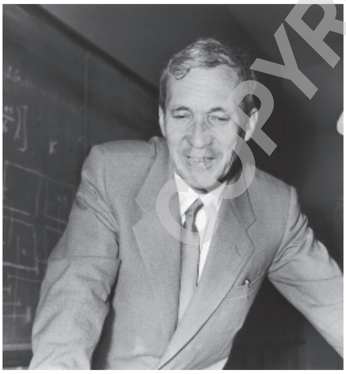
CHAPTER 1

THE CONCEPT OF PROBABILITY

Andrey Nikolaevich Kolmogorov
(Tambov, Russia 1903–Moscow 1987)
***Source*: Keystone-France / Getty Images.**



Regarded as the founder of modern probability theory, Kolmogorov was a Soviet mathematician whose work was also influential in several other scientific areas, notably in topology, constructive logic, classical mechanics, mathematical ecology, and algorithmic information theory.

He earned his Doctor of Philosophy (PhD) degree from Moscow State University in 1929, and two years later, he was appointed a professor in that university. In his book, *Foundations of the Theory of Probability*, which was published in 1933 and which remains a classic text to this day, he built up probability theory from fundamental axioms in a rigorous manner, comparable to Euclid's axiomatic development of geometry.

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

1.1 CHANCE EXPERIMENTS – SAMPLE SPACES

In this chapter, we present the main ideas and the theoretical background to understand what probability is and provide some illustrations of the way it is used to tackle problems in everyday life. It is rather difficult to try to answer the question "what is probability?" in a single sentence. However, from our experience and the use of this word in common language, we understand that it is a way to deal with uncertainty in our lives. In fact, probability theory has been referred to as "the science of uncertainty"; although intuitively most people associate probability with the degree of belief that something may happen, probability theory goes far beyond that as it attempts to formalize uncertainty in a way that is universally accepted and is also subject to rigorous mathematical treatment.

Since the idea of uncertainty is paramount when we discuss probability, we shall first introduce a concept that is broad enough to deal with uncertainties in a wide-ranging context when we consider practical applications. A **chance experiment** or a **random experiment** is any process which leads to an outcome that is not known beforehand. So tossing a coin, selecting a person at random and asking their age, or testing the lifetime of a new machine are all examples of random experiments.

Definition 1.1 A **sample space** Ω of a chance experiment is the set of all possible outcomes that may appear in a realization of this experiment. The elements of Ω are called sample points for this experiment. A subset of Ω is called an **event**.

An event $A = \{\omega\}$, consisting of a single sample point, i.e. a single outcome $\omega \in \Omega$, is called an **elementary event**. We use capital letters A, B, C, and so on to denote events.[1] If an event A consists of more than one outcome, then it is called a **compound event**.

The following simple examples illustrate the above concepts.

Example 1.1 Perhaps the simplest example of a chance experiment is tossing a coin. There are two possible outcomes – Heads (denoted by H) and Tails (denoted by T). In this notation, the sample space of the experiment is

$$\Omega = \{H, T\}.$$

If we toss two coins instead, there are four possible outcomes, represented by the pairs HH, HT, TH, TT. The sample space for this experiment is thus

$$\Omega = \{HH, HT, TH, TT\}.$$

Here, the symbol HH means that both coins land Heads, while HT means that the first coin lands Heads and the second lands Tails. Note in particular that we treat the two events HT and TH as distinguishable, rather than combining them into a single event. The main reason for this is that the events HT and TH are elementary events, while

---

[1] We shall avoid being too pedantic and, for typographical convenience, we shall often use the expression "the event ω" instead of "the event $\{\omega\}$" in the sequel.

the event “one coin lands Heads and the other lands Tails,” which contains both HT and TH, is no longer an elementary event. As we will see later on, when we assign probabilities to the events of a sample space, it is much easier to work with elementary events, since in many cases such events are equally likely, and so it is reasonable the same probability to be assigned to each of them.

Consider now the experiment of tossing three coins. The sample space consists of triplets of the form HHT, HTH, TTH, and so on. Since for each coin toss there are two outcomes, for three coins the number of possible outcomes is $2^3 = 8$. More explicitly, the sample space for this experiment is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}. \tag{1.1}$$

Each of the eight elements of this set forms an elementary event. Note that for events which are not elementary, it is sometimes easier to express them in words, by describing a certain property shared by all elements of the event we consider, rather than by listing all its elements (which may become inconvenient if these elements are too many). For instance, let A be the event “exactly two Heads appear when we toss three coins.” Then,

$$A = \{HHT, HTH, THH\}.$$

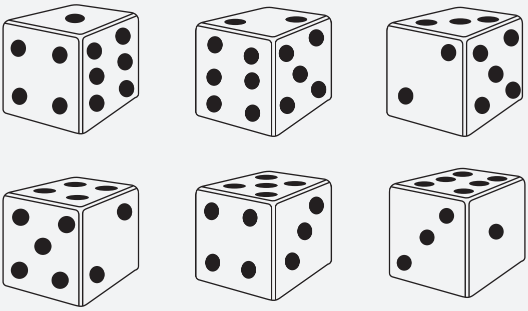
On the other hand, the event

$$B = \{HHH, TTT\}$$

could be described in words as “all three coin outcomes are the same.”

Example 1.2 Another very simple experiment consists of throwing a single die. The die may land on any face with a number i on it, where i takes the values $1, 2, 3, 4, 5, 6$. Therefore, this experiment has sample space

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$



The elementary events are the sets

$$A_1 = \{1\}, \quad A_2 = \{2\}, \quad A_3 = \{3\}, \quad A_4 = \{4\}, \quad A_5 = \{5\}, \quad A_6 = \{6\}.$$

Any other event may again be described either by listing the sample points in it, such as

$$B = \{1, 3, 5, 6\}$$

or, in words, by expressing a certain property of its elements. For instance, the event

$$C = \{2, 4, 6\}$$

may be expressed as "the outcome of the die is an even integer."

For the experiments we considered in the above two examples, the number of sample points was finite in each case. For instance, in Example 1.2, the sample space has six elements, while in the experiment of throwing three coins there are eight sample points as given in (1.1). Such sample spaces which contain a finite number of elements (possible outcomes) are called **finite sample spaces**. It is obvious that any event, i.e. a subset of the sample space, in this case has also finitely many elements.

When dealing with finite sample spaces, the process of enumerating their elements, or the elements of events in such spaces, is often facilitated by the use of **tree diagrams**. Figure 1.1 depicts such a diagram which corresponds to the experiment of tossing three coins, as considered in Example 1.1.

From the "root" of the tree, two segments start, each representing an outcome (H and T, resp.) of the first coin toss. Thus, at the first stage, i.e. after the first throw, there are two nodes. From each of these, in turn, two further segments start corresponding to the two outcomes of the second toss. At the end of the second stage (after the second toss of the coin), there are four nodes. Finally, each of these is associated with two further

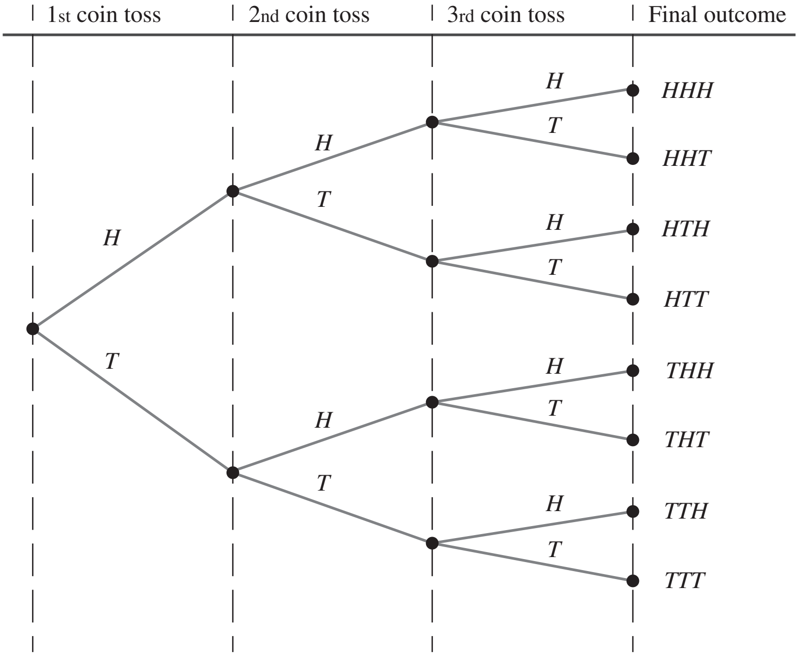


Figure 1.1 Tree diagram for the experiment of tossing three coins.

nodes, which are shown at the next level (end of the three coin tosses). The final column in Figure 1.1 shows the eight possible outcomes for this experiment, i.e. it contains all the elements of the sample space Ω. Each outcome can be traced by connecting the endpoint to the root and writing down the corresponding three-step tree route.

Example 1.3 (An unlimited sequence of coin tosses)
Let us consider the experiment of tossing a coin until "Tails" appear for the first time. In this case, our sample space consists of sequences like $T, HT, HHT, HHHT, \ldots$; that is, $\Omega = \{T, HT, HHT, HHHT, \ldots\}$. The event "Tails appear for the first time at the fifth trial" is then the elementary event

$$A = \{HHHHT\},$$

while the set

$$B = \{T, HT, HHT\}$$

has as its elements all outcomes where Tails appear in the first three tosses. So the event B can be described by saying "the experiment is terminated within the first three coin tosses." Finally, the event "there are at least four tosses until the experiment is terminated" corresponds to the set (event)

$$C = \{HHHT, HHHHT, HHHHHT, \ldots\}.$$

In the previous example, the sample space Ω has infinitely many points. In particular, and since these points can be enumerated, we speak of a **countably infinite sample space**. Examples of sets with countably[2] many points are the set of integers, the set of positive integers, the set of rationals, etc. When a sample space is countably infinite, the events of that sample space may have either finitely many elements (e.g. the event B in Example 1.3) or infinitely many elements (e.g. the event C in Example 1.3).

In contrast, a set whose points cannot be enumerated, is called an *uncountable set*; typical examples of such sets are intervals and unions of intervals on the real line. To illustrate this, we consider the following example.

Example 1.4 In order to monitor the quality of light bulbs that are produced by a manufacturing line, we select a bulb at random, plug it in and record the length of time (in hours) until it fails. In principle, this time length may take any nonnegative real value (however, this presupposes we can take an infinitely accurate measurement of time!). Therefore, the sample space for the experiment whose outcome is the life duration of the bulb is

$$\Omega = \{t : t \geq 0\} = [0, \infty).$$

The subset of Ω

$$A = \{t : 0 \leq t \leq 500\} = [0, 500]$$

---

[2] In mathematics, a countable set is a set which has the same cardinality (number of elements) as the set of natural numbers, $\mathbb{N} = \{1, 2, 3, \ldots\}$. This means that there exists a one-to-one correspondence between the elements of this set and the set $\mathbb{N}$.

describes the event “the life time of the light bulb does not exceed 500 hours,” while the event “the light bulb works for at least 300 hours” corresponds to the set

$$B = \{t : t \geq 300\} = [300, \infty).$$

The sample space Ω in the last example is the half-line of nonnegative real numbers, which is an uncountable set. If Ω is uncountable, then it is usually referred to as a **continuous sample space**. Typically, the study of such sample spaces requires different treatment compared with sample spaces which are either finite or countably infinite. The latter two cases, however, present several similarities and in probability theory the techniques we use are very similar. As a consequence, there is a collective term for sample spaces which have either finitely many or a countably infinite number of elements, called **discrete sample spaces**.

At this point, it is worth noting the difference between an “ideal” continuous sample space and the one we use in practice. With reference to Example 1.4 regarding the lifetime of electric bulbs, such a lifetime does not, in practice, take values such as $\sqrt{12}$ or $(3 + \ln 20)/2$. Since time is measured in hours, it is customary to record a value rounded to the closest integer or, if more precision is required, keep two decimal places, say. In either case, and in contrast to the one used in the example above, the sample space is *countable*. Moreover, if we know that a lifetime of a bulb cannot exceed some (large) value a, the sample space becomes $\Omega = \{0, 0.01, 0.02, \ldots, a\}$ so that it is then in fact *finite*. However, the number of elements in that space is $100a + 1$, so that when a is a large integer, this can be very large. It is often the case that it is much simpler mathematically to assume that Ω is continuous although in practice we can only observe a finite (or infinitely countable) number of outcomes. This convention will be used frequently in the sequel, when we study, for example, weight, age, length, etc.

We conclude this section with an example which demonstrates that for the same experiment, we can define more than one sample space, depending on different aspects we might be interested in studying.

Example 1.5 (Different sample spaces for the same experiment)

Suppose that a store which sells cars has two salespersons. The store has in stock only two cars of a particular make. We are interested in the number of cars which will be sold by each of the two salespersons during next week. Then, a suitable sample space for this experiment is the set of pairs (i,j) for $i,j \in \{0, 1, 2\}$, where i stands for the number of cars sold by the first salesperson and j for the number of cars sold by the second one. However, since there are only two cars available for sale, it must also hold that $i + j \leq 2$, and we thus arrive at the following sample space

$$\Omega_1 = \{(0,0), (0,1), (0,2), (1,0), (1,1), (2,0)\}.$$

Notice that we treat again the pairs (i,j) and (j,i) as being distinguishable; if, however, the store owner is only interested in the `total number of cars` sold during next week, then we could use as a sample space the set

$$\Omega_2 = \{0, 1, 2\}.$$

In this case, an element $\omega \in \Omega_2$ denotes the total number of cars sold. It is worth noting that a specific event of interest is expressed in a completely different manner under these two different sample spaces. Consider, for example, the event

$$A: \text{ the number of cars that will be sold next week is 2.}$$

Viewed as a subset of Ω_1, the event A is a compound event which can be described as

$$A = \{(0,2), (1,1), (2,0)\}.$$

However, when we consider the sample space Ω_2, the event A is an elementary event,

$$A = \{2\}.$$

EXERCISES

Group A

1. Provide suitable sample spaces for each of the following experiments. For each sample space, specify whether it is finite, infinitely countable or uncountable.
   (a) Two throws of a die
   (b) Gender of the children in a family that has three children
   (c) Random selection of a natural number less than 100
   (d) Random selection of a real number from the interval [0, 1]
   (e) Number of telephone calls that someone receives on a mobile phone during a day
   (f) Number of animals living in a certain forest area
   (g) Life duration of an electronic appliance
   (h) Change in the price of a stock during a day
   (i) Percentage change in the price of a stock during a day.
2. John throws a die and subsequently he tosses a coin.
   (i) Suggest a suitable sample space that describes the outcomes of this experiment.
   (ii) Let A be the event that "the outcome of the coin toss is Heads." Which elements of the sample space are included in the event A?
3. We toss a coin until either Heads appear for the first time or we have five tosses which all result in Tails. Give a suitable sample space for this experiment, and then write explicitly (i.e. by listing their elements) each of the following events:

   A: the experiment terminates at the third toss of the coin;

   B: the experiment terminates after the third toss of the coin;

   C: the experiment terminates before the fourth toss of the coin;

D: the experiment terminates with the appearance of H (Heads).

