Part I

PROBABILITY

Chapter 1

Probability

1.1 Trials, Sample Spaces, and Events

The notions of trial, sample space, and event are fundamental to the study of probability theory. Tossing a coin, rolling a die, and choosing a card from a deck of cards are examples that are frequently used to explain basic concepts of probability. Each toss of the coin, roll of the die, or choice of a card is called a *trial* or *experiment*. We shall use the words trial and experiment interchangeably. Each execution of a trial is called a *realization* of the probability experiment.

At the end of any trial involving the examples given above, we are left with a head or a tail, an integer from one through six, or a particular card, perhaps the queen of hearts. The result of a trial is called an *outcome*. The set of all possible outcomes of a probability experiment is called the *sample space* and is denoted by Ω. The outcomes that constitute a sample space are also referred to as *sample points* or *elements*. We shall use ω to denote an element of the sample space.

Example 1.1 The sample space for coin tossing has two sample points, a head (H) and a tail (T). This gives $\Omega = \{H, T\}$, as shown in Figure 1.1.

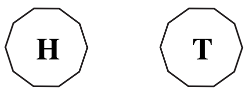

Figure 1.1. Sample space for tossing a coin has two elements $\{H, T\}$.

Example 1.2 For throwing a die, the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$, Figure 1.2, which represents the number of spots on the six faces of the die.

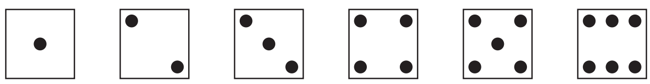

Figure 1.2. Sample space for throwing a die has six elements $\{1, 2, 3, 4, 5, 6\}$.

Example 1.3 For choosing a card, the sample space is a set consisting of 52 elements, one for each of the 52 cards in the deck, from the ace of spades through the king of hearts.

Example 1.4 If an experiment consists of three tosses of a coin, then the sample space is given by

$$\{HHH,\ HHT,\ HTH,\ THH,\ HTT,\ THT,\ TTH,\ TTT\}.$$

Notice that the element *HHT* is considered to be different from the elements *HTH* and *THH*, even though all three tosses give two heads and one tail. The position in which the tail occurs is important.

A sample space may be finite, denumerable (i.e., infinite but countable), or infinite. Its elements depend on the experiment and how the outcome of the experiment is defined. The four illustrative examples given above all have a finite number of elements.

Example 1.5 The sample space derived from an experiment that consists of observing the number of email messages received at a government office in one day may be taken to be denumerable. The sample space is denumerable since we may tag each arriving email message with a unique integer n that denotes the number of emails received prior to its arrival. Thus, $\Omega = \{n | n \in \mathcal{N}\}$, where $\mathcal{N}$ is the set of nonnegative integers.

Example 1.6 The sample space that arises from an experiment consisting of measuring the time one waits at a bus stop is infinite. Each outcome is a nonnegative real number x and the sample space is given by $\Omega = \{x | x \geq 0\}$.

If a finite number of trials is performed, then, no matter how large this number may be, there is no guarantee that every element of its sample space will be realized, even if the sample space itself is finite. This is a direct result of the essential probabilistic nature of the experiment. For example, it is possible, though perhaps not very likely (i.e., not very probable) that after a very large number of throws of the die, the number 6 has yet to appear.

Notice with emphasis that the sample space is a *set*, the set consisting of all the elements of the sample space, i.e., all the possible outcomes, associated with a given experiment. Since the sample space is a set, all permissible set operations can be performed on it. For example, the notion of subset is well defined and, for the coin tossing example, four subsets can be defined: the null subset ϕ; the subsets $\{H\}$, $\{T\}$; and the subset that contains all the elements, $\Omega = \{H, T\}$. The set of subsets of Ω is

$$\phi, \ \{H\}, \ \{T\}, \ \{H, \ T\}.$$

Events

The word *event* by itself conjures up the image of something having happened, and this is no different in probability theory. We toss a coin and get a head, we throw a die and get a five, we choose a card and get the ten of diamonds. Each experiment has an outcome, and in these examples, the outcome is an element of the sample space. These, the elements of the sample space, are called the *elementary events* of the experiment. However, we would like to give a broader meaning to the term event.

Example 1.7 Consider the event of tossing three coins and getting exactly two heads. There are three outcomes that allow for this event, namely, $\{HHT, HTH, THH\}$. The single tail appears on the third, second, or first toss, respectively.

Example 1.8 Consider the event of throwing a die and getting a prime number. Three outcomes allow for this event to occur, namely, $\{2, \ 3, \ 5\}$. This event comes to pass so long as the throw gives neither one, four, nor six spots.

In these last two examples, we have composed an event as a subset of the sample space, the subset $\{HHT, HTH, THH\}$ in the first case and the subset $\{2, \ 3, \ 5\}$ in the second. This is how we define an *event* in general. Rather than restricting our concept of an event to just another name for the elements of the sample space, we think of events as subsets of the sample space. In this case, the *elementary* events are the singleton subsets of the sample space, the subsets $\{H\}$, $\{5\}$, and {10 of diamonds}, for example. More complex events consist of subsets with more than one outcome. Defining an event as a subset of the sample space and not just as a subset that contains a single element provides us with much more flexibility and allows us to define much more general events.

The event is said “to occur” if and only if, the outcome of the experiment is any one of the elements of the subset that constitute the event. They are assigned names to help identify and manipulate them.

Example 1.9 Let $\mathcal{A}$ be the event that a throw of the die gives a number greater than 3. This event consists of the subset $\{4, 5, 6\}$ and we write $\mathcal{A} = \{4, 5, 6\}$. Event $\mathcal{A}$ occurs if the outcome of the trial, the number of spots obtained when the die is thrown, is any one of the numbers 4, 5, and 6. This is illustrated in Figure 1.3.

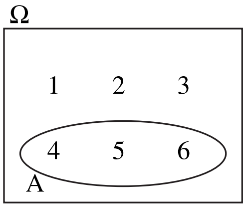

Figure 1.3. Event $\mathcal{A}$: "Throw a number greater than 3."

Example 1.10 Let $\mathcal{B}$ be the event that the chosen card is a 9. Event $\mathcal{B}$ is the subset containing four elements of the sample space: the 9 of spades, the 9 of clubs, the 9 of diamonds, and the 9 of hearts. Event $\mathcal{B}$ occurs if the card chosen is one of these four.

Example 1.11 The waiting time in minutes at a bus stop can be any nonnegative real number. The sample space is $\Omega = \{t \in \Re \mid t \geq 0\}$, and $\mathcal{A} = \{2 \leq t \leq 10\}$ is the event that the waiting time is between 2 and 10 minutes. Event $\mathcal{A}$ occurs if the wait is 2.1 minutes or 3.5 minutes or 9.99 minutes, etc.

To summarize, the standard definition of an event is a subset of the sample space. It consists of a set of outcomes. The null (or empty) subset, which contains none of the sample points, and the subset containing the entire sample space are legitimate events—the first is called the "null" or impossible event (it can never occur); the second is called the "universal" or certain event and is sure to happen no matter what the outcome of the experiments gives. The execution of a trial, or observation of an experiment, *must* yield one and only one of the outcomes in the sample space. If a subset contains *none* of these outcomes, the event it represents cannot happen; if a subset contains *all* of the outcomes, then the event it represents must happen. In general, for each outcome in the sample space, either the event occurs (if that particular outcome is in the defining subset of the event) or it does not occur.

