 \times 16 - 12 \times (4 + 6 \times 44)).$$

Multiplying by 13, the probability of drawing one pair of any denomination is

$$13 \times \frac{6 \times (23 \times 47 \times 16 - 12 \times (4 + 6 \times 44))}{13 \times 17 \times 5 \times 49 \times 48} = \frac{23 \times 47 \times 2 - 3 \times (2 + 3 \times 44)}{17 \times 5 \times 49} \simeq 42\%.$$

From these simple calculations, you can see that if you receive a random hand of five cards, you have a good chance of receiving one pair, a small chance of receiving a three-of-a-kind, and a negligible chance of receiving a four-of-a-kind.

1.14 SIGMA FIELDS*

Definition 1.2 is incomplete as stated. When there are an uncountable infinity of events, it is necessary to restrict the set of allowable events to exclude pathological cases. This is a technicality which has little impact on practical econometrics. However, the terminology is used frequently, so it is prudent to be aware of the following definitions. The correct definition of probability is as follows.

Definition 1.5 A **probability function** $\mathbb{P}$ is a function from a sigma field $\mathscr{B}$ to the real line which satisfies the axioms of probability.

The difference is that Definition 1.5 restricts the domain to a sigma field $\mathscr{B}$. The latter is a collection of sets which is closed under set operations. The restriction means that there are some events for which probability is not defined.

A sigma field is defined as follows.

Definition 1.6 A collection $\mathscr{B}$ of sets is called a **sigma field** if it satisfies the following three properties:

1. $\varnothing \in \mathscr{B}$.
2. If $A \in \mathscr{B}$, then $A^c \in \mathscr{B}$.
3. If $A_1, A_2, \ldots \in \mathscr{B}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathscr{B}$.

The infinite union in part 3 includes all elements which are an element of A_i for some i. An example is $\bigcup_{i=1}^{\infty} [0, 1 - 1/i] = [0, 1)$.

An alternative label for a sigma field is "sigma algebra". The following is a leading example of a sigma field.

Definition 1.7 The **Borel sigma field** is the smallest sigma field on $\mathbb{R}$ containing all open intervals (a, b). It contains all open intervals and closed intervals, and their countable unions, intersections, and complements.

A sigma field can be **generated** from a finite collection of events by taking all unions, intersections, and complements. Take the coin-flip example and start with the event $\{H\}$. Its complement is $\{T\}$, their union is $S = \{H, T\}$, and the union's complement is $\{\varnothing\}$. No further events can be generated. Thus the collection $\{\{\varnothing\}, \{H\}, \{T\}, S\}$ is a sigma field.

For an example on the positive real line, take the sets $[0, 1]$ and $(1, 2]$. Their intersection is $\{\varnothing\}$, their union is $[0, 2]$, and their complements are $(1, \infty)$, $[0, 1] \cup (2, \infty)$, and $(2, \infty)$. A further union is $[0, \infty)$. This collection is a sigma field, as no further events can be generated.

When there are an infinite number of events, then it may not be possible to generate a sigma field through set operations, as pathological counterexamples exist. These counterexamples are difficult to characterize, are nonintuitive, and seem to have no practical implications for econometric practice. Therefore the issue is generally ignored in econometrics.

If the concept of a sigma field seems technical, it is! The concept is not used further in this textbook.

1.15 TECHNICAL PROOFS*

Proof of Theorem 1.1 Take an outcome ω in A. Since $\{B_1, B_2, \cdots\}$ is a partition of S, it follows that $\omega \in B_i$ for some i. Set $A_i = (A \cap B_i)$. Thus $\omega \in A_i \subset \bigcup_{i=1}^{\infty} A_i$. This shows that every element in A is an element of $\bigcup_{i=1}^{\infty} A_i$.

Now take an outcome ω in $\bigcup_{i=1}^{\infty} A_i$. Thus $\omega \in A_i$ for some i. This implies $\omega \in A$. This shows that every element in $\bigcup_{i=1}^{\infty} A_i$ is an element of A.

For $i \neq j$, $A_i \cap A_j = (A \cap B_i) \cap \left(A \cap B_j\right) = A \cap \left(B_i \cap B_j\right) = \varnothing$ since B_i are mutually disjoint. Thus A_i are mutually disjoint. ■

Proof of Theorem 1.2 property 1 A and A^c are disjoint and $A \cup A^c = S$. The second and third axioms imply

$$1 = \mathbb{P}[S] = \mathbb{P}[A] + \mathbb{P}\left[A^c\right]. \tag{1.1}$$

Rearranging, we find $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$ as claimed. ■

Proof of Theorem 1.2 property 2 We have that $\varnothing = S^c$. By Theorem 1.2 and the second axiom of probability, $\mathbb{P}[\varnothing] = 1 - \mathbb{P}[S] = 0$, as claimed. ■