Which, if any, of the events A, B, C, D are elementary events in the sample space that you have considered?

4. A digital scale has an accuracy of two decimal places shown on its screen. Each time a person steps on the scale, we record his/her weight by rounding it to the closest integer (in kilograms). Thus, if the decimal part is 0.50 or greater, we round up to the next integer.

   (i) Give an appropriate sample space for the experiment whose outcome is the rounding error during the above procedure.

   (ii) Write explicitly each of the events

   A: the rounding error is at most 0.10;

   B: the absolute value of the rounding error is at least 0.20.

5. We throw a die twice. Give a suitable sample space for this experiment and then identify the elements each of the following events contains:

   A_1: the outcome of the first throw is 6;

   A_2: the outcome of the second throw is a multiple of 3;

   A_3: the outcome of the first throw is 6 and the outcome of the second throw is a multiple of 3;

   A_4: the sum of the two outcomes is 7;

   A_5: the sum of the two outcomes is at least 9;

   A_6: the two outcomes are identical.

6. A company salesman wants to visit the four cities a, b, c, d wherein his company has stores. If he plans to visit each city once, give a suitable sample space to describe the order in which he visits the cities. Then identify, by an appropriate listing of its elements, each of the following events:

   A_1: the salesman visits first city b;

   A_2: the salesman visits city b first and after that visits city d;

   A_3: the salesman visits city a before he visits city d;

   A_4: the salesman visits the cities a, b, c successively.

7. A box contains 3 red balls and 2 yellow balls. Give a suitable sample space to describe all possible outcomes for the experiment of selecting 4 balls at random, in each of the following schemes:

   (i) For each ball selected, we note its color and return it to the box so that it is available for the next selection (such a scheme is called *selection with replacement*).

   (ii) Every ball selected is subsequently removed from the box (which is called *selection without replacement*).

8. Irène has four books that she wants to put on a library shelf. Three of these books form a 3-volume set of a dictionary, so that they are marked as Volumes *I*, *II*, and *III*, respectively.

   (i) Find an appropriate sample space for all possible ways she can put the books on the shelf.

   (ii) Identify the elements of this sample space that each of the following three events contains:

   B_1: the three volumes of the dictionary are put next to one another;

   B_2: the three volumes of the dictionary are put in the right order (but not necessarily in adjacent places), so that Volume *I* is placed to the left of Volume *II*, which in turn is placed to the left of Volume *III*;

   B_3: the three volumes of the dictionary are placed next to one another and in the right order.

9. Mary has in her wallet three \$1 coins, one \$2 coin and four coins of 25 ¢. She selects four coins at random from her wallet.

   (i) Write down a sample space for the possible selections she can make.

   (ii) Express the following events as subsets of the above sample space:

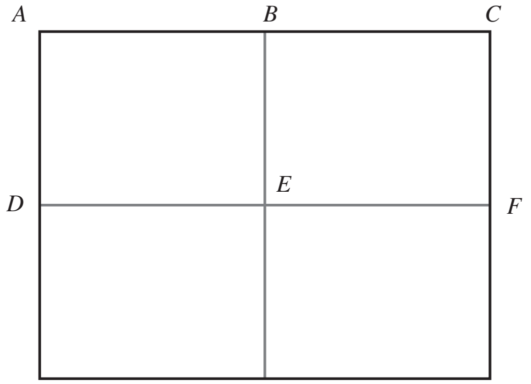
   C_1: exactly three 25 ¢ coins are selected;

   C_2: the total value of the coins selected is \$2.50;

   C_3: the total value of the coins selected is \$3.50.

Group B

10. Bill just visited a friend who lives in Place *A* of the graph below and he wants to return home, which is at Place *I* on the graph. In order to minimize the distance he has to walk, he moves either downwards (e.g. from Place *A* to Place *D*) or to the right (e.g. from Place *E* to Place *F* on the graph). At each time, he makes a choice for his next movement by tossing a coin.



(i) Give a sample space for the different routes Bill can follow to return home.

(ii) Write down explicitly the following events, representing them as subsets of the sample space given in (i):

A_1: he passes through Place E on his way back home;

A_2: he does not pass through Place D;

A_3: on his way back, he has to toss the coin only twice to decide on his next move.

11. A bus, which has a capacity of carrying 50 passengers, passes through a certain bus stop every day at some time point between 10:00 a.m. and 10:30 a.m. In order to study the time the bus arrives at this stop, as well as the number of passengers in the bus at the time of its arrival, we use the following sample space

$$\Omega = \{(k, t) : 0 \leq k \leq 50 \text{ and } 10 \leq t \leq 10.5\},$$

where k above denotes the number of passengers in the bus and t denotes the arrival time at the bus stop (in hours, expressed as a decimal number).

(i) Is the sample space Ω for this experiment countable or uncountable (continuous)?

(ii) Write down explicitly each of the events below:

A_1: the bus arrives at the stop at 10:10 a.m. carrying 25 passengers;

A_2: the bus arrives at the stop at 10:10 a.m. with less than 25 passengers;

A_3: the bus arrives at the stop between 10:10 a.m. and 10:15 a.m.;

A_4: the bus arrives at the stop between 10:10 a.m. and 10:15 a.m., carrying at most 40 passengers.

(iii) Which, if any, of the events in part (ii) is a countable set?

(iv) Suppose now that, in order to describe the experiment of the bus arrival at the bus stop, we use pairs (i, τ), where i is the number of passengers that may get on the bus when it arrives, while τ represents the time after 10:00 a.m. that the bus arrives at the stop. Write down a sample space Ω' for this experiment. Express each of the events in part (ii) as subsets of this new sample space.

12. At a car production line in a factory, each engine produced is tested to examine whether it is ready for use or has some fault. If two consecutive engines that are examined are found faulty, the production process stops and is revised (in such a case, the experiment is terminated). Otherwise, the process continues.

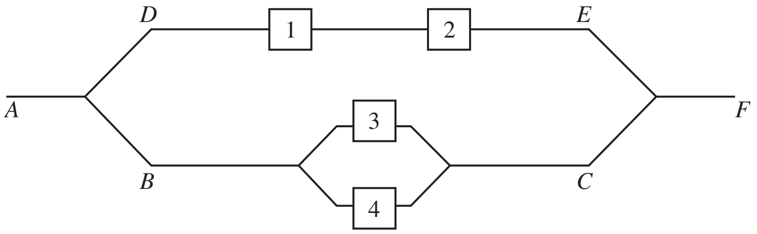
(i) Provide a suitable sample space to describe the inspection process of the engines.

(ii) Find an expression to describe each of the following events

A_i: the production line will be revised after i engines have been examined

for $i = 2, 3, 4, 5, 6$.

13. In a water supply network, depicted below, the water is transferred from point A to point F through water tubes. At the positions marked with the numbers $1, 2, 3$, and 4 on the graph, there are four switches which, if turned off, stop the water supply passing through the tube.



   (i) Find a sample space for the experiment which describes the positions of the four switches (ON or OFF).

   (ii) Identify each of the following events as a suitable subset of the sample space given in part (i):

   A_1: there is water flow from point D to point E;

   A_2: there is water flow from point B to point C;

   A_3: there is water flow from point A to point F.

1.2 OPERATIONS BETWEEN EVENTS

In the preceding section, we made a distinction between discrete (i.e. finite or countably infinite) and continuous sample spaces. When a sample space Ω is discrete, then any subset of Ω is an event. For continuous sample spaces, however, some theoretical difficulties appear if we assume that all subsets of Ω are events. There are cases where certain[3] sets have to be excluded from the "family of events" related to a sample space Ω. The treatment of such technical difficulties is beyond the scope of the present book and in all applications we will consider, we assume that any subset of the sample space is an event.

Suppose that Ω is a sample space for a chance experiment and $A \subseteq \Omega$ is an event. If, in a realization of the experiment we are studying, we observe the outcome $\omega \in \Omega$ which belongs to A, then we say that A **has occurred** or that A has appeared. For example, if we toss a coin three times and we observe the outcome HTH, then (with reference to Example 1.1) we may say that

- the event $A = \{HHT, HTH, THH\}$ has occurred, but
- the event $B = \{HHH, TTT\}$ has not occurred.

For the study of events associated with a certain experiment, and the assignment of probabilities to these events later on, it is essential to consider various relations among the events of a sample space, as well as operations among them. Recall that each event is

---

[3] Such sets are studied in a specific branch of mathematics called *Measure Theory*. Subsets of Ω which cannot be considered as events for the sample space are called "nonmeasurable" sets.

mathematically represented by a set (a subset of Ω); thus, it is no surprise that the relations and operations we consider are borrowed from mathematical set theory.

To begin with, assume that A and B are events on the same sample space Ω. If every element (sample point) of A is also a member of B, then we use the standard notation for subsets and write $A \subseteq B$ (A is a subset of B). In words, this means that whenever A occurs, B occurs too. For instance, in a single throw of a die (see Example 1.2), consider the events:

$$A\colon \text{ the outcome of the die is 4,}$$

$$B\colon \text{ the outcome of the die is an even integer.}$$

Then, expressing A and B as sets, we have $A = \{4\}, B = \{2, 4, 6\}$ and it is clear that $A \subseteq B$. On the other hand, if we know that the outcome is 4 (A has occurred), then necessarily the outcome is even, so that B occurs, too.

If $A \subseteq B$ and $B \subseteq A$, then obviously A occurs iff (if and only if) B occurs in which case we have the following definition.

Definition 1.2 Two events A and B, defined on a sample space Ω, are called equivalent if when A appears, then B appears, and vice versa. In this case, we shall write $A = B$.

The entire sample space Ω itself is an event (we have trivially $\Omega \subseteq \Omega$) and, since Ω contains all possible experiment outcomes, it is called a **certain event**. On the other hand, if we are certain that an event cannot happen, then we call this an **impossible event** for which we use the empty set symbol, $\emptyset$.

Coming back to the experiment of throwing a die, consider again the event $A = \{4\}$ and, instead of B, the event

$$C\colon \text{ the outcome of the die is 5.}$$

Suppose now that Nick, who likes gambling, throws a die and wins a bet if the outcome of the die is *either* 4 *or* 5. Then, the event

$$D\colon \text{ Nick wins the bet}$$

occurs if and only if at least one of the events A and C occur. The event D is the same as the event "at least one of A and C occur" (more precisely, and according to Definition 1.2, these two events are equivalent), and so, using set notation, we can write $D = \{4, 5\}$. We thus see that, expressed as a set, D coincides with the union of the two sets A and C.

Definition 1.3 The **union** of two events A and B, denoted by $A \cup B$, is the event which occurs if and only if at least one of A and B occur.

The union operation can be easily extended when more than two sets are involved. More specifically, if $A_1, A_2, \ldots, A_n$ are events on a sample space, then the event "at least one of the A_i's occur" is called the union of the events $A_1, A_2, \ldots, A_n$ and is expressed in symbols as

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

To illustrate the next concept, viz., the intersection between two events, we return to the example of throwing a die. Suppose we have two gamblers who play against an opponent.

In particular, each of them throws a die: the first player wins if the outcome of the die is greater than or equal to 4, while the other one wins if the outcome is an even integer. How do we describe the event that "both players win their bets?"

Let A be the event that the first gambler wins his bet, B the event that the second gambler wins, and C the event that they both win their bets. Then, clearly, C occurs if and only if both A and B occur. In order to find C explicitly, we resort to set notation again; A and B can be written as follows:

$$A = \{4, 5, 6\}, \quad B = \{2, 4, 6\}.$$

It is now apparent that the event "both A and B occur" contains exactly those elements of the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ which belong to both A and B, that is,

$$C = \{4, 6\}.$$

In set theory, the set which contains exactly those elements which are common to two other sets A and B is called the *intersection* between A and B. We thus arrive at the following definition.

Definition 1.4 The **intersection** between two events A and B is the event which occurs whenever both A and B occur. For the intersection between A and B, we write $A \cap B$ or simply AB.

It should be mentioned that in most probability books, in contrast to set theoretic books, the notation AB is more common than $A \cap B$ and this is the notation which will largely be followed in the present book.

As with the union before, we can extend the last definition to cover more than two sets; in particular, assume again that $A_1, A_2, \dots, A_n$ are events defined on a sample space. Then, the event "all the events A_i occur" is called the intersection of the events $A_1, A_2, \dots, A_n$ and we denote it by

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

or, more commonly,

$$A_1 A_2 \cdots A_n = \prod_{i=1}^{n} A_i,$$

where the product symbol $\prod$ is used as a generic symbol for the intersection among the A_i's.

When we consider an infinite sequence of events $A_1, A_2, \dots$, their union and intersection are denoted by

$$\bigcup_{i=1}^{\infty} A_i \quad \text{and} \quad \bigcap_{i=1}^{\infty} A_i = \prod_{i=1}^{\infty} A_i,$$

respectively. When two events cannot occur simultaneously, then they are called *disjoint* or *mutually exclusive events*. An equivalent definition of this concept is the following.