Two events $\mathcal{A}$ and $\mathcal{B}$ defined on the same sample space are said to be *equivalent* or *identical* if $\mathcal{A}$ occurs if and only if $\mathcal{B}$ occurs. Events $\mathcal{A}$ and $\mathcal{B}$ may be specified differently, but the elements in their defining subsets are identical. In set terminology, two sets $\mathcal{A}$ and $\mathcal{B}$ are equal (written $\mathcal{A} = \mathcal{B}$) if and only if $\mathcal{A} \subset \mathcal{B}$ and $\mathcal{B} \subset \mathcal{A}$.

Example 1.12 Consider an experiment that consists of simultaneously throwing two dice. The sample space consists of all pairs of the form (i, j) for $i = 1, 2, \ldots, 6$ and $j = 1, 2, \ldots, 6$. Let $\mathcal{A}$ be the event that the sum of the number of spots obtained on the two dice is even, i.e., $i + j$ is an even number, and let $\mathcal{B}$ be the event that both dice show an even number of spots or both dice show an odd number of spots, i.e., i and j are even *or* i and j are odd. Although event $\mathcal{A}$ has been stated differently from $\mathcal{B}$, a moment's reflection should convince the reader that the sample points in both defining subsets must be exactly the same, and hence $\mathcal{A} = \mathcal{B}$.

Viewing events as subsets allows us to apply typical set operations to them, operations such as set union, set intersection, set complementation, and so on.

1. If $\mathcal{A}$ is an event, then the *complement* of $\mathcal{A}$, denoted $\mathcal{A}^c$, is also an event. $\mathcal{A}^c$ is the subset of all sample points of Ω that are *not* in $\mathcal{A}$. Event $\mathcal{A}^c$ occurs only if $\mathcal{A}$ does *not* occur.
2. The *union* of two events $\mathcal{A}$ and $\mathcal{B}$, denoted $\mathcal{A} \cup \mathcal{B}$, is the event consisting of all the sample points in $\mathcal{A}$ and in $\mathcal{B}$. It occurs if *either* $\mathcal{A}$ or $\mathcal{B}$ occurs.

3. The *intersection* of two events $\mathcal{A}$ and $\mathcal{B}$, denoted $\mathcal{A} \cap \mathcal{B}$, is also an event. It consists of the sample points that are in both $\mathcal{A}$ and $\mathcal{B}$ and occurs if *both* $\mathcal{A}$ and $\mathcal{B}$ occur.
4. The *difference* of two events $\mathcal{A}$ and $\mathcal{B}$, denoted by $\mathcal{A} - \mathcal{B}$, is the event that $\mathcal{A}$ occurs and $\mathcal{B}$ does not occur. It consists of the sample points that are in $\mathcal{A}$ but not in $\mathcal{B}$. This means that

$$\mathcal{A} - \mathcal{B} = \mathcal{A} \cap \mathcal{B}^c.$$

It follows that $\Omega - \mathcal{B} = \Omega \cap \mathcal{B}^c = \mathcal{B}^c$.

5. Finally, notice that if $\mathcal{B}$ is a subset of $\mathcal{A}$, i.e., $\mathcal{B} \subset \mathcal{A}$, then the event $\mathcal{B}$ implies the event $\mathcal{A}$. In other words, if $\mathcal{B}$ occurs, it must follow that $\mathcal{A}$ has also occurred.

Example 1.13 Let event $\mathcal{A}$ be "throw a number greater than 3" and let event $\mathcal{B}$ be "throw an odd number." Event $\mathcal{A}$ occurs if a 4, 5, or 6 is thrown, and event $\mathcal{B}$ occurs if a 1, 3, or 5 is thrown. Thus both events occur if a 5 is thrown (this is the event that is the *intersection* of events $\mathcal{A}$ and $\mathcal{B}$) and neither event occurs if a 2 is thrown (this is the event that is the *complement* of the union of $\mathcal{A}$ and $\mathcal{B}$). These are represented graphically in Figure 1.4. We have

$$\mathcal{A}^c = \{1, 2, 3\}; \quad \mathcal{A} \cup \mathcal{B} = \{1, 3, 4, 5, 6\}; \quad \mathcal{A} \cap \mathcal{B} = \{5\}; \quad \mathcal{A} - \mathcal{B} = \{4, 6\}.$$

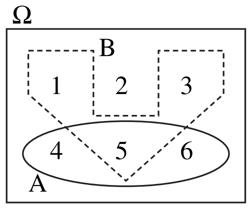

Figure 1.4. Two events on the die-throwing sample space.

Example 1.14 Or again, consider the card-choosing scenario. The sample space for the deck of cards contains 52 elements, each of which constitutes an elementary event. Now consider two events. Let event $\mathcal{A}$ be the subset containing the 13 elements corresponding to the diamond cards in the deck. Event $\mathcal{A}$ occurs if any one of these 13 cards is chosen. Let event $\mathcal{B}$ be the subset that contains the elements representing the four queens. This event occurs if one of the four queens is chosen. The event $\mathcal{A} \cup \mathcal{B}$ contains 16 elements, the 13 corresponding to the 13 diamonds plus the queens of spades, clubs, and hearts. The event $\mathcal{A} \cup \mathcal{B}$ occurs if any one of these 16 cards is chosen: i.e., if one of the 13 diamond cards is chosen *or* if one of the four queens is chosen (logical OR). On the other hand, the event $\mathcal{A} \cap \mathcal{B}$ has a single element, the element corresponding to the queen of diamonds. The event $\mathcal{A} \cap \mathcal{B}$ occurs only if a diamond card is chosen *and* that card is a queen (logical AND). Finally, the event $\mathcal{A} - \mathcal{B}$ occurs if any diamond card, *other than the queen of diamonds*, occurs.

Thus, as these examples show, the union of two events is also an event. It is the event that consists of all of the sample points in the two events. Likewise, the intersection of two events is the event that consists of the sample points that are simultaneously in both events. It follows that the union of an event and its complement is the universal event Ω, while the intersection of an event and its complement is the null event ϕ.

The definitions of union and intersection may be extended to more than two events. For n events $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$, they are denoted, respectively, by

$$\bigcup_{i=1}^{n} \mathcal{A}_i \quad \text{and} \quad \bigcap_{i=1}^{n} \mathcal{A}_i.$$

In the first case, the event $\bigcup_{i=1}^{n} \mathcal{A}_i$ occurs if any one of the events $\mathcal{A}_i$ occurs, while the second event, $\bigcap_{i=1}^{n} \mathcal{A}_i$ occurs only if all the events $\mathcal{A}_i$ occur. The entire logical algebra is available for use with events, to give an "algebra of events." Commutative, associative, and distributive laws, the laws of DeMorgan and so on, may be used to manipulate events. Some of the most important of these are as follows (where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are subsets of the universal set Ω):

Intersection	Union
$\mathcal{A} \cap \Omega = \mathcal{A}$	$\mathcal{A} \cup \Omega = \Omega$
$\mathcal{A} \cap \mathcal{A} = \mathcal{A}$	$\mathcal{A} \cup \mathcal{A} = \mathcal{A}$
$\mathcal{A} \cap \phi = \phi$	$\mathcal{A} \cup \phi = \mathcal{A}$
$\mathcal{A} \cap \mathcal{A}^c = \phi$	$\mathcal{A} \cup \mathcal{A}^c = \Omega$
$\mathcal{A} \cap (\mathcal{B} \cap \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) \cap \mathcal{C}$	$\mathcal{A} \cup (\mathcal{B} \cup \mathcal{C}) = (\mathcal{A} \cup \mathcal{B}) \cup \mathcal{C}$
$\mathcal{A} \cap (\mathcal{B} \cup \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{C})$	$\mathcal{A} \cup (\mathcal{B} \cap \mathcal{C}) = (\mathcal{A} \cup \mathcal{B}) \cap (\mathcal{A} \cup \mathcal{C})$
$(\mathcal{A} \cap \mathcal{B})^c = \mathcal{A}^c \cup \mathcal{B}^c$	$(\mathcal{A} \cup \mathcal{B})^c = \mathcal{A}^c \cap \mathcal{B}^c$

Venn diagrams can be used to illustrate these results and can be helpful in establishing proofs. For example, an illustration of DeMorgan's laws is presented in Figure 1.5.