Proof of Theorem 1.2 property 3 The first axiom implies $\mathbb{P}[A^c] \geq 0$. This and equation (1.1) imply

$$\mathbb{P}[A] = 1 - \mathbb{P}\left[A^c\right] \leq 1$$

as claimed. ■

Proof of Theorem 1.2 property 4 The assumption $A \subset B$ implies $A \cap B = A$. By the partitioning theorem (Theorem 1.1) $B = (B \cap A) \cup (B \cap A^c) = A \cup (B \cap A^c)$ where A and $B \cap A^c$ are disjoint. The third axiom implies

$$\mathbb{P}[B] = \mathbb{P}[A] + \mathbb{P}\left[B \cap A^c\right] \geq \mathbb{P}[A]$$

where the inequality is $\mathbb{P}[B \cap A^c] \geq 0$ which holds by the first axiom. Thus, $\mathbb{P}[B] \geq \mathbb{P}[A]$, as claimed. ■

Proof of Theorem 1.2 property 5 $\{A \cup B\} = A \cup \{B \cap A^c\}$ where A and $\{B \cap A^c\}$ are disjoint. Also $B = \{B \cap A\} \cup \{B \cap A^c\}$ where $\{B \cap A\}$ and $\{B \cap A^c\}$ are disjoint. These two relationships and the third axiom imply

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}\left[B \cap A^c\right]$$

$$\mathbb{P}[B] = \mathbb{P}[B \cap A] + \mathbb{P}\left[B \cap A^c\right].$$

Subtracting,

$$\mathbb{P}[A \cup B] - \mathbb{P}[B] = \mathbb{P}[A] - \mathbb{P}[B \cap A].$$

Rearranging, we obtain the result. ■

Proof of Theorem 1.2 property 6 From the Inclusion-Exclusion Principle and $\mathbb{P}[A \cap B] \geq 0$ (the first axiom)

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B] \leq \mathbb{P}[A] + \mathbb{P}[B]$$

as claimed. ■

Proof of Theorem 1.2 property 7 Rearranging the Inclusion-Exclusion Principle and using $\mathbb{P}[A \cup B] \leq 1$ (Theorem 1.2 property 3), we have

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cup B] \geq \mathbb{P}[A] + \mathbb{P}[B] - 1$$

which is the stated result. ■

Proof of Theorem 1.11 (Binomial Theorem) Multiplying out, the expression

$$(a+b)^N = (a+b) \times \cdots \times (a+b) \tag{1.2}$$

is a polynomial in a and b with 2^N terms. Each term takes the form of the product of K of the a and $N-K$ of the b, thus is of the form $a^K b^{N-K}$. The number of terms of this form is equal to the number of combinations of the a's, which is $\binom{N}{K}$. Consequently, expression (1.2) equals $\sum_{K=0}^{N} \binom{N}{K} a^K b^{N-K}$, as stated. ■

1.16 EXERCISES

Exercise 1.1 Let $A = \{a, b, c, d\}$ and $B = \{a, c, e, f\}$.

(a) Find $A \cap B$.

(b) Find $A \cup B$.

Exercise 1.2 Describe the sample space S for the following experiments.

(a) Flip a coin.
(b) Roll a six-sided die.
(c) Roll two six-sided dice.
(d) Shoot six free throws (in basketball).

Exercise 1.3 From a 52-card deck of playing cards, draw five cards to make a hand.

(a) Let A be the event "The hand has two Kings". Describe A^c.
(b) A **straight** is five cards in sequence, for example, $\{5, 6, 7, 8, 9\}$. A **flush** is five cards of the same suit. Let A be the event "The hand is a straight" and B be the event "The hand is 3-of-a-kind". Are A and B disjoint or not disjoint?
(c) Let A be the event "The hand is a straight" and B be the event "The hand is flush". Are A and B disjoint or not disjoint?

Exercise 1.4 For events A and B, express the probability of "either A or B but not both" as a formula in terms of $\mathbb{P}[A]$, $\mathbb{P}[B]$, and $\mathbb{P}[A \cap B]$.

Exercise 1.5 If $\mathbb{P}[A] = 1/2$ and $\mathbb{P}[B] = 2/3$, can A and B be disjoint? Explain.

Exercise 1.6 Prove that $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$.

Exercise 1.7 Show that $\mathbb{P}[A \cap B] \leq \mathbb{P}[A] \leq \mathbb{P}[A \cup B] \leq \mathbb{P}[A] + \mathbb{P}[B]$.

Exercise 1.8 Suppose $A \cap B = A$. Can A and B be independent? If so, give the appropriate condition.

Exercise 1.9 Prove that

$$\mathbb{P}[A \cap B \cap C] = \mathbb{P}[A \mid B \cap C]\, \mathbb{P}[B \mid C]\, \mathbb{P}[C].$$

Assume $\mathbb{P}[C] > 0$ and $\mathbb{P}[B \cap C] > 0$.

Exercise 1.10 Is $\mathbb{P}[A \mid B] \leq \mathbb{P}[A]$, $\mathbb{P}[A \mid B] \geq \mathbb{P}[A]$, or is neither necessarily true?

Exercise 1.11 Give an example where $\mathbb{P}[A] > 0$, yet $\mathbb{P}[A \mid B] = 0$.