Definition 1.5 Two events A and B, defined on a sample space Ω, are called **disjoint** or **mutually exclusive events** if their intersection is the empty set (impossible event), i.e. $AB = \emptyset$.

In the case where we have more than two events, $A_1, A_2, \ldots, A_n$, and that

$$A_i A_j = \emptyset \quad \text{for any } i \neq j \; (i, j = 1, 2, \ldots, n),$$

then the events A_i are said to be **pairwise disjoint**. A similar definition applies when the number of the A_i's is infinite.

Two further important concepts useful in probability theory, which arise again from similar concepts in set theory, are given in the next two definitions.

Definition 1.6 Let A be an event on a sample space. The **complement**, or the **complementary event** of A, is the event which occurs if and only if A does not occur.

An equivalent definition is that the complement of A contains exactly those elements of the sample space which are not elements of A.

For the complement of an event A, there are at least three different symbols which are used quite commonly, and they are

$$A' \quad \text{or} \quad \overline{A} \quad \text{or} \quad A^c.$$

In this book, we prefer to use the first of these three symbols.

Definition 1.7 The **difference** of an event B from an event A refers to the event which occurs when A occurs but B does not. The symbol we use for the difference of B from A is $A - B$.

Two useful expressions, which link the above concepts and follow easily from the definitions above, are the following:

$$A' = \Omega - A \quad \text{and} \quad A - B = AB'.$$

Example 1.6 (Emission of digital signals)

A digital signal is a stream of numbers, usually in binary form (0 or 1). Such signals are used, for example, in electrical engineering, telecommunications, biomedicine, seismology and so on. Assume that a source emits a sequence of four binary digits, such as

$$0010, \quad 0001, \quad 1010, \quad 1100, \quad \ldots$$

As can be seen from the following tree diagram, for each 4-digit emission, the sample space consists of 16 elements, given explicitly in the final column (which has been labeled as “result”) of the diagram (Figure 1.2).

Suppose we are interested in the following events:

A_i: the signal contains exactly i “0” digits, for $i = 0, 1, 2, 3, 4$;

B: the signal contains at least two “0” digits;

C: the signal contains exactly three digits which are the same;

D: the signal has at least one “0” digit and one “1” digit;

E: the signal contains exactly three successive digits which are the same.

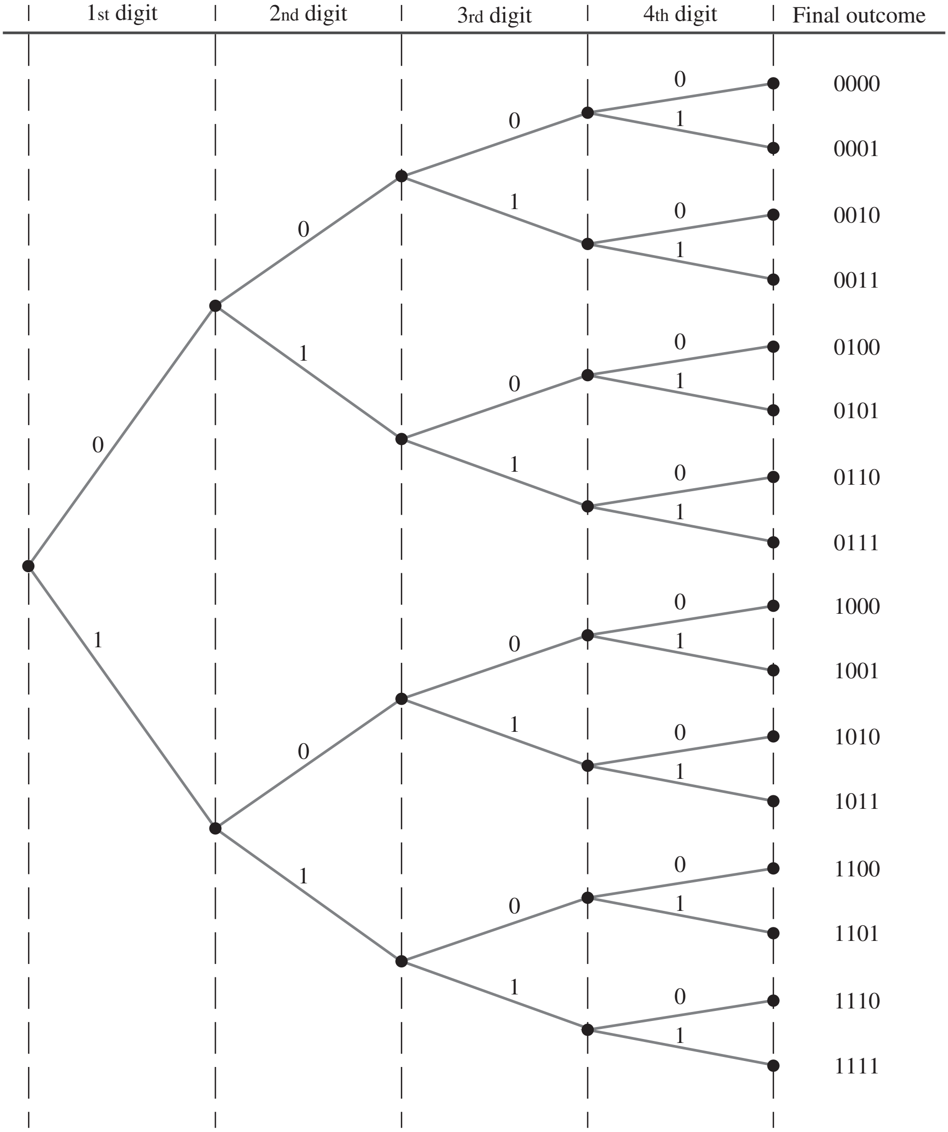


Figure 1.2 Tree diagram for the emission of a signal consisting of four binary digits.

After a careful look through the tree diagram, we conclude that the events A_i $(i = 0, 1, 2, 3, 4)$ can be written explicitly as follows:

$$A_0 = \{1111\}, \quad A_1 = \{0111, 1011, 1101, 1110\},$$

$$A_2 = \{0011, 0101, 0110, 1001, 1010, 1100\},$$

$$A_3 = \{0001, 0010, 0100, 1000\}, \quad A_4 = \{0000\}.$$

It is apparent that the A_i's are pairwise disjoint, i.e.

$$A_iA_j = \emptyset \quad \text{for any } i \neq j \ (i, j = 0, 1, 2, 3, 4).$$

Note that, in addition, we have

$$A_0 \cup A_1 \cup A_2 \cup A_3 \cup A_4 = \Omega$$

(a family of events with these two properties is called a **partition** of the sample space Ω). The events B, C, D can then be expressed in terms of the events A_i as follows:

$$B = A_2 \cup A_3 \cup A_4, \quad C = A_1 \cup A_3, \quad D = A_1 \cup A_2 \cup A_3 = A_0' A_4'.$$

Finally, from the description of the event E, it is obvious that

$$E \subseteq A_1 \cup A_3.$$

More precisely, the event E contains the following elements

$$E = \{0111, 1110, 0001, 1000\}$$

and there seems to be no simple way in which it can be expressed in terms of the events A_i, $i = 0, 1, 2, 3, 4$, by employing the usual operations between events.

Another tool from set theory that is useful often to describe and understand better relations between events is a Venn diagram. In such a diagram, the sample space Ω is represented by a rectangle and, in it, we plot circles or other geometrical shapes to represent certain events in that space (see Figure 1.3).

Figures 1.4–1.11 illustrate graphically the various concepts (relations and operations among events) which have been discussed so far.

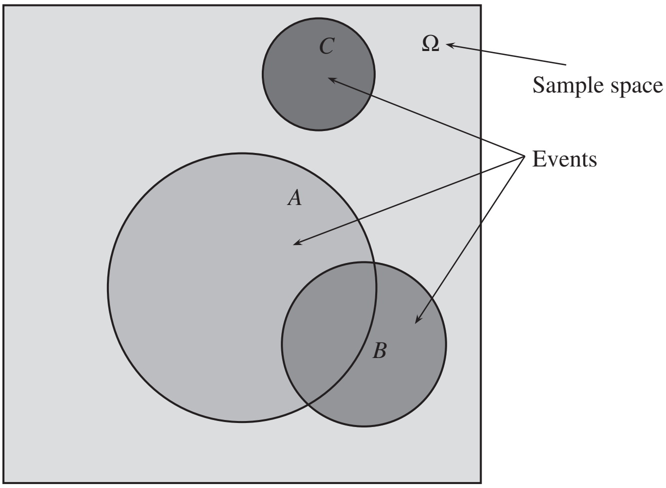


Figure 1.3 A Venn diagram.

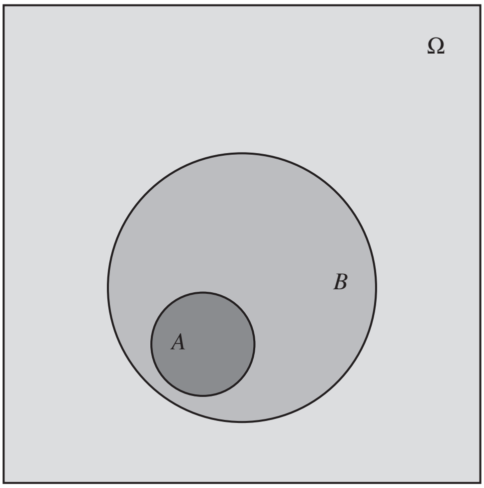


Figure 1.4 $A \subseteq B$.

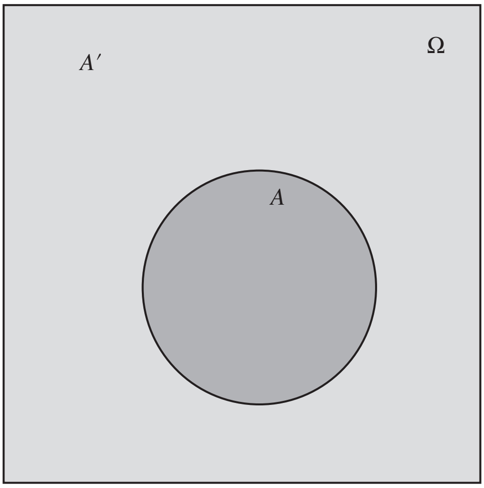


Figure 1.5 Complement of an event.

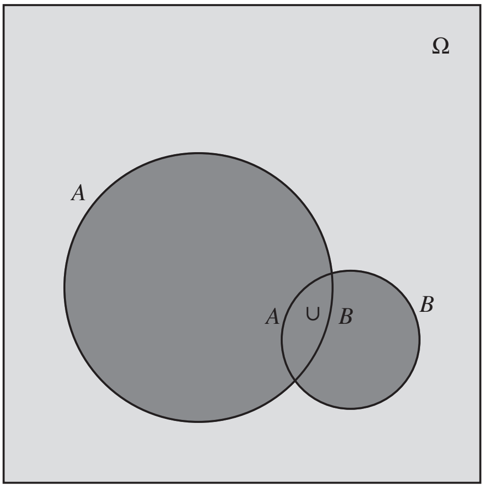


Figure 1.6 Union of events.

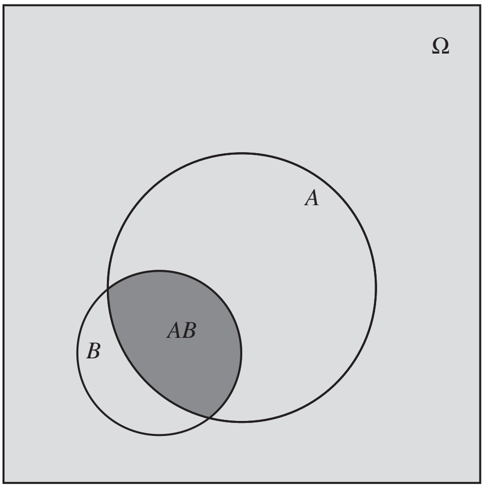


Figure 1.7 Intersection of events.

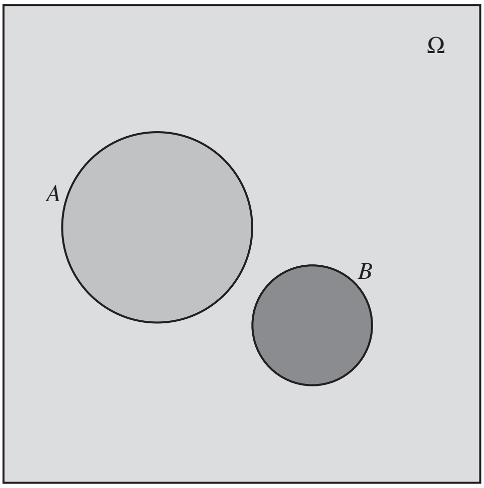


Figure 1.8 Disjoint events.

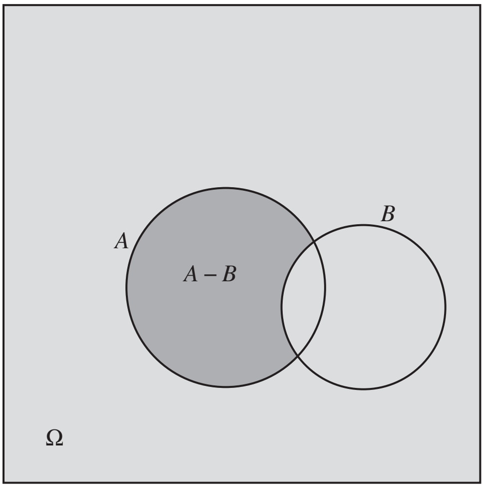


Figure 1.9 Difference between two events.

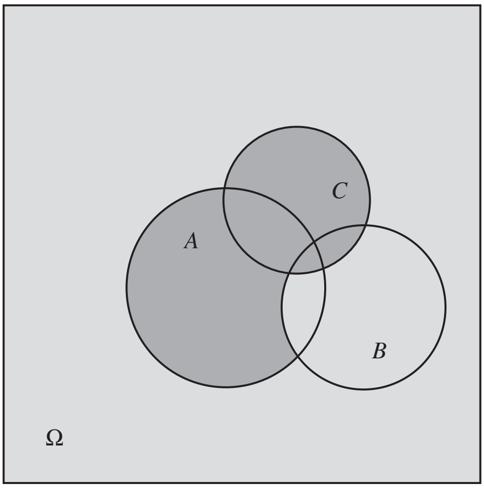


Figure 1.10 $(A - B) \cup C$.