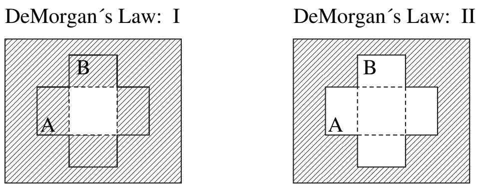

Figure 1.5. DeMorgan's laws: $(\mathcal{A} \cap \mathcal{B})^c = \mathcal{A}^c \cup \mathcal{B}^c$ and $(\mathcal{A} \cup \mathcal{B})^c = \mathcal{A}^c \cap \mathcal{B}^c$.

Mutually Exclusive and Collectively Exhaustive Events

When two events $\mathcal{A}$ and $\mathcal{B}$ contain no element of the sample space in common (i.e., $\mathcal{A} \cap \mathcal{B}$ is the null set), the events are said to be *mutually exclusive* or *incompatible*. The occurrence of one of them precludes the occurrence of the other. In the case of multiple events, $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ are mutually exclusive if and only if $\mathcal{A}_i \cap \mathcal{A}_j = \phi$ for all $i \neq j$.

Example 1.15 Consider the four events, $\mathcal{A}_1,\ \mathcal{A}_2,\ \mathcal{A}_3,\ \mathcal{A}_4$ corresponding to the four suits in a deck of cards, i.e., $\mathcal{A}_1$ contains the 13 elements corresponding to the 13 diamonds, $\mathcal{A}_2$ contains the 13 elements corresponding to the 13 hearts, etc. Then none of the sets $\mathcal{A}_1$ through $\mathcal{A}_4$ has any element in common. The four sets are mutually exclusive.

Example 1.16 Similarly, in the die-throwing experiment, if we choose $\mathcal{B}_1$ to be the event "throw a number greater than 5," $\mathcal{B}_2$ to be the event "throw an odd number," and $\mathcal{B}_3$ to be the event "throw a 2," then the events $\mathcal{B}_1$ through $\mathcal{B}_3$ are mutually exclusive.

In all cases, an event $\mathcal{A}$ and its complement $\mathcal{A}^c$ are mutually exclusive. In general, a *list of events* is said to be mutually exclusive if no element in their sample space is in more than one event. This is illustrated in Figure 1.6.

When all the elements in a sample space can be found in at least one event in a list of events, then the list of events is said to be *collectively exhaustive*. In this case, no element of the sample space is omitted and a single element may be in more than one event. This is illustrated in Figure 1.7.

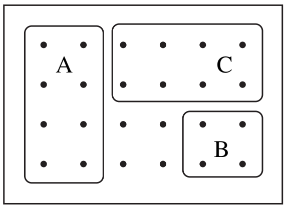

Figure 1.6. Events $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are mutually exclusive.

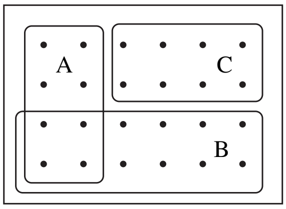

Figure 1.7. Events $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are collectively exhaustive.

Events that are both mutually exclusive and collectively exhaustive, such as those illustrated in Figure 1.8, are said to form a *partition* of the sample space. Additionally, the previously defined four events on the deck of cards, $\mathcal{A}_1,\ \mathcal{A}_2,\ \mathcal{A}_3,\ \mathcal{A}_4$, are both mutually exclusive and collectively exhaustive and constitute a partition of the sample space. Furthermore, since the elementary events (or outcomes) of a sample space are mutually exclusive and collectively exhaustive they too constitute a partition of the sample space. Any set of mutually exclusive and collectively exhaustive events is called an *event space*.

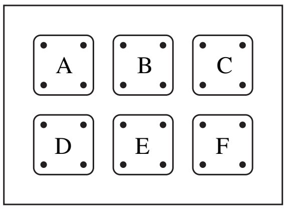

Figure 1.8. Events $\mathcal{A}$–$\mathcal{F}$ constitute a partition.

Example 1.17 Bit sequences are transmitted over a communication channel in groups of five. Each bit may be received correctly or else be modified in transit, which occasions an error. Consider an experiment that consists in observing the bit values as they arrive and identifying them with the letter c if the bit is correct and with the letter e if the bit is in error.

The sample space consists of 32 outcomes from *ccccc* through *eeeee*, from zero bits transmitted incorrectly to all five bits being in error. Let the event $\mathcal{A}_i,\ i = 0, 1, \ldots, 5$, consist of all outcomes in which i bits are in error. Thus $\mathcal{A}_0 = \{ccccc\}$, $\mathcal{A}_1 = \{ecccc, ceccc, ccecc, cccec, cccce\}$, and so on up to $\mathcal{A}_5 = \{eeeee\}$. The events $\mathcal{A}_i,\ i = 0, 1, \ldots, 5$, partition the sample space and therefore constitute an event space. It may be much easier to work in this small *event space* rather than in the larger *sample space*, especially if our only interest is in knowing the number of bits transmitted in error. Furthermore, when the bits are transmitted in larger groups, the difference becomes even more important. With 16 bits per group instead of five, the event space now contains 17 events, whereas the sample space contains 2^{16} outcomes.

1.2 Probability Axioms and Probability Space

Probability Axioms

So far our discussion has been about trials, sample spaces, and events. We now tackle the topic of probabilities. Our concern will be with assigning probabilities to events, i.e., providing some measure of the relative likelihood of the occurrence of the event. We realize that when we toss a fair coin, we have a 50–50 chance that it will give a head. When we throw a fair die, the chance of getting a 1 is the same as that of getting a 2, or indeed any of the other four possibilities. If a deck of cards is well shuffled and we pick a single card, there is a one in 52 chance that it will be the queen of hearts. What we have done in these examples is to associate probabilities with the elements of the sample space; more correctly, we have assigned probabilities to the elementary events, the events consisting of the singleton subsets of the sample space.

Probabilities are real numbers in the closed interval [0, 1]. The greater the value of the probability, the more likely the event is to happen. If an event has probability zero, that event cannot occur; if it has probability one, then it is certain to occur.

Example 1.18 In the coin-tossing example, the probability of getting a head in a single toss is 0.5, since we are equally likely to get a head as we are to get a tail. This is written as

$$\text{Prob}\{H\} = 0.5 \quad \text{or} \quad \text{Prob}\{\mathcal{A}_1\} = 0.5,$$

where $\mathcal{A}_1$ is the event $\{H\}$.

Similarly, the probability of throwing a 6 with a die is 1/6 and the probability of choosing the queen of hearts is 1/52. In these cases, the elementary events of each sample space all have equal probability, or equal likelihood, of being the outcome on any given trial. They are said to be *equiprobable* events and the outcome of the experiment is said to be *random*, since each event has the same chance of occurring. In a sample space containing n equally likely outcomes, the probability of any particular outcome occurring is $1/n$. Naturally, we can assign probabilities to events other than elementary events.

Example 1.19 Find the probability that should be associated with the event $\mathcal{A}_2 = \{1, 2, 3\}$, i.e., throwing a number smaller than 4 using a fair die. This event occurs if any of the numbers 1, 2, or 3 is the outcome of the throw. Since each has a probability of 1/6 and there are three of them, the probability of event $\mathcal{A}_2$ is the sum of the probabilities of these three elementary events and is therefore equal to 0.5.

This holds in general: the probability of any event is simply the sum of the probabilities associated with the (elementary) elements of the sample space that constitute that event.