Exercise 1.12 Calculate the following probabilities concerning a standard 52-card playing deck.

(a) Drawing a King with one card.
(b) Drawing a King on the second card, conditional on a King on the first card.
(c) Drawing two Kings with two cards.
(d) Drawing a King on the second card, conditional on the first card is not a King.
(e) Drawing a King on the second card, when the first card is placed face down (so is unknown).

Exercise 1.13 You are on a game show, and the host shows you five doors marked A, B, C, D, and E. The host says that a prize is behind one of the doors, and you win the prize if you select the correct door. Given the stated information, what probability distribution would you use for modeling the distribution of the correct door?

Exercise 1.14 Calculate the following probabilities, assuming fair coins and dice.

(a) Getting three heads in a row from three coin flips.
(b) Getting a heads given that the previous coin was a tails.
(c) From two coin flips getting two heads given that at least one coin is a heads.
(d) Rolling a six from a pair of dice.
(e) Rolling "snakes eyes" from a pair of dice. (Getting a pair of ones.)

Exercise 1.15 If four random cards are dealt from a deck of playing cards, what is the probability that all four are Aces?

Exercise 1.16 Suppose that the unconditional probability of a disease is 0.0025. A screening test for this disease has a detection rate of 0.9, and has a false positive rate of 0.01. Given that the screening test returns positive, what is the conditional probability of having the disease?

Exercise 1.17 Suppose that 1% of athletes use banned steroids. Suppose that a drug test has a detection rate of 40% and a false positive rate of 1%. If an athlete tests positive, what is the conditional probability that the athlete has taken banned steroids?

Exercise 1.18 Sometimes we use the concept of **conditional independence**. The definition is as follows. Let A, B, C be three events with positive probabilities. Then A and B are conditionally independent given C if $\mathbb{P}[A \cap B \mid C] = \mathbb{P}[A \mid C]\,\mathbb{P}[B \mid C]$. Consider the experiment of tossing two dice. Let $A = \{\text{First die is 6}\}$, $B = \{\text{Second die is 6}\}$, and $C = \{\text{Both dice are the same}\}$. Show that A and B are independent (unconditionally), but A and B are dependent given C.

Exercise 1.19 **Monte Hall**. This is a famous (and surprisingly difficult) problem based on an old U.S. television game show "Let's Make a Deal" hosted by Monte Hall. A standard part of the show ran as follows: A contestant was asked to select from one of three identical doors: A, B, and C. Behind one of the three doors was a prize. If the contestant selected the correct door, they would receive the prize. The contestant picked one door (say, A) but it is not immediately opened. To increase the drama, the host opened one of the two remaining doors (say, door B) revealing that that door does not have the prize. The host then made the offer: "You have the option to switch your choice" (e.g., to switch to door C). You can imagine that the contestant may have made one of reasonings (a)–(c) below. Comment on each of these three reasonings. Are they correct?

(a) "When I selected door A, the probability that it has the prize was 1/3. No information was revealed. So the probability that Door A has the prize remains 1/3."
(b) "The original probability was 1/3 on each door. Now that door B is eliminated, doors A and C each have each probability of 1/2. It does not matter whether I stay with A or switch to C."

(c) "The host inadvertently revealed information. If door C had the prize, he was forced to open door B. If door B had the prize, he would have been forced to open door C. Thus it is quite likely that door C has the prize."

(d) Assume a prior probability for each door of 1/3. Calculate the posterior probabilities that door A and door C have the prize, respectively. What choice do you recommend for the contestant?

Exercise 1.20 In the game of blackjack, you are dealt two cards from a standard playing deck. Your score is the sum of the value of the two cards, where numbered cards have the value given by their number, face cards (Jack, Queen, King) each receive 10 points, and an Ace either 1 or 11 (player can choose). A **blackjack** is receiving a score of 21 from two cards, thus an Ace and any card worth 10 points.

(a) What is the probability of receiving a blackjack?

(b) The dealer is dealt one of their cards face down and one face up. Suppose the "show" card is an Ace. What is the probability that the dealer has a blackjack? (For simplicity, assume you have not seen any other cards.)

Exercise 1.21 Consider drawing five cards at random from a standard deck of playing cards. Calculate the following probabilities.

(a) A straight (five cards in sequence, suit not relevant).

(b) A flush (five cards of the same suit, order not relevant).

(c) A full house (3-of-a-kind and a pair, e.g., three Kings and two "3's").

Exercise 1.22 In the poker game "Five Card Draw", a player first receives five cards drawn at random. The player decides to discard some of their cards and then receives replacement cards. Assume a player is dealt a hand with one pair and three unrelated cards and decides to discard the three unrelated cards to obtain replacements. Calculate the following conditional probabilities for the resulting hand after the replacements are made.

(a) Obtaining a four-of-a-kind.

(b) Obtaining a three-of-a-kind.

(c) Obtaining two pairs.

(d) Obtaining a straight or a flush.

(e) Ending with one pair.

INDEX