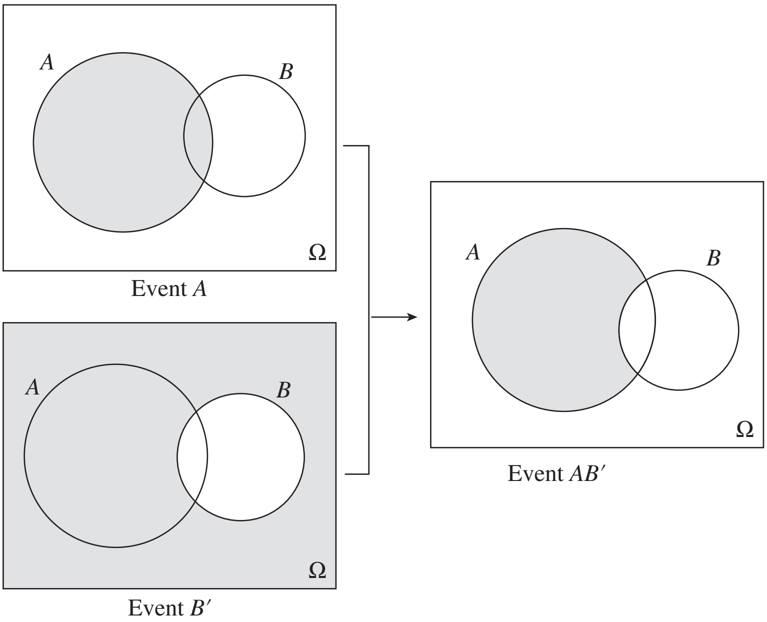


Figure 1.11 Venn diagram for the event AB'.

The use of such diagrams offers, typically quite useful, visualizations of the various relations among events, and enables us to depict graphically rather complicated events. For example, let A, B, C be three events on a sample space and consider the event

C: either A occurs but not B, or else C occurs.

From the definitions given earlier, we see that this event corresponds to the set $(A - B) \cup C$; this event is represented by the shadowed area in Figure 1.10.

Considering the various operations among events that have been introduced, we may state numerous properties that are particularly useful when we wish to simplify complicated expressions. Some of these properties are listed in the following proposition.

Proposition 1.1 *The following properties hold true for the operations among events:*

*SP*1. $A \cup A = A \qquad AA = A$

*SP*2. $A \cup \emptyset = A \qquad A\emptyset = \emptyset$

*SP*3. $A \cup \Omega = \Omega \qquad A\Omega = A$

*SP*4. $A \cup B = B \cup A \qquad AB = BA$

*SP*5. $A \cup (B \cup C) = (A \cup B) \cup C \qquad A(BC) = (AB)C$

*SP*6. $A \cup (BC) = (A \cup B)(A \cup C) \qquad A(B \cup C) = (AB) \cup (AC)$

*SP*7. $A \cup A' = \Omega \qquad AA' = \emptyset$

*SP*8. $(A')' = A$

*SP*9. *If* $A \subseteq B$ *and* $B \subseteq C$, *then* $A \subseteq C$

*SP*10. *If* $A \subseteq B$, *then* $B' \subseteq A'$ *and vice versa*

*SP*11. *If* $A \subseteq B$, *then* $AB = A$ *and* $A \cup B = B$.

Property *SP*5, known as the *associativity property,* allows us to omit the brackets and use a notation like $A \cup B \cup C$ or ABC, because the order in which the unions or intersections are performed is immaterial. Also, *SP*4 in the above proposition (commutative property) shows that when the union or intersection operators are used, we can change the order of the events considered.

We further note that, those properties in Proposition 1.1 which are not immediate, can be verified easily with the use of a Venn diagram; such diagrams can also be used to discover and study a number of additional relationships among events. As an illustration of this, which is of particular interest, we demonstrate the force of an identity concerning the intersection between the complements of two sets. Let A and B be these two sets (events). The shaded area in the two Venn diagrams on the left of Figure 1.12 present their complements, A' and B', respectively, while the graph on the right shows their intersection, $A'B'$.

From the diagram shown in Figure 1.13, it is apparent that the same result could be obtained if we consider the complement of the union, $A \cup B$.

It therefore seems from Figures 1.12 and 1.13 that the following identity holds:

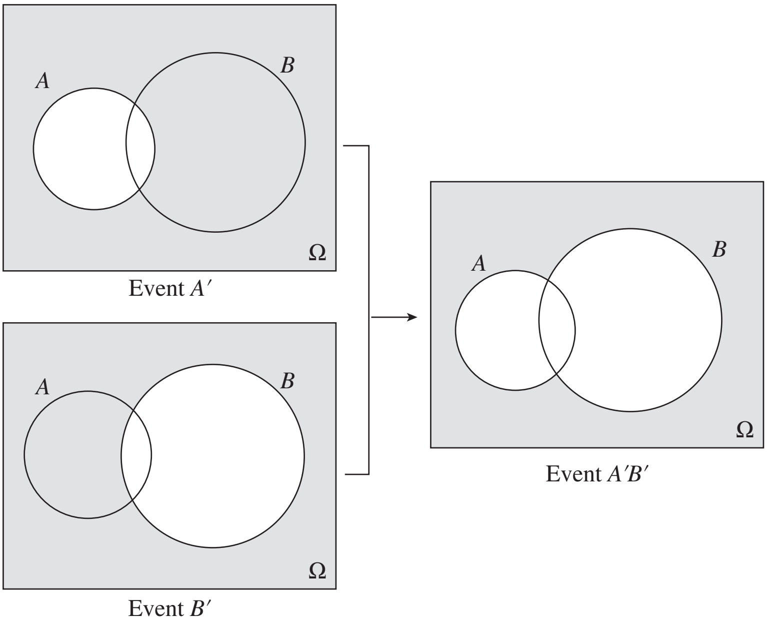
$$(A \cup B)' = A'B'.$$

However, we stress that, no matter how illustrative they might be, Venn diagrams alone do not offer a rigorous mathematical proof for an identity like the one above. In order to prove such an identity formally, we must show that the sets on the two sides of an equation contain exactly the same elements. As an illustration of how this can be done, we present a detailed mathematical proof of the following proposition.

Proposition 1.2 ***(De Morgan formulas)*** *Suppose* A *and* B *are events on the same sample space* Ω. *Then, the following identities hold:*

$$(A \cup B)' = A'B', \quad (AB)' = A' \cup B'.$$

These identities are known as De Morgan identities or De Morgan formulas.

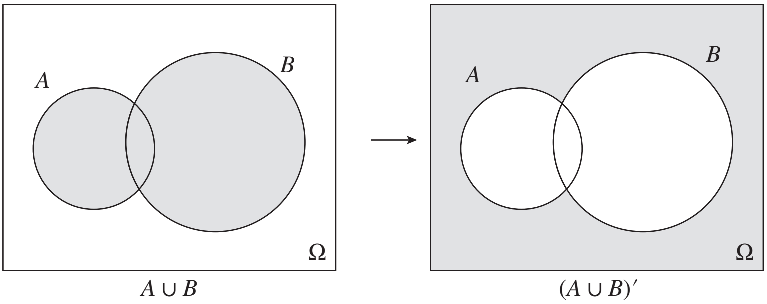


Event A'

Event B'

Event $A'B'$

Figure 1.12 Venn diagram for the event $A'B'$.



$A \cup B$

$(A \cup B)'$

Figure 1.13 Venn diagram for the event $(A \cup B)'$.

Proof: To prove the first identity, it is sufficient to show that both $(A \cup B)' \subseteq A'B'$ and $A'B' \subseteq (A \cup B)'$ hold. To begin with, let ω be a sample point which belongs to the set $(A \cup B)'$. This means that $\omega \notin A \cup B$, and since the set $A \cup B$ contains all elements which are either in A or in B, we conclude that ω does not belong to any of them. Thus, ω belongs to both A' and B', and therefore it is also a member of their intersection, $A'B'$. The above arguments show that $(A \cup B)' \subseteq A'B'$.

Next, to establish the reverse relationship, suppose now that $\omega \in A'B'$. Since the set $A'B'$ contains exactly those elements of the sample space which belong to both A' and B', we see that ω is not a member of either A or B. Thus, it does not belong to their union, $A \cup B$, which shows immediately that $\omega \in (A \cup B)'$. We therefore see that $A'B' \subseteq (A \cup B)'$, and this completes the proof for the first assertion of the proposition.

The proof of the second identity is carried out in a similar fashion. □

De Morgan identities can be generalized for the case when more than two sets are involved. More specifically, if $A_1, A_2, \ldots, A_n$ are events on a sample space, then we have the following identities:

$$(A_1 \cup A_2 \cup \cdots \cup A_n)' = A_1' A_2' \cdots A_n', \quad (A_1 A_2 \cdots A_n)' = A_1' \cup A_2' \cup \cdots \cup A_n'.$$

A similar result holds true if we have an infinite number of events.

We close this section with an example to demonstrate how we can use the various properties we have discussed so far in order to simplify complicated expressions involving sets and their operations.

Example 1.7 Suppose A, B, and C are three events on the same sample space Ω. Using the properties of the operations among events, simplify the expression

$$(A'B')' \cup (AB'C')'.$$

SOLUTION Using the second De Morgan identity from Proposition 1.2 (applied to the sets A' and B') and Property $SP8$, we obtain

$$(A'B')' = (A')' \cup (B')' = A \cup B.$$

In a similar fashion, we get

$$(AB'C')' = A' \cup (B')' \cup (C')' = A' \cup (B \cup C).$$

Inserting these two relations into the expression given in the statement of the example, we obtain

$$\begin{aligned}(A'B')' \cup (AB'C')' &= (A \cup B) \cup (A' \cup (B \cup C)) \\ &= A \cup B \cup A' \cup B \cup C \\ &= (A \cup A') \cup (B \cup B) \cup C \quad \text{(Properties SP4, SP5)} \\ &= \Omega \cup B \cup C \quad \text{(Properties SP1, SP7)} \\ &= \Omega \cup (B \cup C) \\ &= \Omega. \quad \text{(Property SP3)}\end{aligned}$$

EXERCISES

Group A

1. Let A, B, and C be three events in a sample space Ω. Express each of the following events by the use of the operators (unions, intersections, complements) among sets:
   (i) all three events occur;
   (ii) at least one of the three events occur;
   (iii) A occurs, but not A and B;

(iv) A and C occur, but not B;

(v) at least one of the three events occur, but C does not occur;

(vi) exactly two among the three events A, B, and C occur.

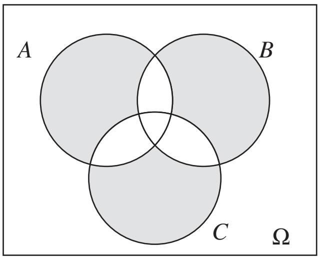
2. Consider the experiment of throwing a die twice, and define the following events:

   A: the sum of the two outcomes is 6;
   B: the two outcomes are equal;
   C: the first outcome is an even integer;
   D: the first outcome is an odd integer.

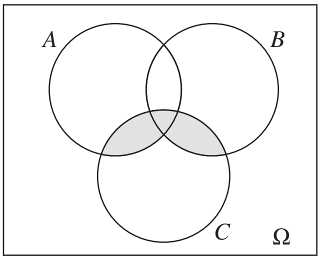
   Write down the 36 sample elements for this experiment, and then identify which of these elements each of the sets A, B, C, D contains. Then describe the following events, both by expressing in words what each of them represents and by listing their elements:

$$AB, \quad AC', \quad AB'C', \quad AC \cup AD, \quad A \cup (CD), \quad BC' \cup BD, \quad (B \cup C)(B \cup D).$$

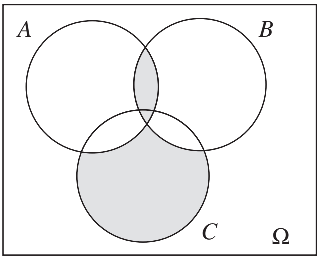
3. For each of the following graphs (a)–(d), express the event in the shaded area in terms of the events A, B, and C and state, in words, what this event represents.



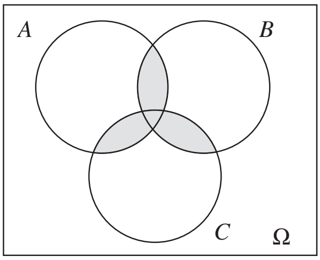
(a)



(b)



(c)



(d)

4. What conclusions can we draw about the events A and B if the following relations hold?

   (i) $A \cup B = A$; (ii) $A - B = A$;
   (iii) $AB = A$; (iv) $A - B = B - A$.

5. Suppose for the events A, B, C, D, we have $A \subseteq B$ and $C \subseteq D$. Then, arguing as in the proof of Proposition 1.2, show that

$$A \cup C \subseteq B \cup D \quad \text{and} \quad AC \subseteq BD.$$

6. If, for the events A and B, we know that $A \subseteq B$ and $A \subseteq B'$, show that $A = \emptyset$. (*Hint*: Use the result of Exercise 5.)

7. Suppose the events A and B of a sample space Ω are such that $A \subseteq B$ and $A' \subseteq B$. Then prove that $B = \Omega$. (*Hint*: You may use the result of Exercise 5.)

8. Let A, B, and C be three events on a sample space Ω. Examine, possibly with the aid of Venn diagrams, which of the following results are always true:

   (i) $(A - AB)B = AB$;

   (ii) $(A \cup B)'C = A'B'C$;

   (iii) $(AB)'C' = (ABC)'$;

   (iv) $AB \subseteq A \subseteq A \cup B \cup C$;

   (v) $A'B \cup C'B = (A \cup C)'B$;

   (vi) $(A - B) - C = A - (B - C)$.

Group B

9. A box contains 15 balls numbered $1, 2, \ldots, 15$. The balls numbered 1–5 are white and those numbered 6–15 are black. We select a ball at random, and record its color and the number on it.