Example 1.20 Consider Figure 1.6 once again (reproduced here as Figure 1.9), and assume that each of the 24 points or elements of the sample space is equiprobable.

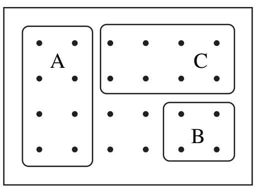

Figure 1.9. Sample space with 24 equiprobable elements.

Then event $\mathcal{A}$ contains eight elements, and so the probability of this event is

$$\text{Prob}\{\mathcal{A}\} = \frac{1}{24} + \frac{1}{24} + \frac{1}{24} + \frac{1}{24} + \frac{1}{24} + \frac{1}{24} + \frac{1}{24} + \frac{1}{24} = 8 \times \frac{1}{24} = \frac{1}{3}.$$

Similarly, $\text{Prob}\{\mathcal{B}\} = 4/24 = 1/6$ and $\text{Prob}\{\mathcal{C}\} = 8/24 = 1/3$.

Assigning probabilities to events is an extremely important part of developing probability models. In some cases, we know in advance the probabilities to associate with elementary events, while in other cases they must be estimated. If we assume that the coin and the die are fair and the deck of cards completely shuffled, then it is easy to associate probabilities with the elements of the sample space and subsequently to the events described on these sample spaces. In other cases, the probabilities must be guessed at or estimated.

Two approaches have been developed for defining probabilities: the *relative frequency* approach and the *axiomatic* approach. The first, as its name implies, consists in performing the probability experiment a great many times, say N, and counting the number of times a certain event occurs, say n. An estimate of the probability of the event may then be obtained as the relative frequency n/N with which the event occurs, since we would hope that, in the limit (limit in a probabilistic sense) as $N \to \infty$, the ratio n/N tends to the correct probability of the event. In mathematical terms, this is stated as follows: Given that the probability of an event is p, then

$$\lim_{N\to\infty} \text{Prob}\left\{ \left| \frac{n}{N} - p \right| > \epsilon \right\} = 0$$

for any small $\epsilon > 0$. In other words, no matter how small we choose ϵ to be, the probability that the difference between n/N and p is greater than ϵ tends to zero as $N \to \infty$. Use of relative frequencies as estimates of probability can be justified mathematically, as we shall see later.

The axiomatic approach sets up a small number of laws or axioms on which the entire theory of probability is based. Fundamental to this concept is the fact that it is possible to manipulate probabilities using the same logic algebra with which the events themselves are manipulated. The three basic axioms are as follows.

Axiom 1: For any event $\mathcal{A}$, $0 \le \text{Prob}\{\mathcal{A}\} \le 1$; i.e., probabilities are real numbers in the interval [0, 1].

Axiom 2: $\text{Prob}\{\Omega\} = 1$; The universal or certain event is assigned probability 1.

Axiom 3: For any *countable collection* of events $\mathcal{A}_1, \mathcal{A}_2, \ldots$ that are mutually exclusive,

$$\text{Prob}\left\{ \bigcup_{i=1}^{\infty} \mathcal{A}_i \right\} \equiv \text{Prob}\{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n \cup \cdots\} = \sum_{i=1}^{\infty} \text{Prob}\{\mathcal{A}_i\}.$$

In some elementary texts, the third axiom is replaced with the simpler

Axiom 3†: For two mutually exclusive events $\mathcal{A}$ and $\mathcal{B}$, $\text{Prob}\{\mathcal{A} \cup \mathcal{B}\} = \text{Prob}\{\mathcal{A}\} + \text{Prob}\{\mathcal{B}\}$

and a comment included stating that this extends in a natural sense to any finite or denumerable number of mutually exclusive events.

These three axioms are very natural; the first two are almost trivial, which essentially means that all of probability is based on unions of mutually exclusive events. To gain some insight, consider the following examples.

Example 1.21 If $\text{Prob}\{\mathcal{A}\} = p_1$ and $\text{Prob}\{\mathcal{B}\} = p_2$ where $\mathcal{A}$ and $\mathcal{B}$ are two mutually exclusive events, then the probability of the events $\mathcal{A} \cup \mathcal{B}$ and $\mathcal{A} \cap \mathcal{B}$ are given by

$$\text{Prob}\{\mathcal{A} \cup \mathcal{B}\} = p_1 + p_2 \quad \text{and} \quad \text{Prob}\{\mathcal{A} \cap \mathcal{B}\} = 0.$$

Example 1.22 If the sets $\mathcal{A}$ and $\mathcal{B}$ are not mutually exclusive, then the probability of the event $\mathcal{A} \cup \mathcal{B}$ will be less than $p_1 + p_2$ since some of the elementary events will be present in both $\mathcal{A}$ and $\mathcal{B}$, but can only be counted once. The probability of the event $\mathcal{A} \cap \mathcal{B}$ will be greater than zero; it will be the sum of the probabilities of the elementary events found in the intersection of the two subsets. It follows then that

$$\text{Prob}\{\mathcal{A} \cup \mathcal{B}\} = \text{Prob}\{\mathcal{A}\} + \text{Prob}\{\mathcal{B}\} - \text{Prob}\{\mathcal{A} \cap \mathcal{B}\}.$$

Observe that the probability of an event $\mathcal{A}$, formed from the union of a set of mutually exclusive events, is equal to the sum of the probabilities of those mutually exclusive events, i.e.,

$$\mathcal{A} = \text{Prob}\left\{\bigcup_{i=1}^{n} \mathcal{A}_i\right\} = \text{Prob}\{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n\} = \sum_{i=1}^{n} \text{Prob}\{\mathcal{A}_i\}.$$

In particular, the probability of *any* event is equal to the sum of the probabilities of the outcomes in the sample space that constitute the event since outcomes are elementary events which are mutually exclusive.

A number of the most important results that follow from these definitions are presented below. The reader should make an effort to prove these independently.

- For any event $\mathcal{A}$, $\text{Prob}\{\mathcal{A}^c\} = 1 - \text{Prob}\{\mathcal{A}\}$. Alternatively, $\text{Prob}\{\mathcal{A}\} + \text{Prob}\{\mathcal{A}^c\} = 1$.
- For the impossible event ϕ, $\text{Prob}\{\phi\} = 0$ (since $\text{Prob}\{\Omega\} = 1$).
- If $\mathcal{A}$ and $\mathcal{B}$ are any events, not necessarily mutually exclusive,

  $$\text{Prob}\{\mathcal{A} \cup \mathcal{B}\} = \text{Prob}\{\mathcal{A}\} + \text{Prob}\{\mathcal{B}\} - \text{Prob}\{\mathcal{A} \cap \mathcal{B}\}.$$

  Thus $\text{Prob}\{\mathcal{A} \cup \mathcal{B}\} \leq \text{Prob}\{\mathcal{A}\} + \text{Prob}\{\mathcal{B}\}$.
- For arbitrary events $\mathcal{A}$ and $\mathcal{B}$,

  $$\text{Prob}\{\mathcal{A} - \mathcal{B}\} = \text{Prob}\{\mathcal{A}\} - \text{Prob}\{\mathcal{A} \cap \mathcal{B}\}.$$

- For arbitrary events $\mathcal{A}$ and $\mathcal{B}$ with $\mathcal{B} \subset \mathcal{A}$,

  $$\text{Prob}\{\mathcal{B}\} \leq \text{Prob}\{\mathcal{A}\}.$$

It is interesting to observe that an event having probability zero does not necessarily mean that this event cannot occur. The probability of no heads appearing in an infinite number of throws of a fair coin is zero, but this event can occur.