   (i) Write down a suitable sample space for this experiment.

   (ii) We define the events

   A_i: the ball selected has a number less than or equal to i, for $1 \le i \le 15$;

   B_i: the ball selected has a number greater than or equal to i on it, for $1 \le i \le 15$;

   C: the selected ball is white;

   D: the selected ball is black.

   Examine which of the following assertions are correct and which are false:

   (a) $A_5 = C$;

   (b) $A_4 \subseteq C$;

   (c) $A_iB_i = \emptyset$ for any $i = 1, 2, \ldots, 15$;

   (d) $A_{i-1}B_i = \emptyset$ for any $i = 2, 3, \ldots, 15$;

   (e) $A_iB_{i+1} = \emptyset$ for any $i = 1, 2, \ldots, 14$;

   (f) the events C and D are mutually exclusive;

   (g) $A_{10}B_5 \subseteq C$;

(h) $A_7D = \emptyset$;

(i) $A'_5 = D$;

(j) $A_i \cup B_{i+1} = A_i$ for any $i = 1, 2, \ldots, 14$;

(k) $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_{15}$;

(l) $A_i \cup B_i = \Omega$ for any $i = 1, 2, \ldots, 15$;

(m) $B_1 \subseteq B_2 \subseteq \cdots \subseteq B_{15}$;

(n) $A_i \cup B_{i+1} = \emptyset$ for any $i = 1, 2, \ldots, 15$;

(o) $A'_i = B_{i+1}$ for $i = 1, 2, \ldots, 14$;

(p) $(A_{10} - C)B_6 = \emptyset$;

(q) $(A_{12} - D) \subseteq B_5$;

(r) $D - B_{11} = A_{10} - A_5$.

10. Express each of the following events in terms of the events A_i, B_i, C, D defined in Exercise 9.

(i) The number on the ball selected is greater than 4 and less than 10.

(ii) The ball selected is either white or has a number greater than 10 on it.

(iii) The ball selected is black or has an even number on it.

(iv) The ball selected is black and has an odd number on it.

(v) The selected ball is either white with a number less than 3 on it, or else it is black with a number less than 10.

11. In order to describe a chance experiment, we have used the following (continuous) sample space

$$\Omega = \{(x, y) : -5 \leq x \leq 5 \text{ and } -3 \leq y \leq 7\}.$$

On this space, we define the following events:

$$\begin{array}{ll} A = \{(x, y) \in \Omega : x = y\}, & B = \{(x, y) \in \Omega : x^2 = y^2\}, \\ C = \{(x, y) \in \Omega : x + y = 0\}, & D = \{(x, y) \in \Omega : x \leq y\}, \\ E = \{(x, y) \in \Omega : x \geq y\}. & \end{array}$$

Which of the following statements are correct and which are false?

(a) $B = AC$;

(b) $AC \subseteq B$;

(c) D and E are mutually exclusive events;

(d) $A = \{(x, x) : |x| \leq 5\}$;

(e) $A = DE$;

(f) $D' = E - A$;

(g) $A = \{(x, x) : -3 \leq x \leq 3\}$.

12. Let $A_i, i = 1, 2, \ldots, n$, be events on the same sample space Ω. Express in words what conclusions can be drawn about these events in each of the following cases:

    (i) $A_1 \cup A_2 \cup \cdots \cup A_n = A_1$;

    (ii) $A_1 A_2 \cdots A_n = A_1$;

    (iii) $A_1 \cup A_2 \cup \cdots \cup A_n = A_1 \cup A_n$;

    (iv) $A_1 A_2 \cdots A_n = A_1 A_2$.

13. Simplify each of the expressions below by the use of properties among event operators:

    (i) $A'B'AB$;

    (ii) $(A' \cup B') \cup AB$;

    (iii) $(A \cup B)(A \cup B')(A' \cup B)$;

    (iv) $(A'B')(A \cup B)$;

    (v) $[(A \cup B)(A \cup B')] \cup [(A' \cup B)(A' \cup B')]$;

    (vi) $(A \cup B \cup C)(A \cup B \cup C')(A \cup B')$;

    (vii) $A \cup (B - AB) \cup (C - AC)$;

    (viii) $(A \cup B \cup C)(A \cup BC')$;

    (ix) $[ABC(A \cup B \cup C')] \cup (A \cup B')$.

14. From an ordinary card deck with 52 cards, we select n cards successively. We consider the events

    A_1: the first card drawn is an ace;

    A_n: in the first $n - 1$ selections ($n \geq 2$), neither an ace nor a face card is drawn and in the nth selection an ace is drawn,

    where $n = 2, 3, \ldots$ Write down an expression, in terms of $A_1, A_2, \ldots$, for the event that an ace is drawn before a face card.

15. Let A, B, and C be three events in a sample space Ω. In each case below, find an event X such that the union operator on the right hand side is applied between events which are disjoint (for example, in (i) we seek X such that AB and X are disjoint sets):

    (i) $B = (AB) \cup X$;

    (ii) $A \cup B = (A'B) \cup X$;

    (iii) $A \cup B \cup C = A \cup (A'B) \cup X$.

16. For any events A, B, and C in a sample space, verify the truth of the following relations:

    (i) $(A - B) - C = (A - C) - (B - C)$;

    (ii) $(B - A) \cup (C - A) = (B \cup C) - A$;

    (iii) $A - (B - C) = (A - B) \cup (AC)$;

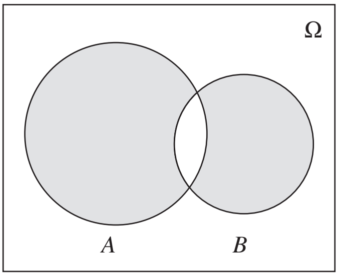
(iv) $(A - B) \cup B = A \cup B$;

(v) $A(B - C) = AB - AC$;

(vi) $A \cup B \cup C = (A - B) \cup (B - C) \cup (C - A) \cup (ABC)$.

17. For any events A and B in the same sample space Ω, we define the **symmetric difference** of A and B to be the event (see next figure)

$$A \Delta B = (A - B) \cup (B - A).$$



(i) Express in words what this event represents;

(ii) Show that the following hold:

$$A \Delta B = B \Delta A, \quad A \Delta B = (A \cup B) - AB, \quad (A \Delta B) \Delta B = A.$$

1.3 PROBABILITY AS RELATIVE FREQUENCY

As already mentioned, the main feature of a chance (random) experiment is that we do not know what its outcome will be in each realization of it. We therefore do not know whether a specific event, A, associated with this experiment will occur in a particular realization of the experiment. Yet, however, we very often want to know how likely it is for A to occur. For instance, what are the chances that it will rain tomorrow, or that I will pass a certain University exam, or that a political party will win the forthcoming elections? Today, hundreds of decisions are made on a daily basis, using the “degree of belief” we have that something will happen, or not, at some future point; just imagine plainly the persons who purchase stocks, or other financial products, hoping that their prices will go up. Probability theory emerged from the need to put this “degree of belief” in a proper mathematical framework, so that the concept of “how likely” an event is can be quantified on one hand, but on the other hand to be able to make logical deductions and draw conclusions about complicated events. More generally, probability theory enables us not only to make decisions but to deal with uncertainty in everyday life, and for this reason it has been referred to as “the science of uncertainty.” From this viewpoint, it may seem paradoxical that the entire history of probability theory and its development, until the nineteenth century at least, is entangled and motivated by games of chance and gambling. Even so, it is curious that, although the origins of games of chance seem to have been lost in history,[4] the first systematic use of mathematics to deal with the uncertainty in games

---

[4] An archaeological excavation in south-eastern Iran unearthed a backgammon, which dates back to 3000BC.

occurs in fifteenth century, while the first worthwhile achievements were established in mid-seventeenth century. And, since then, it took nearly three centuries until a sound and coherent theory was developed so that probability theory became recognized as an independent branch of mathematics.

The first definition of probability to be widely used is attributed to the French mathematician Pierre Simon Laplace (1749–1827) and is usually referred to as the classical definition of probability. In his words, it is as follows:

> "The probability of an event is the ratio of the number of cases favorable to it, divided by the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, equally possible."

Thus, when we throw a die and consider the event

A: the outcome is an even integer,

it seems plausible to assume that all six faces of the die are "equally possible." The number of possible outcomes (the number of elements in the sample space, to conform with the terminology used in the previous sections) is 6, while the "favorable" outcomes are all those in which an even integer results in the throw. These are the elements of the set $\{2, 4, 6\}$ and we therefore obtain from the definition that the probability of the event A is $3/6$, that is, a half.

Although the definition seems to give plausible answers which agree with our intuition, it has at least two serious drawbacks:

(i) What happens if all the "cases" that may occur (that is, the elementary events in the sample space) can not be considered as being "equally possible?"

(ii) The definition assumes that the number of possible outcomes is finite. What happens if this is not the case for the experiment that we consider?

In view of the above problems, mathematicians searched for alternative ways to define probability. The first serious attempt to do this was made by R. von Mises in 1917. This attempt relies on the notion of a relative frequency which is discussed in the rest of this section. The approach taken by von Mises was rather empirical in nature, and this has been considered both as an advantage and a disadvantage; on the one hand, it is nice to link our "degree of belief" with our experience from the past. However, relying on experimental data to define probability has several drawbacks, as we shall see in the next section. Thus, and with the increasing recognition of the importance of probability theory in early twentieth century, scientists felt that there was a need to "axiomatize" probability theory, in the same way as was done with geometry for example. That is, to find a set of axioms which would seem universally acceptable and, from these, derive a chain of theoretical results and also the answer to practical problems encompassing uncertainty. In the next section, we present an approach which followed this path, due to the Russian probabilist A. N. Kolmogorov.

Let us agree, to begin with, that certain events are intuitively more likely than others; for instance,

- when two basketball teams, which are more or less at the same level, play against each other, we typically think that the home team has better chance to win the game;
- our experience suggests that it is "a lot more likely" to rain on a winter day in England than on a summer day in Las Vegas.

Both these assertions seem to be universally accepted, and this can be attributed to our common experience from what has happened in the past. This is more obvious in the second case above, wherein if we collect data say from the last 50 years, we may see for instance that the number of winter days with rain in England is 20 times as much as the number of rainy days in summer in Las Vegas. The same argument may work for the first case since if we select at random a large number of basketball games, it is very likely that home wins occur more often than away wins. For example, looking at the 2010–2011 NBA regular season, we see that among the 1230 matches which were played, there were 743 home wins and 487 away wins.

When the classical definition of probability cannot be used, it seems reasonable to use past experience to assign a degree of credibility to an event. To begin with, we may say that an event A is more likely than another event B, both associated with the same experiment, if in a large number of realizations of this experiment, A occurred more often than B. At a next level, we may attempt to quantify this difference, saying for instance that if A occurs twice as often as B does, then A is twice more likely to occur than B. However, we would wish to use a single number to tell us how likely an event is, rather than considering this always in relation to some other event. Following the above, it seems plausible that, if an experiment is conducted n times and A occurs n_A times out of them, then the ratio n_A/n can be used as a measure of our degree of belief for its appearance. Of course, for this to be valid, and consistent, all n repetitions have to be performed under the same conditions. We thus arrive at the following definition.

Definition 1.8 (Relative frequency) If in n repetitions of an experiment, under identical conditions, the event A occurs n_A times, the ratio

$$f_A = \frac{n_A}{n}$$

is called the *relative frequency* of A (in these n experiments).

From this definition, we have immediately the following properties for a relative frequency.

Proposition 1.3 ***(Properties of relative frequencies)*** *Let Ω be a sample space for a chance experiment. Then:*

(i) $f_A \geq 0$ *for any event A defined on Ω;*

(ii) $f_\Omega = 1$;

(iii) *for any disjoint events A and B, we have*

$$f_{A \cup B} = f_A + f_B.$$

Clearly, the third property can be generalized to more than two events. Thus, if $A_1, A_2, \ldots, A_k$ are pairwise disjoint events, we have

$$f_{A_1 \cup A_2 \cup \cdots \cup A_k} = f_{A_1} + f_{A_2} + \cdots + f_{A_k} = \sum_{i=1}^{k} f_{A_i}.$$

Furthermore, if $A = \{\omega_1, \omega_2, \ldots, \omega_k\} \subseteq \Omega$ is a (finite) event on Ω and we denote by

$$f_i = f_{\omega_i}, \quad i = 1, 2, \ldots, k,$$

the relative frequencies of the elementary events $\{\omega_i\}$, $i = 1, 2, \ldots, k$, then we may write

$$f_A = f_1 + f_2 + \cdots + f_k = \sum_{i=1}^{k} f_i.$$

Example 1.8 (Repetitions of a chance experiment)
Table 1.1 gives the results of 100 realizations for the random experiment of throwing a die twice. In the last three columns of this table, we have recorded the relative frequency (based on the first i trials, $i = 1, 2, \ldots, 100$) for each of the events

A: at least one of the two outcomes is 6;

B: the outcome of the first die is 4;

C: the sum of the two outcomes is 7.

We see from Table 1.1 that the relative frequency for each of the three events A, B, C varies along with the number of throws and, in fact, changes at each step. However, it seems that as the number of throws gets larger, the fluctuations of the values for the three relative frequencies tend to become smaller. Further, throughout Table 1.1, it is apparent that event A occurs more often than the other two, which have about the same rate of occurrence.