Probability Space

The set of subsets of a given set, which includes the empty subset and the complete set itself, is sometimes referred to as the *superset* or *power set* of the given set. The superset of a set of elements in a sample space is therefore the set of all possible events that may be defined on that space. When the sample space is finite, or even when it is countably infinite (denumerable), it is possible to assign probabilities to each event in such a way that all three axioms are satisfied. However, when the sample space is not denumerable, such as the set of points on a segment of the real line, such an assignment of probabilities may not be possible. To avoid difficulties of this nature, we restrict the set of events to those to which probabilities satisfying all three axioms can be assigned. This is the basis of "measure theory:" for a given application, there is a particular family of events (a class of subsets of Ω), to which probabilities can be assigned, i.e., given a "measure." We shall call this family of subsets $\mathcal{F}$. Since we will wish to apply set operations, we need to insist that $\mathcal{F}$ be closed under countable unions, intersections, and complementation. A collection of subsets of a given set Ω that is closed under countable unions and complementation is called a σ*-field* of subsets of Ω.

The term σ*-algebra* is also used. Using DeMorgan's law, it may be shown that countable intersections of subsets of a σ-field $\mathcal{F}$ also lie in $\mathcal{F}$.

Example 1.23 The set $\{\Omega, \ \phi\}$ is the smallest σ-field defined on a sample space. It is sometimes called the trivial σ-field over Ω and is a subset of every other σ-field over Ω. The superset of Ω is the largest σ-field over Ω.

Example 1.24 If $\mathcal{A}$ and $\mathcal{B}$ are two events, then the set containing the events $\Omega, \ \phi, \ \mathcal{A}, \ \mathcal{A}^c$, $\mathcal{B}$, and $\mathcal{B}^c$ is a σ-field.

Example 1.25 In a die-rolling experiment having sample space $\{1, 2, 3, 4, 5, 6\}$, the following are all σ-fields:

$$\mathcal{F} = \{\Omega, \ \phi\},$$
$$\mathcal{F} = \{\Omega, \ \phi, \ \{2, 4, 6\}, \ \{1, 3, 5\}\},$$
$$\mathcal{F} = \{\Omega, \ \phi, \ \{1, 2, 4, 6\}, \ \{3, 5\}\},$$

but the sets $\{\Omega, \ \phi, \ \{1, 2\}, \ \{3, 4\}, \ \{5, 6\}\}$ and $\{\Omega, \ \phi, \ \{1, 2, 4, 6\}, \ \{3, 4, 5\}\}$ are not.

We may now define a *probability space* or *probability system*. This is defined as the triplet $\{\Omega, \mathcal{F}, \text{Prob}\}$, where Ω is a set, $\mathcal{F}$ is a σ-field of subsets of Ω that includes Ω, and Prob is a probability measure on $\mathcal{F}$ that satisfies the three axioms given above. Thus, $\text{Prob}\{\cdot\}$ is a function with domain $\mathcal{F}$ and range [0, 1] which satisfies axioms 1–3. It assigns a number in [0, 1] to events in $\mathcal{F}$.

1.3 Conditional Probability

Before performing a probability experiment, we cannot know precisely the particular outcome that will occur, nor whether an event $\mathcal{A}$, composed of some subset of the outcomes, will actually happen. We may know that the event is likely to take place, if $\text{Prob}\{\mathcal{A}\}$ is close to one, or unlikely to take place, if $\text{Prob}\{\mathcal{A}\}$ is close to zero, but we cannot be sure until after the experiment has been conducted. $\text{Prob}\{\mathcal{A}\}$ is the *prior probability* of $\mathcal{A}$. We now ask how this prior probability of an event $\mathcal{A}$ changes if we are informed that some other event, $\mathcal{B}$, has occurred. In other words, a probability experiment has taken place and one of the outcomes that constitutes an event $\mathcal{B}$ observed to have been the result. We are not told which particular outcome in $\mathcal{B}$ occurred, just that this event was observed to occur. We wish to know, given this additional information, how our knowledge of the probability of $\mathcal{A}$ occurring must be altered.

Example 1.26 Let us return to the example in which we consider the probabilities obtained on three throws of a fair coin. The elements of the sample space are

$$\{HHH, \ HHT, \ HTH, \ THH, \ HTT, \ THT, \ TTH, \ TTT\}$$

and the probability of each of these events is $1/8$. Suppose we are interested in the probability of getting three heads, $\mathcal{A} = \{HHH\}$. The prior probability of this event is $\text{Prob}\{\mathcal{A}\} = 1/8$. Now, how do the probabilities change if we know the result of the first throw?

If the first throw gives tails, the event $\mathcal{B}$ is constituted as $\mathcal{B} = \{THH, \ THT, \ TTH, \ TTT\}$ and we know that we are not going to get our three heads! Once we know that the result of the first throw is tails, the event of interest becomes impossible, i.e., has probability zero.

If the first throw gives a head, i.e., $\mathcal{B} = \{HHH, \ HHT, \ HTH, \ HTT\}$, then the event $\mathcal{A} = \{HHH\}$ is still possible. The question we are now faced with is to determine the probability of getting $\{HHH\}$ given that we know that the first throw gives a head. Obviously the probability must now be greater than 1/8. All we need to do is to get heads on the second and third throws, each of which is obtained with probability 1/2. Thus, given that the first throw yields heads, the probability of getting the event *HHH* is 1/4. Of the original eight elementary events, only four of them can now be assigned positive probabilities. From a different vantage point, the event $\mathcal{B}$ contains four equiprobable outcomes and is known to have occurred. It follows that the probability of any one of these four equiprobable outcomes, and in particular that of *HHH*, is 1/4.

The effect of knowing that a certain event has occurred changes the original probabilities of other events defined on the sample space. Some of these may become zero; for some others, their associated probability is increased. For yet others, there may be no change.

Example 1.27 Consider Figure 1.10 which represents a sample space with 24 elements all with probability 1/24. Suppose that we are told that event $\mathcal{B}$ has occurred. As a result the prior probabilities associated with elementary events outside $\mathcal{B}$ must be reset to zero and the sum of the probabilities of the elementary events inside $\mathcal{B}$ must sum to 1. In other words, the probabilities of the elementary events must be renormalized so that only those that can possibly occur have strictly positive probability and these probabilities must be coherent, i.e., they must sum to 1. Since the elementary events in $\mathcal{B}$ are equiprobable, after renormalization, they must each have probability $1/12$.

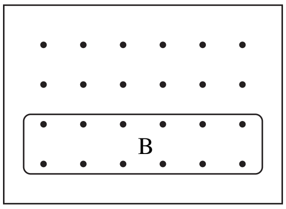

Figure 1.10. Sample space with 24 equiprobable elements.

We let $\text{Prob}\{\mathcal{A}|\mathcal{B}\}$ denote the probability of $\mathcal{A}$ given that event $\mathcal{B}$ has occurred. Because of the need to renormalize the probabilities so that they continue to sum to 1 after this given event has taken place, we must have

$$\text{Prob}\{\mathcal{A}|\mathcal{B}\} = \frac{\text{Prob}\{\mathcal{A} \cap \mathcal{B}\}}{\text{Prob}\{\mathcal{B}\}}. \tag{1.1}$$

Since it is known that event $\mathcal{B}$ occurred, it must have positive probability, i.e., $\text{Prob}\{\mathcal{B}\} > 0$, and hence the quotient in Equation (1.1) is well defined. The quantity $\text{Prob}\{\mathcal{A}|\mathcal{B}\}$ is called the *conditional probability of event $\mathcal{A}$ given the hypothesis $\mathcal{B}$*. It is defined only when $\text{Prob}\{\mathcal{B}\} \neq 0$. Notice that a rearrangement of Equation (1.1) gives

$$\text{Prob}\{\mathcal{A} \cap \mathcal{B}\} = \text{Prob}\{\mathcal{A}|\mathcal{B}\}\text{Prob}\{\mathcal{B}\}. \tag{1.2}$$

Similarly,

$$\text{Prob}\{\mathcal{A} \cap \mathcal{B}\} = \text{Prob}\{\mathcal{B}|\mathcal{A}\}\text{Prob}\{\mathcal{A}\}$$

provided that $\text{Prob}\{\mathcal{A}\} > 0$.