Based on the total of 100 throws, we have the following relative frequencies for A, B, C:

$$f_A = 0.29, \quad f_B = 0.16, \quad f_C = 0.15.$$

Suppose now we define the events

$$B_i : \quad \text{the outcome of the first throw is } i$$

for $i = 5, 6$. Taking into account the result for the event B, it seems reasonable to state that

$$f_{B_5} = 0.16, \quad f_{B_6} = 0.16.$$

Further, for the event

$$E : \quad \text{the outcome of the first throw is greater than 4,}$$

since $E = B_5 \cup B_6$, we expect to have (by Proposition 1.3(iii)) that

$$f_E = f_{B_5} + f_{B_6} = 0.32.$$

Throw no.	Outcome		Frequency			Relative Frequency		
	First throw	Second throw	n_A	n_B	n_C	f_A	f_B	f_C
1	5	4	0	0	0	0.0000	0.0000	0.0000
2	6	3	1	0	0	0.5000	0.0000	0.0000
3	4	4	1	1	0	0.3333	0.3333	0.0000
4	5	2	1	1	1	0.2500	0.2500	0.2500
5	3	3	1	1	1	0.2000	0.2000	0.2000
6	1	3	1	1	1	0.1667	0.1667	0.1667
7	5	6	2	1	1	0.2857	0.1429	0.1429
8	2	1	2	1	1	0.2500	0.1250	0.1250
9	2	4	2	1	1	0.2222	0.1111	0.1111
10	3	2	2	1	1	0.2000	0.1000	0.1000
11	2	3	2	1	1	0.1818	0.0909	0.0909
12	4	4	2	2	1	0.1667	0.1667	0.0833
13	5	4	2	2	1	0.1538	0.1538	0.0769
14	6	2	3	2	1	0.2143	0.1429	0.0714
15	2	5	3	2	2	0.2000	0.1333	0.1333
16	3	3	3	2	2	0.1875	0.1250	0.1250
17	6	3	4	2	2	0.2353	0.1176	0.1176
18	1	1	4	2	2	0.2222	0.1111	0.1111
19	4	3	4	3	3	0.2105	0.1579	0.1579
20	2	3	4	3	3	0.2000	0.1500	0.1500
21	1	4	4	3	3	0.1905	0.1429	0.1429
22	3	2	4	3	3	0.1818	0.1364	0.1364
23	6	5	5	3	3	0.2174	0.1304	0.1304
24	3	4	5	3	4	0.2083	0.1250	0.1667
25	4	5	5	4	4	0.2000	0.1600	0.1600
26	2	2	5	4	4	0.1923	0.1538	0.1538
27	4	5	5	5	4	0.1852	0.1852	0.1481
28	6	6	6	5	4	0.2143	0.1786	0.1429
29	6	3	7	5	4	0.2414	0.1724	0.1379
30	1	1	7	5	4	0.2333	0.1667	0.1333
31	5	3	7	5	4	0.2258	0.1613	0.1290
32	5	5	7	5	4	0.2188	0.1563	0.1250
33	4	3	7	6	4	0.2121	0.1818	0.1212
34	1	1	7	6	4	0.2059	0.1765	0.1176
35	4	3	7	7	5	0.2000	0.2000	0.1429
36	6	1	8	7	6	0.2222	0.1944	0.1667
37	3	5	8	7	6	0.2162	0.1892	0.1622
38	5	2	8	7	7	0.2105	0.1842	0.1842
39	6	3	9	7	7	0.2308	0.1795	0.1795
40	1	5	9	7	7	0.2250	0.1750	0.1750

Table 1.1 Dice outcomes of two throws (100 repetitions).

Throw no.	Outcome		Frequency			Relative Frequency		
	First throw	Second throw	n_A	n_B	n_C	f_A	f_B	f_C
41	4	6	10	8	7	0.2439	0.1951	0.1707
42	5	4	10	8	7	0.2381	0.1905	0.1667
43	1	3	10	8	7	0.2326	0.1860	0.1628
44	1	3	10	8	7	0.2273	0.1818	0.1591
45	6	4	11	8	7	0.2444	0.1778	0.1556
46	2	2	11	8	7	0.2391	0.1739	0.1522
47	6	1	12	8	8	0.2553	0.1702	0.1702
48	1	3	12	8	8	0.2500	0.1667	0.1667
49	5	3	12	8	8	0.2449	0.1633	0.1633
50	4	2	12	9	8	0.2400	0.1800	0.1600
51	6	5	13	9	8	0.2549	0.1765	0.1569
52	2	6	14	9	8	0.2692	0.1731	0.1538
53	3	3	14	9	8	0.2642	0.1698	0.1509
54	1	4	14	9	8	0.2593	0.1667	0.1481
55	5	1	14	9	8	0.2545	0.1636	0.1455
56	3	6	15	9	8	0.2679	0.1607	0.1429
57	5	1	15	9	8	0.2632	0.1579	0.1404
58	3	4	15	9	9	0.2586	0.1552	0.1552
59	3	6	16	9	9	0.2712	0.1525	0.1525
60	6	1	17	9	10	0.2833	0.1500	0.1667
61	2	3	17	9	10	0.2787	0.1475	0.1639
62	1	4	17	9	10	0.2742	0.1452	0.1613
63	4	5	17	10	10	0.2698	0.1587	0.1587
64	6	4	18	10	10	0.2813	0.1563	0.1563
65	3	4	18	10	11	0.2769	0.1538	0.1692
66	2	6	18	10	11	0.2727	0.1515	0.1667
67	4	2	18	11	11	0.2687	0.1642	0.1642
68	4	1	18	12	11	0.2647	0.1765	0.1618
69	6	6	19	12	11	0.2754	0.1739	0.1594
70	6	3	20	12	11	0.2857	0.1714	0.1571
71	3	2	20	12	11	0.2817	0.1690	0.1549
72	2	5	20	12	12	0.2778	0.1667	0.1667
73	5	5	20	12	12	0.2740	0.1644	0.1644
74	5	1	20	12	12	0.2703	0.1622	0.1622
75	6	1	21	12	13	0.2800	0.1600	0.1733
76	4	4	21	13	13	0.2763	0.1711	0.1711
77	1	3	21	13	13	0.2727	0.1688	0.1688
78	5	6	22	13	13	02821.	0.1667	0.1667
79	6	5	23	13	13	0.2911	0.1646	0.1646
80	3	4	23	13	14	0.2875	0.1625	0.1750

Table 1.1 *(continued)*

Throw no.	Outcome		Frequency			Relative Frequency		
	First throw	Second throw	n_A	n_B	n_C	f_A	f_B	f_C
81	5	1	23	13	14	0.2840	0.1605	0.1728
82	6	5	24	13	14	0.2927	0.1585	0.1707
83	5	4	24	13	14	0.2892	0.1566	0.1687
84	4	4	24	14	14	0.2857	0.1667	0.1667
85	2	4	24	14	14	0.2824	0.1647	0.1647
86	4	1	24	15	14	0.2791	0.1744	0.1628
87	4	3	24	16	15	0.2759	0.1839	0.1724
88	1	2	24	16	15	0.2727	0.1818	0.1705
89	5	4	24	16	15	0.2697	0.1798	0.1685
90	3	2	24	16	15	0.2667	0.1778	0.1667
91	3	6	25	16	15	0.2747	0.1758	0.1648
92	2	4	25	16	15	0.2717	0.1739	0.1630
93	3	3	25	16	15	0.2688	0.1720	0.1613
94	4	2	25	16	15	0.2660	0.1702	0.1596
95	1	5	25	16	15	0.2632	0.1684	0.1579
96	1	1	25	16	15	0.2604	0.1667	0.1563
97	6	4	26	16	15	0.2680	0.1649	0.1546
98	5	6	27	16	15	0.2755	0.1633	0.1531
99	6	5	28	16	15	0.2828	0.1616	0.1515
100	3	6	29	16	15	0.2900	0.1600	0.1500

Table 1.1 (*continued*)

Looking at the relative frequencies of the events A, B, and C in Table 1.1, we observe that, although these tend to fluctuate initially a lot, as the number of trials performed grows, they seem to be concentrated around a fixed value. If we assume for now that the relative frequency converges to a certain limit, then this limit is called the **limiting relative frequency** of the event. It should be mentioned here that the fact that relative frequencies do converge to a limit is a consequence of one of the key theorems in probability theory, the *"law of large numbers"* (a visual illustration of this is given in Figure 1.14).

In the beginning of the twentieth century, the limit of the relative frequency was used by mathematicians for the definition of probability. In particular, Richard von Mises (1883–1953) gave the following definition of probability, known as the **statistical, or frequentist, definition of probability**.

Definition 1.9 (Frequentist definition of probability) Let Ω be a sample space for a chance experiment and A be an event on that space. If n_A denotes the number of times that A has appeared in n repetitions of the experiment ($0 \leq n_A \leq n$), then we define as the **probability of the event** A the limit

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n} = \lim_{n\to\infty} f_A.$$

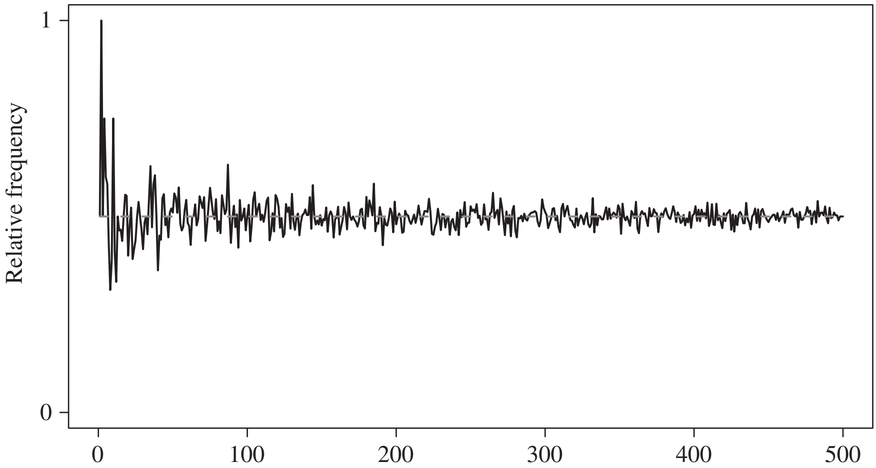


Figure 1.14 Long-run behavior of relative frequencies.

It is worth noting that, although not explicit in the notation, the quantities n_A in the above definition, depend on the number n of repetitions which have been carried out for the experiment that we consider.

We may assume for now that Definition 1.9 can be used to associate probabilities to events on a sample space. Then, by an appeal to the properties listed in Proposition 1.3, and in particular by taking the limit $n \to \infty$ there (that is, assuming that the number of experiments grows indefinitely), we obtain the following properties for probability:

FP1. $P(A) \geq 0$ for any event A on the sample space Ω;

FP2. $P(\Omega) = 1$;

FP3. For any pair of events A, B which are disjoint, we have

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

As we will see in the next section, these three properties form the basis on which a rigorous mathematical framework can be built; this framework enables us to tackle both theoretical and practical problems associated with chance experiments.

Example 1.9 In a small city, there are just two local newspapers, called "The Weekly News" and the "News Herald." We asked 2000 inhabitants of this city which, if any, of these newspapers they read regularly. 460 persons answered they read "The Weekly News," 580 answered they read the "News Herald," while 140 answered that they read both. Assuming that the number of repetitions, $n = 2000$, is large enough so that the relative frequency of an event has approached its true value (limit), find the probability that if we select a person at random from this city, then he or she will

(i) be a reader of "The Weekly News";

(ii) be a reader of "News Herald";

(iii) read both newspapers;

(iv) read at least one newspaper;

(v) read the “The Weekly News” only;

(vi) be a reader of “News Herald” only;

(vii) will be a reader of exactly one local newspaper.

SOLUTION We consider the experiment of asking a person at random which newspaper(s) he/she reads. Let us define the events

A: a person is a reader of “The Weekly News”

B: a person is a reader of “News Herald.”

The data in the example come from 2000 repetitions of the experiment. Based on this information, we have for the frequencies of the events A, B and their intersection

$$n_A = 460, \quad n_B = 580$$

and

$$n_{AB} = 140.$$

For the relative frequencies, we have

$$f_A = \frac{460}{2000} = 0.23 = 23\%, \quad f_B = \frac{580}{2000} = 0.29 = 29\%, \quad f_{AB} = \frac{140}{2000} = 0.07 = 7\%.$$

Since we have assumed that $n = 2000$ is a sufficiently large number of repetitions, we may estimate the various probabilities by the corresponding relative frequencies. We thus have the following results:

(i) The probability that someone reads the “The Weekly News” is $P(A) = 0.23$.

(ii) The probability that someone reads the “News Herald” is $P(B) = 0.29$.

(iii) The probability that someone reads both newspapers is $P(AB) = 0.07$.

(iv) Here we seek the probability of the event $A \cup B$. We use again the relative frequency as an estimate for this. The number of people who read exactly one newspaper, out of the 2000 persons we asked, is

$$n_{A \cup B} = 460 + 580 - 140 = 900$$

(because when we add $460 + 580$ we have counted the persons who read both newspapers twice). Thus,

$$P(A \cup B) = f_{A \cup B} = \frac{n_{A \cup B}}{n} = \frac{900}{2000} = 0.45;$$

(v) The number of persons who read “The Weekly News” only equals the number of persons who said that they read that newspaper minus the 140 persons who read both newspapers. Consequently,

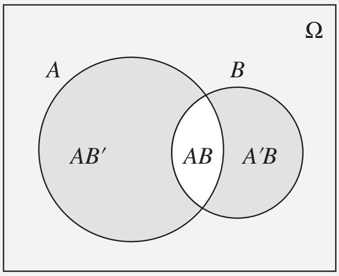
$$P(AB') = f_{AB'} = \frac{n_{AB'}}{n} = \frac{460 - 140}{2000} = 0.16.$$

(vi) Similarly, we have for the number of persons who read the “News Herald” only

$$P(A'B) = f_{A'B} = \frac{n_{A'B}}{n} = \frac{580 - 140}{2000} = 0.22;$$

(vii) The event “someone reads only one of the two newspapers” is expressed as $(AB') \cup (A'B)$. Since the events AB' and $A'B$ are disjoint, we obtain from Property *FP3* that

$$P(AB' \cup A'B) = 0.16 + 0.22 = 0.38.$$



EXERCISES

In the following exercises, assume that the number of repetitions is large enough to secure that the probability in question can be found using the associated relative frequency.

1. Use the first 50 throws from the data in Table 1.1 to calculate the probabilities of the events A, B, and C in Example 1.8. Then, repeat the same calculations using the second half of the throws (numbered 51–100 on the table). What do you observe?

2. For the experiment of throwing a die twice, use the results from Table 1.1 to calculate the probabilities of the following events:

   D: at least one of the two outcomes is 3;

   E: the outcome of the second throw is 5;

   F: the results of the two throws are the same.

   How do the probabilities of the events D and E compare to those of the events A and B in Example 1.8?

3. In a study of religious habits of persons living in a city, 500 persons were asked whether they go to church regularly. 240 of them were men, out of which 40 said that they attend a church service regularly, while among women, 80 said that they attend a church service regularly. Consider the following events, related to a randomly selected person among the persons who took part in this study:

   A: the person selected is male;

   B: the person selected goes to church.

   Express, in terms of the events A and B, each of the following events, and calculate their probabilities:

   (a) the person selected is male and attends church services regularly;

   (b) the person selected is female and attends church services regularly;

   (c) the person selected does not go to church.

4. In a survey conducted to examine whether there is correlation between gender and physical exercise, 700 persons were selected at random and their gender and whether they did some form of physical exercise at least twice a week or not were recorded. The outcomes are given in the following table:

	Physical exercise?	
Gender	Yes	No
Male	80	300
Female	75	245

   Consider the following events, associated with the experiment of selecting randomly a person among the 700 persons who took part in this survey:

   A: the person selected is a man;

   B: the person selected does physical exercise at least twice a week.

   Based on the data from the above table, calculate the probabilities of the events

$$A, B, AB, A', B', A'B', AB', A'B, A'B \cup AB', AB \cup A'B'.$$

5. In a University class, there are 200 students, who at the end of the last semester took three exams: Algebra (A), Calculus (C), and Probability (P). The numbers of students who obtained a first class mark were 24, 12, and 33 for A, C, and P, respectively. The numbers of students who obtained a first class mark in both A and C was 7, 10 students got a first class mark in A and P, and 6 students got a first class mark in C and P. If we select a student at random, find the probability that this student will have a first class mark

   (i) only in Algebra;

   (ii) only in Probability;

(iii) in both Probability and Algebra, but not in Calculus;

(iv) in both Calculus and Probability, but not in Algebra;

(v) in all three courses;

(vi) in none of these courses;

(vii) in exactly one of the three courses.

6. When Jenny returns home from her local supermarket, she has to pass through three sets of traffic lights. During the last 50 times she did that, she had to stop her car at the first set of traffic lights on 30 occasions, while 10 times the first traffic light was green but she had to stop at the second one. Furthermore, 6 times when both the first and the second lights were green, she had to stop at the third. Next time she returns from the supermarket, what is the probability that she has to stop in at least one of the three sets of traffic lights?

1.4 AXIOMATIC DEFINITION OF PROBABILITY

We have seen in the previous section that Definition 1.9 overcomes the difficulties encountered with Laplace's definition of probability. However, the approach by von Mises could not form the basis for a formal mathematical development of probability theory, as it is based on experimental data. Some of the problems which may arise if we use Definition 1.9 to define probability are as follows:

- On several occasions, the nature of the experiment is such that it may not be reasonable or even possible to repeat it a large number of times. Consider, for instance, the probability of the event that a political party wins the forthcoming election; or that a new spacecraft will land successfully on Mars. In some other occasions, when repetitions of an experiment are possible, it may be too costly or time consuming.
- When it is feasible to have a large number of repetitions, there appears to be no systematic way to control the error which arises when we use a relative frequency as an approximation for the probability of an event.
- In Definition 1.9, probability is defined as a limit as $n \to \infty$. How do we know that such a limit always exists? One way to get around this is to *assume* that the limit exists, as part of that definition; however, this is a rather strong assumption to be used as an axiom. Moreover, in practice, we can never have an infinite number of experiments and, especially when precision is essential, it may not be easy to "guess" what the value of the limit

$$\lim_{n\to\infty} f_A = \lim_{n\to\infty} \frac{n_A}{n}$$

  is, when we have at hand only a finite number of repetitions n for the experiment.

The aforementioned concerns initiated a search for alternative definitions of probability. In fact, in a famous lecture delivered in 1900 at the *International Congress of Mathematicians* held in Paris, the German mathematician David Hilbert (1862–1943), who was perhaps the leading mathematician of his time, proposed 23 problems whose solution presented a major challenge and would form a breakthrough for mathematics of the twentieth century. One of these problems concerned the axiomatic foundation of probability theory. It was only in 1933 when the Russian mathematician Andrei Kolmogorov (1903–1989) developed a consistent framework within which any problem regarding probabilities could be, if not solved, at least formulated. In Kolmogorov's work, three "obvious" and "undoubtable" properties of probability are taken as axioms. Based on them, the entire theory of probability can be developed, making a series of mathematical and logical deductions. It is worth noting at this point that, using tools from Kolmogorov's axiomatic foundation of probability, we can prove the existence of the limit (in an appropriate sense) for the relative frequency of an event when the number of repetitions tends to infinity. Thus, Kolmogorov's theory does not contradict the statistical definition by von Mises; rather, it incorporates it by providing a formal reasoning for its validity.

We are now ready to give the definition of probability, based on Kolmogorov's approach, which will be used throughout this book.

Definition 1.10 Assume that Ω is a sample space for a chance experiment. Assume also that to each event A in Ω, there corresponds a real number, $P(A)$. If the function $P(\cdot)$ satisfies the following three axioms, then P will be called a probability on the sample space Ω, while the number $P(A)$ will be referred to as the probability of the event A:

*P*1. $P(A) \geq 0$ for any event A defined in the sample space Ω;

*P*2. $P(\Omega) = 1$;

*P*3. If $A_1, A_2, \ldots$ is a sequence of events in Ω which are pairwise disjoint (that is, $A_i A_j = \emptyset$ for any $i \neq j$), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \tag{1.2}$$

According to this definition, the probability is considered as a **set function**, i.e. a function which associates any event A defined in Ω to a real number. Looking at Kolmogorov's axioms, and comparing them with Properties (*FP1–FP3*) stated in the previous section as direct consequences of von Mises' definition of probability, we see that they are not that different! In fact, the first two are the same, while *P*3 in Definition 1.10 includes *FP3* as a special case (just take $A_3 = A_4 = \cdots = \emptyset$).

Property *P*3 is known as the **additivity property of probability**. More precisely, as it involves an infinite summation (a sum of countably many terms), it is usually called the **countable additivity** axiom of probability. If we put $A_1 = A_2 = A_3 = \cdots = \emptyset$ in (1.2), we obtain immediately

$$P(\emptyset) = P(\emptyset) + P(\emptyset) + P(\emptyset) + \cdots ,$$

so that subtracting $P(\emptyset)$ on either side yields

$$0 = P(\emptyset) + P(\emptyset) + \cdots .$$

Since $P(\emptyset)$ is finite and uniquely defined, it has to be zero, that is,

$$P(\emptyset) = 0.$$

Next, if we put

$$A_{n+1} = A_{n+2} = \cdots = \emptyset$$

in (1.2), we arrive at the following result, known as the **finite additivity property**.

Proposition 1.4 *If* $A_1, A_2, \ldots, A_n$ *are n pairwise disjoint events in a sample space* Ω *(that is,* $A_iA_j = \emptyset$ *for any* $i \neq j$*), then*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i). \tag{1.3}$$

For $n = 2$, as already mentioned, we obtain Property *FP3* of the previous section; namely, if A and B are disjoint events, then

$$P(A \cup B) = P(A) + P(B). \tag{1.4}$$

We have already seen some examples of obtaining new results with the aid of the three axioms in Definition 1.10. However, the definition seems to offer no obvious way to *calculate* the probability $P(A)$ associated with a certain event A. This will be studied and illustrated in great detail in subsequent chapters of the book.

In the next propositions, we develop some useful properties for probabilities, which are easily deduced from the axioms of Definition 1.10.

Proposition 1.5 *Let* $A = \{a_1, a_2, \ldots\}$ *be a countably infinite subset of the sample space* Ω*. Then, the probability of the event A is*

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

where $P(A_i)$ *denotes the probability of the elementary event* $\{a_i\}$*. If A has finitely many elements, so that* $A = \{a_1, a_2, \ldots, a_n\}$*, then*

$$P(A) = \sum_{i=1}^{n} P(A_i).$$

Proof: The first statement is obvious from (1.2) by putting $A_i = \{a_i\}$ for $i = 1, 2, \ldots$, and observing that

$$\bigcup_{i=1}^{\infty} A_i = A \quad \text{and} \quad A_iA_j = \emptyset \quad \text{for } i \neq j.$$

For the second statement, put similarly $A_i = \{a_i\}$ for $i = 1, 2, \ldots, n$ in Proposition 1.4.

□

Applying the previous proposition to the case $A = \Omega$, when Ω is a finite or countably infinite sample space, we immediately obtain the following.

Corollary 1.1 *Let $\Omega = \{\omega_1, \omega_2, \dots\}$ be a countably infinite sample space. Then the probabilities $p_i = P(\{\omega_i\})$ associated with the elementary events $\{\omega_i\}$ satisfy the condition*

$$\sum_{i=1}^{\infty} p_i = 1.$$

For the case when Ω is finite so that $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, this reduces to

$$\sum_{i=1}^{n} p_i = 1.$$

There are certain cases where it is rather complicated to calculate the probability for an event A, but is easier to find the probability of its complement, A'. The following result is then useful.

Proposition 1.6 *Let A be an arbitrary event in a sample space Ω. Then, the probability of its complement, A', is given by*

$$P(A') = 1 - P(A). \tag{1.5}$$

Proof: This is obvious from (1.4) and the fact that

$$P(\Omega) = P(A \cup A') = P(A) + P(A'),$$

which holds since $A \cup A' = \Omega$ and A and A' are disjoint sets. □

Example 1.10 Suppose the probability of an event A is larger by 0.5 than the probability of its complement. Find $P(A)$.

SOLUTION We are given that

$$P(A) = P(A') + 0.5,$$

and replacing $P(A')$ by Proposition 1.5 gives

$$P(A) = 1 - P(A) + 0.5 = 1.5 - P(A).$$

Solving this, we immediately obtain $P(A) = 0.75$.

Example 1.11 Suppose the possible outcomes of an experiment are all the positive integers, while for $i = 1, 2, \dots$, the probability that the outcome of the experiment is i is twice as much as the probability that the outcome is $i + 1$. Then, calculate

(i) the probabilities of all elementary events of this experiment;

(ii) the probability that the outcome is an even integer;

(iii) the probability that the outcome is an odd integer.

SOLUTION The sample space for this experiment is the set $\Omega = \{1, 2, \dots\}$, and so an elementary event is a set $\{i\}$, where i is a positive integer. For any such i, we are given that

$$P(\{i\}) = 2P(\{i+1\}).$$

Denoting $p_i = P(\{i\})$ for simplicity in notation, this reads

$$p_i = 2p_{i+1}, \quad i = 1, 2, \dots .$$

Thus, we have, for example, $p_1 = 2p_2, p_2 = 2p_3$, and so on.

(i) Applying the last equation recursively, we can express the probability of any elementary event, p_i, in terms of p_1. More specifically,

$$p_1 = 2p_2 = 2(2p_3) = 4p_3 = 4(2p_4) = 8p_4 = \dots ,$$

and in general $p_1 = 2^{i-1}p_i$ for $i = 1, 2, \dots$, which can be rewritten as

$$p_i = 2^{-(i-1)}p_1 = \left(\frac{1}{2}\right)^{i-1} p_1, \quad i = 1, 2, \dots .$$

Since Ω is countably infinite, we may use the first part of Corollary 1.1. Replacing the probabilities p_i there by the last expression, we deduce that

$$1 = \sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{i-1} p_1 = p_1 \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{i-1} .$$

The quantity

$$\sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{i-1}$$

is the sum of a geometric series,[5] and so we obtain

$$p_1 \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{i-1} = \left(1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots \right) p_1 = \frac{1}{1 - \frac{1}{2}} p_1 = 2p_1 .$$

Since this must equal 1, we get $p_1 = 1/2$, and the probabilities of the elementary events are then given by

$$p_i = \left(\frac{1}{2}\right)^{i-1} p_1 = \left(\frac{1}{2}\right)^{i}, \quad i = 1, 2, \dots .$$

---

[5] The main results for a geometric series, for readers not familiar with it, are listed in the Appendix A, along with some other formulas from calculus which are needed for the discussions in this book.

(ii) The probability that the outcome of the experiment is an even integer is given by the sum

$$\sum_{i=1}^{\infty} p_{2i} = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{2i} = \sum_{i=1}^{\infty} \left(\frac{1}{4}\right)^{i}.$$

This is another sum of a geometric series, and so is equal to

$$\frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots = \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{1}{3}.$$

(iii) If A denotes the event considered in part (ii), viz., that the outcome is an even integer, then the event that the outcome is an odd integer is simply A', and so we immediately have

$$P(A') = 1 - P(A) = 1 - \frac{1}{3} = \frac{2}{3}.$$

Alternatively, the probability of the event A' could be obtained directly as follows:

$$\begin{aligned} P(A') &= p_1 + p_3 + p_5 + \cdots = \frac{1}{2} + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \cdots \\ &= \frac{1}{2}\left(1 + \frac{1}{2^2} + \frac{1}{2^4} + \cdots\right) = \frac{1}{2}(1 + P(A)) \\ &= \frac{1}{2} \cdot \frac{4}{3} = \frac{2}{3}. \end{aligned}$$

We close this section by a brief reference to a rather subtle issue of the axiomatic foundation of probability, stemming from Definition 1.10. We have mentioned already that if the sample space Ω has at most countably many elements, then we can safely associate a probability to *any* subset A of Ω; for sample spaces with uncountably many elements, there may be subsets of Ω for which this is not possible. Such subsets are called *nonmeasurable* and are not considered here. Excluding such sets from the domain of the probability function $P(\cdot)$, we see that this set function is defined on a family of subsets of Ω, say $\mathcal{A}$, the members of which are events on the sample space. The minimum requirements that $\mathcal{A}$ must satisfy so that Definition 1.10 can be applied are the following:

Σ1: $\Omega \in \mathcal{A}$;

Σ2: If $A \in \mathcal{A}$, then $A' \in \mathcal{A}$;

Σ3: If $A_i \in \mathcal{A}$ for $i = 1, 2, \ldots$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

A family of subsets of the sample space Ω with the above properties is called a **σ-field** or a **σ-algebra**. Many advanced books on probability theory begin by considering such families in order to introduce probability in a more rigorous manner from a mathematical viewpoint. Several such textbooks can be found in the list of references at the end of this book, and the interested reader may refer to any of these books for more details.