Since conditional probabilities are probabilities in the strictest sense of the term, they satisfy all the properties that we have seen so far concerning ordinary probabilities. In addition, the following hold:

- Let $\mathcal{A}$ and $\mathcal{B}$ be two mutually exclusive events. Then $\mathcal{A} \cap \mathcal{B} = \phi$ and hence $\text{Prob}\{\mathcal{A}|\mathcal{B}\} = 0$.
- If event $\mathcal{B}$ implies event $\mathcal{A}$, (i.e., $\mathcal{B} \subset \mathcal{A}$), then $\text{Prob}\{\mathcal{A}|\mathcal{B}\} = 1$.

Example 1.28 Let $\mathcal{A}$ be the event that a red queen is pulled from a deck of cards and let $\mathcal{B}$ be the event that a red card is pulled. Then Prob$\{\mathcal{A}|\mathcal{B}\}$, the probability that a red queen is pulled *given* that a red card is chosen, is

$$\text{Prob}\{\mathcal{A}|\mathcal{B}\} = \frac{\text{Prob}\{\mathcal{A} \cap \mathcal{B}\}}{\text{Prob}\{\mathcal{B}\}} = \frac{2/52}{1/2} = 1/13.$$

Notice in this example that Prob$\{\mathcal{A} \cap \mathcal{B}\}$ and Prob$\{\mathcal{B}\}$ are *prior probabilities*. Thus the event $\mathcal{A} \cap \mathcal{B}$ contains two of the 52 possible outcomes and the event $\mathcal{B}$ contains 26 of the 52 possible outcomes.

Example 1.29 If we observe Figure 1.11 we see that Prob$\{\mathcal{A} \cap \mathcal{B}\} = 1/6$, that Prob$\{\mathcal{B}\} = 1/2$, and that Prob$\{\mathcal{A}|\mathcal{B}\} = (1/6)/(1/2) = 1/3$ as expected. We know that $\mathcal{B}$ has occurred and that event $\mathcal{A}$ will occur if one of the four outcomes in $\mathcal{A} \cap \mathcal{B}$ is chosen from among the 12 equally probable outcomes in $\mathcal{B}$.

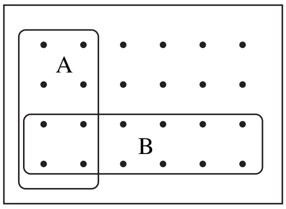

Figure 1.11. Prob$\{\mathcal{A}|\mathcal{B}\} = 1/3$.

Equation (1.2) can be generalized to multiple events. Let $\mathcal{A}_i,\ i = 1, 2, \ldots, k$, be k events for which Prob$\{\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_k\} > 0$. Then

$$\text{Prob}\{\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_k\} = \text{Prob}\{\mathcal{A}_1\}\text{Prob}\{\mathcal{A}_2|\mathcal{A}_1\}\text{Prob}\{\mathcal{A}_3|\mathcal{A}_1 \cap \mathcal{A}_2\} \cdots \\ \times \text{Prob}\{\mathcal{A}_k|\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_{k-1}\}.$$

The proof is by induction. The base clause ($k = 2$) follows from Equation (1.2):

$$\text{Prob}\{\mathcal{A}_1 \cap \mathcal{A}_2\} = \text{Prob}\{\mathcal{A}_1\}\text{Prob}\{\mathcal{A}_2|\mathcal{A}_1\}.$$

Now let $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_k$ and assume the relation is true for k, i.e., that

$$\text{Prob}\{\mathcal{A}\} = \text{Prob}\{\mathcal{A}_1\}\text{Prob}\{\mathcal{A}_2|\mathcal{A}_1\}\text{Prob}\{\mathcal{A}_3|\mathcal{A}_1 \cap \mathcal{A}_2\} \cdots \text{Prob}\{\mathcal{A}_k|\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_{k-1}\}.$$

That the relation is true for $k + 1$ follows immediately, since

$$\text{Prob}\{\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_k \cap \mathcal{A}_{k+1}\} = \text{Prob}\{\mathcal{A} \cap \mathcal{A}_{k+1}\} = \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{A}_{k+1}|\mathcal{A}\}.$$

Example 1.30 In a first-year graduate level class of 60 students, ten students are undergraduates. Let us compute the probability that three randomly chosen students are all undergraduates. We shall let $\mathcal{A}_1$ be the event that the first student chosen is an undergraduate student, $\mathcal{A}_2$ be the event that the second one chosen is an undergraduate, and so on. Recalling that the intersection of two events $\mathcal{A}$ and $\mathcal{B}$ is the event that occurs when both $\mathcal{A}$ and $\mathcal{B}$ occur, and using the relationship

$$\text{Prob}\{\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3\} = \text{Prob}\{\mathcal{A}_1\}\text{Prob}\{\mathcal{A}_2|\mathcal{A}_1\}\text{Prob}\{\mathcal{A}_3|\mathcal{A}_1 \cap \mathcal{A}_2\},$$

we obtain

$$\text{Prob}\{\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3\} = \frac{10}{60} \times \frac{9}{59} \times \frac{8}{58} = 0.003507.$$

1.4 Independent Events

We saw previously that two events are *mutually exclusive* if and only if the probability of the *union* of these two events is equal to the *sum* of the probabilities of the events, i.e., if and only if

$$\text{Prob}\{\mathcal{A} \cup \mathcal{B}\} = \text{Prob}\{\mathcal{A}\} + \text{Prob}\{\mathcal{B}\}.$$

Now we investigate the probability associated with the *intersection* of two events. We shall see that the probability of the intersection of two events is equal to the *product* of the probabilities of the events if and only if the outcome of one event does not influence the outcome of the other, i.e., if and only if the two events are *independent* of each other.

Let $\mathcal{B}$ be an event with positive probability, i.e., $\text{Prob}\{\mathcal{B}\} > 0$. Then event $\mathcal{A}$ is said to be *independent* of event $\mathcal{B}$ if

$$\text{Prob}\{\mathcal{A}|\mathcal{B}\} = \text{Prob}\{\mathcal{A}\}. \tag{1.3}$$

Thus the fact that event $\mathcal{B}$ occurs with positive probability has no effect on event $\mathcal{A}$. Equation (1.3) essentially says that the probability of event $\mathcal{A}$ occurring, given that $\mathcal{B}$ has already occurred, is just the same as the unconditional probability of event $\mathcal{A}$ occurring. It makes no difference at all that event $\mathcal{B}$ has occurred.

Example 1.31 Consider an experiment that consists in rolling two colored (and hence distinguishable) dice, one red and one green. Let $\mathcal{A}$ be the event that the sum of spots obtained is 7, and let $\mathcal{B}$ be the event that the red die shows 3. There are a total of 36 outcomes, each represented as a pair (i, j), where i denotes the number of spots on the red die, and j the number of spots on the green die. Of these 36 outcomes, six, namely, (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1), result in event $\mathcal{A}$ and hence $\text{Prob}\{\mathcal{A}\} = 6/36$. Also six outcomes result in the occurrence of event $\mathcal{B}$, namely, (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), but only one of these gives event $\mathcal{A}$. Therefore $\text{Prob}\{\mathcal{A}|\mathcal{B}\} = 1/6$. Events $\mathcal{A}$ and $\mathcal{B}$ must therefore be independent since

$$\text{Prob}\{\mathcal{A}|\mathcal{B}\} = 1/6 = 6/36 = \text{Prob}\{\mathcal{A}\}.$$

If event $\mathcal{A}$ is independent of event $\mathcal{B}$, then event $\mathcal{B}$ must be independent of event $\mathcal{A}$; i.e., independence is a *symmetric* relationship. Substituting $\text{Prob}\{\mathcal{A}\cap\mathcal{B}\}/\text{Prob}\{\mathcal{B}\}$ for $\text{Prob}\{\mathcal{A}|\mathcal{B}\}$ it must follow that, for independent events

$$\text{Prob}\{\mathcal{A}|\mathcal{B}\} = \frac{\text{Prob}\{\mathcal{A} \cap \mathcal{B}\}}{\text{Prob}\{\mathcal{B}\}} = \text{Prob}\{\mathcal{A}\}$$

or, rearranging terms, that

$$\text{Prob}\{\mathcal{A} \cap \mathcal{B}\} = \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}\}.$$

Indeed, this is frequently taken as a definition of independence. Two events $\mathcal{A}$ and $\mathcal{B}$ are said to be *independent* if and only if

$$\text{Prob}\{\mathcal{A} \cap \mathcal{B}\} = \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}\}.$$

Pursuing this direction, it then follows that, for two independent events

$$\text{Prob}\{\mathcal{A}|\mathcal{B}\} = \frac{\text{Prob}\{\mathcal{A} \cap \mathcal{B}\}}{\text{Prob}\{\mathcal{B}\}} = \frac{\text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}\}}{\text{Prob}\{\mathcal{B}\}} = \text{Prob}\{\mathcal{A}\},$$

which conveniently brings us back to the starting point.

Example 1.32 Suppose a fair coin is thrown twice. Let $\mathcal{A}$ be the event that a head occurs on the first throw, and $\mathcal{B}$ the event that a head occurs on the second throw. Are $\mathcal{A}$ and $\mathcal{B}$ independent events?

Obviously $\text{Prob}\{\mathcal{A}\} = 1/2 = \text{Prob}\{\mathcal{B}\}$. The event $\mathcal{A} \cap \mathcal{B}$ is the event of a head occurring on the first throw *and* a head occurring on the second throw. Thus, $\text{Prob}\{\mathcal{A} \cap \mathcal{B}\} = 1/2 \times 1/2 = 1/4$ and since $\text{Prob}\{\mathcal{A}\} \times \text{Prob}\{\mathcal{B}\} = 1/4$, the events $\mathcal{A}$ and $\mathcal{B}$ must be independent, since

$$\text{Prob}\{\mathcal{A} \cap \mathcal{B}\} = 1/4 = \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}\}.$$

Example 1.33 Let $\mathcal{A}$ be the event that a card pulled randomly from a deck of 52 cards is red, and let $\mathcal{B}$ be the event that this card is a queen. Are $\mathcal{A}$ and $\mathcal{B}$ independent events? What happens if event $\mathcal{B}$ is the event that the card pulled is the queen of hearts?

The probability of pulling a red card is 1/2 and the probability of pulling a queen is 1/13. Thus $\text{Prob}\{\mathcal{A}\} = 1/2$ and $\text{Prob}\{\mathcal{B}\} = 1/13$. Now let us find the probability of the event $\mathcal{A} \cap \mathcal{B}$ (the probability of pulling a red queen) and see if it equals the product of these two. Since there are two red queens, the probability of choosing a red queen is 2/52, which is indeed equal to the product of $\text{Prob}\{\mathcal{A}\}$ and $\text{Prob}\{\mathcal{B}\}$ and so the events are independent.

If event $\mathcal{B}$ is now the event that the card pulled is the queen of hearts, then $\text{Prob}\{\mathcal{B}\} = 1/52$. But now the event $\mathcal{A} \cap \mathcal{B}$ consists of a single outcome: there is only one red card that is the queen of hearts, and so $\text{Prob}\{\mathcal{A} \cap \mathcal{B}\} = 1/52$. Therefore the two events are *not* independent since

$$1/52 = \text{Prob}\{\mathcal{A} \cap \mathcal{B}\} \neq \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}\} = 1/2 \times 1/52.$$

We may show that, if $\mathcal{A}$ and $\mathcal{B}$ are independent events, then the pairs $(\mathcal{A}, \mathcal{B}^c)$, $(\mathcal{A}^c, \mathcal{B})$, and $(\mathcal{A}^c, \mathcal{B}^c)$ are also independent. For example, to show that $\mathcal{A}$ and $\mathcal{B}^c$ are independent, we proceed as follows. Using the result, $\text{Prob}\{\mathcal{A}\} = \text{Prob}\{\mathcal{A} \cap \mathcal{B}\} + \text{Prob}\{\mathcal{A} \cap \mathcal{B}^c\}$ we obtain

$$\begin{aligned}\text{Prob}\{\mathcal{A} \cap \mathcal{B}^c\} &= \text{Prob}\{\mathcal{A}\} - \text{Prob}\{\mathcal{A} \cap \mathcal{B}\}\\ &= \text{Prob}\{\mathcal{A}\} - \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}\}\\ &= \text{Prob}\{\mathcal{A}\}(1 - \text{Prob}\{\mathcal{B}\}) = \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}^c\}.\end{aligned}$$

The fact that, given two independent events $\mathcal{A}$ and $\mathcal{B}$, the four events $\mathcal{A}$, $\mathcal{B}$, $\mathcal{A}^c$, and $\mathcal{B}^c$ are *pairwise independent*, has a number of useful applications.

Example 1.34 Before being loaded onto a distribution truck, packages are subject to two independent tests, to ensure that the truck driver can safely handle them. The weight of the package must not exceed 80 lbs and the sum of the three dimensions must be less than 8 feet. It has been observed that 5% of packages exceed the weight limit and 2% exceed the dimension limit. What is the probability that a package that meets the weight requirement fails the dimension requirement?

The sample space contains four possible outcomes: (ws, ds), (wu, ds), (ws, du), and (wu, du), where w and d represent weight and dimension, respectively, and s and u represent satisfactory and unsatisfactory, respectively. Let $\mathcal{A}$ be the event that a package satisfies the weight requirement, and $\mathcal{B}$ the event that it satisfies the dimension requirement. Then $\text{Prob}\{\mathcal{A}\} = 0.95$ and $\text{Prob}\{\mathcal{B}\} = 0.98$. We also have $\text{Prob}\{\mathcal{A}^c\} = 0.05$ and $\text{Prob}\{\mathcal{B}^c\} = 0.02$.

The event of interest is the single outcome $\{(ws, du)\}$, which is given by $\text{Prob}\{\mathcal{A} \cap \mathcal{B}^c\}$. Since $\mathcal{A}$ and $\mathcal{B}$ are independent, it follows that $\mathcal{A}$ and $\mathcal{B}^c$ are independent and hence

$$\text{Prob}\{(ws, du)\} = \text{Prob}\{\mathcal{A} \cap \mathcal{B}^c\} = \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}^c\} = 0.95 \times 0.02 = 0.0019.$$

Multiple Independent Events

Consider now multiple events. Let $\mathcal{Z}$ be an arbitrary class of events, i.e.,

$$\mathcal{Z} = \mathcal{A}_1,\ \mathcal{A}_2, \ldots,\ \mathcal{A}_n, \ldots.$$

These events are said to be *mutually independent* (or simply independent), if, for every finite subclass $\mathcal{A}_1,\ \mathcal{A}_2, \ldots,\ \mathcal{A}_k$ of $\mathcal{Z}$,

$$\text{Prob}\{\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_k\} = \text{Prob}\{\mathcal{A}_1\}\text{Prob}\{\mathcal{A}_2\} \cdots \text{Prob}\{\mathcal{A}_k\}.$$

In other words, any pair of events $(\mathcal{A}_i,\ \mathcal{A}_j)$ must satisfy

$$\text{Prob}\{\mathcal{A}_i \cap \mathcal{A}_j\} = \text{Prob}\{\mathcal{A}_i\}\text{Prob}\{\mathcal{A}_j\};$$

any triplet of events $(\mathcal{A}_i,\ \mathcal{A}_j,\ \mathcal{A}_k)$ must satisfy

$$\text{Prob}\{\mathcal{A}_i \cap \mathcal{A}_j \cap \mathcal{A}_k\} = \text{Prob}\{\mathcal{A}_i\}\text{Prob}\{\mathcal{A}_j\}\text{Prob}\{\mathcal{A}_k\};$$

and so on, for quadruples of events, for quintuples of events, etc.

Example 1.35 The following example shows the need for this definition. Figure 1.12 shows a sample space with 16 equiprobable elements and on which three events $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, each with probability 1/2, are defined. Also, observe that

$$\begin{aligned}\text{Prob}\{\mathcal{A} \cap \mathcal{B}\} &= \text{Prob}\{\mathcal{A} \cap \mathcal{C}\} = \text{Prob}\{\mathcal{B} \cap \mathcal{C}\}\\ &= \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}\} = \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{C}\} = \text{Prob}\{\mathcal{B}\}\text{Prob}\{\mathcal{C}\} = 1/4\end{aligned}$$

and hence $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are *pairwise* independent.

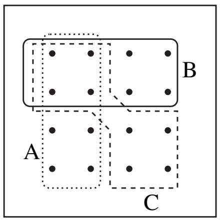

Figure 1.12. Sample space with 16 equiprobable elements.

However, they are not *mutually* independent since

$$\text{Prob}\{\mathcal{C}|\mathcal{A} \cap \mathcal{B}\} = \frac{\text{Prob}\{\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}\}}{\text{Prob}\{\mathcal{A} \cap \mathcal{B}\}} = \frac{1/4}{1/4} = 1 \neq \text{Prob}\{\mathcal{C}\}.$$

Alternatively,

$$1/4 = \text{Prob}\{\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}\} \neq \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}\}\text{Prob}\{\mathcal{C}\} = 1/8.$$

In conclusion, we say that the three events $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ defined above, are *not* independent; they are simply pairwise independent. Events $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are *mutually* independent only if *all* the following conditions hold:

$$\begin{aligned}\text{Prob}\{\mathcal{A} \cap \mathcal{B}\} &= \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}\},\\ \text{Prob}\{\mathcal{A} \cap \mathcal{C}\} &= \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{C}\},\\ \text{Prob}\{\mathcal{B} \cap \mathcal{C}\} &= \text{Prob}\{\mathcal{B}\}\text{Prob}\{\mathcal{C}\},\\ \text{Prob}\{\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}\} &= \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}\}\text{Prob}\{\mathcal{C}\}.\end{aligned}$$

Example 1.36 Consider a sample space that contains four equiprobable outcomes denoted a, b, c, and d. Define three events on this sample space as follows: $\mathcal{A} = \{a, b\}$, $\mathcal{B} = \{a, b, c\}$, and $\mathcal{C} = \phi$. This time

$$\text{Prob}\{\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}\} = 0 \quad \text{and} \quad \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}\}\text{Prob}\{\mathcal{C}\} = 1/2 \times 3/4 \times 0 = 0$$

but

$$1/2 = \text{Prob}\{\mathcal{A} \cap \mathcal{B}\} \neq \text{Prob}\{\mathcal{A}\}\text{Prob}\{\mathcal{B}\} = 1/2 \times 3/4.$$

The events $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are not independent, nor even pairwise independent.

1.5 Law of Total Probability

If $\mathcal{A}$ is any event, then it is known that the intersection of $\mathcal{A}$ and the universal event Ω is $\mathcal{A}$. It is also known that an event $\mathcal{B}$ and its complement $\mathcal{B}^c$ constitute a partition. Thus

$$\mathcal{A} = \mathcal{A} \cap \Omega \quad \text{and} \quad \mathcal{B} \cup \mathcal{B}^c = \Omega.$$

Substituting the second of these into the first and then applying DeMorgan's law, we find

$$\mathcal{A} = \mathcal{A} \cap (\mathcal{B} \cup \mathcal{B}^c) = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{B}^c). \tag{1.4}$$

Notice that the events $(\mathcal{A} \cap \mathcal{B})$ and $(\mathcal{A} \cap \mathcal{B}^c)$ are mutually exclusive. This is illustrated in Figure 1.13 which shows that, since $\mathcal{B}$ and $\mathcal{B}^c$ cannot have any outcomes in common, the *intersection* of $\mathcal{A}$ and $\mathcal{B}$ cannot have any outcomes in common with the *intersection* of $\mathcal{A}$ and $\mathcal{B}^c$.

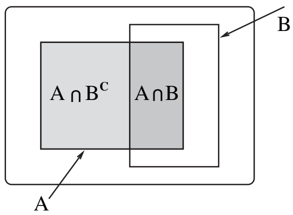

Figure 1.13. Events $(\mathcal{A} \cap \mathcal{B})$ and $(\mathcal{A} \cap \mathcal{B}^c)$ are mutually exclusive.

Returning to Equation (1.4), using the fact that $(\mathcal{A} \cap \mathcal{B})$ and $(\mathcal{A} \cap \mathcal{B}^c)$ are mutually exclusive, and applying Axiom 3, we obtain

$$\text{Prob}\{\mathcal{A}\} = \text{Prob}\{\mathcal{A} \cap \mathcal{B}\} + \text{Prob}\{\mathcal{A} \cap \mathcal{B}^c\}.$$

This means that to evaluate the probability of the event $\mathcal{A}$, it is sufficient to find the probabilities of the intersection of $\mathcal{A}$ with $\mathcal{B}$ and $\mathcal{A}$ with $\mathcal{B}^c$ and to add them together. This is frequently easier than trying to find the probability of $\mathcal{A}$ by some other method.

The same rule applies for any partition of the sample space and not just a partition defined by an event and its complement. Recall that a partition is a set of events that are mutually exclusive and collectively exhaustive. Let the n events $\mathcal{B}_i,\ i = 1, 2, \ldots, n$, be a partition of the sample space Ω. Then, for any event $\mathcal{A}$, we can write

$$\text{Prob}\{\mathcal{A}\} = \sum_{i=1}^{n} \text{Prob}\{\mathcal{A} \cap \mathcal{B}_i\}, \quad n \geq 1.$$

This is the *law of total probability*. To show that this law must hold, observe that the sets $\mathcal{A} \cap \mathcal{B}_i,\ i = 1, 2, \ldots, n$, are mutually exclusive (since the $\mathcal{B}_i$ are) and the fact that $\mathcal{B}_i,\ i = 1, 2, \ldots, n$, is a partition of Ω implies that

$$\mathcal{A} = \bigcup_{i=1}^{n} \mathcal{A} \cap \mathcal{B}_i, \quad n \geq 1.$$

Hence, using Axiom 3,

$$\text{Prob}\{\mathcal{A}\} = \text{Prob}\left\{ \bigcup_{i=1}^{n} \mathcal{A} \cap \mathcal{B}_i \right\} = \sum_{i=1}^{n} \text{Prob}\{\mathcal{A} \cap \mathcal{B}_i\}.$$

Example 1.37 As an illustration, consider Figure 1.14, which shows a partition of a sample space containing 24 equiprobable outcomes into six events, $\mathcal{B}_1$ through $\mathcal{B}_6$.