EXERCISES

Group A

1. If for the event A we have $2P(A) = P(A') + 0.5$, find the probability $P(A)$.
2. At 08:00 a.m., the probability that John is in bed is 0.2, while the probability of him having breakfast is 0.5. What is the probability that on a particular day he is neither in bed nor having breakfast at 08:00 a.m.?
3. The probability that the price of a certain stock increases during a day is 50% higher than the probability that the price of the stock decreases, while it is also three times as much as the probability that the price remains the same. Find the probability that, during a day, the price of the stock (i) increases; (ii) decreases; (iii) does not change.
4. When a salesman visits a certain town, he stays in one of three available hotels, H_1, H_2, H_3. Let A_i be the event that he stays in Hotel H_i, for $i = 1, 2, 3$. If it is known that

$$P(A_1 \cup A_2) = 2P(A_3) \quad \text{and} \quad 3P(A_2) = 1 - \frac{P(A_3)}{2},$$

   find the probability that he stays in Hotel H_i, for $i = 1, 2, 3$.

5. Let $A_1, A_2, \ldots, A_n$ be a finite collection of events on a sample space Ω such that $A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$ and the A_i's are pairwise disjoint. If it is known that $P(A_i) = 3P(A_{i+1})$ for $i = 1, 2, \ldots, n-1$, find $P(A_3)$ and $P(A_5)$ (assuming $n \geq 5$).
6. Let A and B be two disjoint events in a sample space Ω. Prove that

$$P(A'B') = P(A') + P(B') - 1.$$

7. Let A, B, and C be three events in a sample space Ω such that $AB = \emptyset$. Then:
   (i) Verify that the events AC and BC are disjoint.
   (ii) Calculate the probability of the event

$$(A' \cup C')(B' \cup C')$$

   in terms of the probabilities $P(AB)$ and $P(AC)$.

Group B

8. Consider the infinite countable sample space $\Omega = \{0, 1, 2, \ldots\}$. On this space, assume that the probabilities of the elementary events $\{i\}$, for $i = 0, 1, \ldots$, are given by

$$P(\{i\}) = \frac{3^i}{4^{i+1}}, \quad i = 0, 1, 2, \ldots.$$

   (i) Show that the probability that the outcome of the experiment is at least k, for $k = 1, 2, \ldots$, is given by $(3/4)^k$.
   (ii) What is the probability that we get an outcome less than k for $k = 1, 2, \ldots$?

9. Let A_1, A_2, A_3, and A_4 be four events in a sample space Ω that are pairwise disjoint and such that

$$\bigcup_{i=1}^{4} A_i = \Omega.$$

If it is known that

$$P(A_2) = 2P(A_1), \quad P(A_4) = 3P(A_2), \quad P(A_2) = 4P(A_3),$$

calculate the probabilities of the union of any two and any three events among A_1, A_2, A_3, A_4.

10. For two disjoint events A and B in a sample space, suppose

$$P(A \cup B) = \frac{1}{2}, \quad 3P(A') + 2P(B) = 3.$$

Then, find the probabilities $P(A)$ and $P(B)$.

11. Consider the continuous sample space $\Omega = [1, 1000]$. For each event A in Ω ($A \subseteq \Omega$), we define the probability $P(A)$ to be $k/1000$, where k is the number of integers included in the set A. Then:

   (i) Show that $P(\cdot)$ satisfies the three properties of Definition 1.10.

   (ii) Calculate the probabilities of the events

$$A = [100, 200], \quad B = [1, 100] \cup [900, 1000], \quad C = \bigcup_{i=1}^{9} \left[100i, 100i + \frac{100}{2^{i+1}}\right].$$

1.5 PROPERTIES OF PROBABILITY

In this section, we look at various properties of probability. We have already seen, in the last section, results which follow from the conditions we set in Definition 1.10. Further results in this section, which are important for subsequent developments in this book, may illustrate the development of a chain of results from Kolmogorov's axioms. A major tool for the proofs in this section is the additive property of probability, given in Proposition 1.4. Note, however, that this applies to pairwise disjoint events. Therefore, in order to use that proposition, we must first express our event of interest as the union of two or more events which are pairwise disjoint. This is in fact the technique used in the proofs of the ensuing propositions.

Proposition 1.7 *For any events A and B in a sample space Ω, we have*

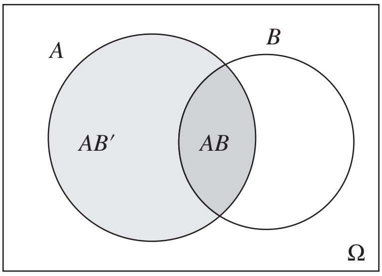
$$P(A - B) = P(AB') = P(A) - P(AB). \tag{1.6}$$

For the special case where $B \subseteq A$, we have that

$$P(A - B) = P(A) - P(B).$$

Proof: Recall that the set $A - B$ contains all elements of A which are not contained in B. This means that the events $A - B$ and AB' are identical, and so they must have the same probability, and this establishes the first part of (1.6). For the second part, note that the events AB and AB' are disjoint, while their union equals

$$(AB) \cup (AB') = A.$$



In view of the finite additivity property, we thus have

$$P(AB) + P(AB') = P(A),$$

and the second equation in (1.6) is now obvious.

Turning to the case where $B \subseteq A$, we then have $AB = B$, and so (1.6) immediately implies that $P(A - B) = P(A) - P(B)$. □

Proposition 1.8 ***(Monotonicity of probability)***

(i) *Let A and B be events in a sample space Ω such that $B \subseteq A$. Then,*

$$P(B) \leq P(A).$$

(ii) *For any event A on Ω, we have*

$$P(A) \leq 1.$$

Proof:

(i) When $B \subseteq A$, Proposition 1.7 yields

$$P(A - B) = P(A) - P(B).$$

Since the left-hand side is nonnegative, so must be the right-hand side. Thus $P(A) - P(B) \geq 0$, that is,

$$P(B) \leq P(A).$$

(ii) This is obvious from Part (i), the fact that for any event A on the sample space we have $A \subseteq \Omega$, and Property $P2$ of Definition 1.10. □

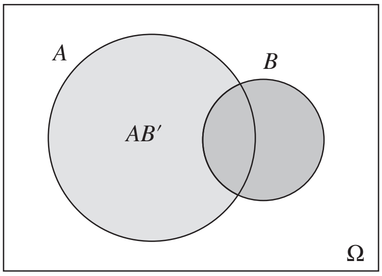
The following result is a key identity which generalizes (1.4), as it enables us to calculate the probability of a union of two events which are not necessarily disjoint.

Proposition 1.9 *For any events A and B on a sample space, we have*

$$P(A \cup B) = P(A) + P(B) - P(AB). \tag{1.7}$$

Proof: As stated above, we know how to deal with the probability of the union between two events only when these events are disjoint. We therefore try to express $A \cup B$ as the union of two such events. This is easily accomplished by considering the events B and AB'; these have no common element between them, while

$$A \cup B = AB' \cup B.$$



From the additivity property, we readily have

$$P(A \cup B) = P(AB' \cup B) = P(AB') + P(B).$$

Upon replacing $P(AB')$ from the result stated in Proposition 1.7, we get

$$P(A \cup B) = P(A) + P(B) - P(AB),$$

as required. □

Example 1.12 A store for women's clothes sells both shoes and handbags. The store manager has estimated that 20% of the customers who enter the shop buy a pair of shoes, 30% purchase a handbag while 10% of the customers buy both a pair of shoes and a handbag. Based on these assertions, estimate the proportions of customers who buy

(i) a pair of shoes only,
(ii) a handbag only,
(iii) at least one of these two items,
(iv) exactly one of these items.

SOLUTION Let us define the events

A: a person buys a pair of shoes from the store,

B: a person buys a handbag from the store.

Then, based on the manager's assertions, we have

$$P(A) = 0.20, \quad P(B) = 0.30, \quad P(AB) = 0.10.$$

The required probabilities are thus found as follows:

(i) $P(AB') = P(A) - P(AB) = 0.20 - 0.10 = 0.10 = 10\%$;

(ii) $P(A'B) = P(B) - P(AB) = 0.30 - 0.10 = 0.20 = 20\%$;

(iii) $P(A \cup B) = P(A) + P(B) - P(AB) = 0.20 + 0.30 - 0.10 = 0.40 = 40\%$;

(iv) $P(AB' \cup A'B) = P(AB') + P(A'B) = 0.10 + 0.20 = 0.30 = 30\%$.

Proposition 1.9 enables us to calculate the probability that at least one of two events occurs. Often, in practice, we have more than two events and we are interested in the probability that at least one of them occurs. When we have three events, say A_1, A_2, and A_3, we may apply Proposition 1.9 twice to obtain the corresponding result. More explicitly, we have

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P[A_1 \cup (A_2 \cup A_3)] \\ &= P(A_1) + P(A_2 \cup A_3) - P[A_1(A_2 \cup A_3)] \quad \text{by Proposition 1.9} \\ &= P(A_1) + P(A_2 \cup A_3) - P(A_1A_2 \cup A_1A_3) \quad \text{by } SP6 \text{ of Proposition 1.1} \\ &= P(A_1) + P(A_2) + P(A_3) - P(A_2A_3) \\ &\quad - [P(A_1A_2) + P(A_1A_3) - P(A_1A_2A_1A_3)] \quad \text{by Proposition 1.9} \\ &= P(A_1) + P(A_2) + P(A_3) - [P(A_1A_2) + P(A_1A_3) + P(A_2A_3)] \\ &\quad + P(A_1A_2A_3). \end{aligned}$$

For the general case when we have n events, the following result, known as the **Poincaré formula**[6] or the **inclusion–exclusion formula**, holds.

Proposition 1.10 ***(Inclusion–exclusion formula)*** *Let $A_1, A_2, \ldots, A_n$ be n events in a sample space Ω. Then, the probability that at least one of them occur is given by*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = S_1 - S_2 + S_3 - \cdots + (-1)^{n-1}S_n = \sum_{r=1}^{n} (-1)^{r-1}S_r,$$

where

$$S_1 = \sum_{i=1}^{n} P(A_i),$$

$$S_2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_iA_j),$$

---

[6] The formula is named after Jules Henri Poincaré (1854–1912), a French mathematician, physicist, engineer as well as a distinguished philosopher of science and mathematics.

$$S_3 = \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} P(A_i A_j A_k),$$

$$\vdots$$

$$S_r = \sum_{i_1=1}^{n-r+1} \sum_{i_2=i_1+1}^{n-r+2} \cdots \sum_{i_r=i_{r-1}+1}^{n} P(A_{i_1} A_{i_2} \cdots A_{i_r}),$$

$$\vdots$$

$$S_n = P(A_1 A_2 \cdots A_n).$$

Proof: The proof can be established by induction, using the result of Proposition 1.9. You may try proving it! □

Example 1.13 An insurance company offers its clients coverage against three different risks, labeled *I*, *II*, and *III*. The following table shows the percentage of insured clients who had opted to insure against one of these risks (diagonal elements in the table), as well as those who had insured against more than one of these risks (for instance, 17% of the clients had insured against risks *I* and *II*). We also know that the percentage of clients who had insured for all three risks is 5%.

	I	*II*	*III*
I	32%	17%	12%
II		23%	9%
III			18%

Calculate the percentage of clients who had insured

(i) for at least one of the three risks;

(ii) for none of the three risks;

(iii) for exactly two among the three risks.

SOLUTION We consider the events

A_i: a customer is insured against risk i,

for $i = 1, 2, 3$. According to the values in the table, we have

$$\begin{array}{lll} P(A_1) = 0.32, & P(A_2) = 0.23, & P(A_3) = 0.18, \\ P(A_1A_2) = 0.17, & P(A_1A_3) = 0.12, & P(A_2A_3) = 0.09. \end{array}$$

In addition, we have $P(A_1A_2A_3) = 0.05$.

(i) For this part, we use the inclusion–exclusion formula, since we seek the probability $P(A_1 \cup A_2 \cup A_3)$ (of course, this agrees with the formula for $n = 3$, given just before Proposition 1.10, so we could have used this formula instead). In this way, we get

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{r=1}^{3} (-1)^{r-1} S_r = S_1 - S_2 + S_3.$$

We now find S_r, for $r = 1, 2, 3$, as follows:

$$S_1 = \sum_{i=1}^{3} P(A_i) = P(A_1) + P(A_2) + P(A_3) = 0.32 + 0.23 + 0.18 = 0.73,$$

$$\begin{aligned} S_2 &= \sum_{i=1}^{2} \sum_{j=i+1}^{3} P(A_i A_j) = P(A_1 A_2) + P(A_1 A_3) + P(A_2 A_3) \\ &= 0.17 + 0.12 + 0.09 = 0.38, \end{aligned}$$

and finally

$$S_3 = P(A_1 A_2 A_3) = 0.05.$$

We thus find the required probability to be

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 0.73 - 0.38 + 0.05 = 0.40.$$

(ii) Here, we want to calculate the probability of the event $A_1' A_2' A_3'$. But, by De Morgan identity, this is the complement of the event $A_1 \cup A_2 \cup A_3$, and consequently

$$P(A_1' A_2' A_3') = 1 - P(A_1 \cup A_2 \cup A_3) = 1 - 0.40 = 0.60.$$

(iii) The event we are interested in can be expressed as

$$E = \underbrace{A_1 A_2 A_3'}_{\text{only } A_1 \text{ and } A_2 \text{ occur}} \cup \underbrace{A_1 A_2' A_3}_{\text{only } A_1 \text{ and } A_3 \text{ occur}} \cup \underbrace{A_1' A_2 A_3}_{\text{only } A_2 \text{ and } A_3 \text{ occur}}.$$

This is the union of three events which are pairwise disjoint; therefore,

$$P(E) = P(A_1 A_2 A_3') + P(A_1 A_2' A_3) + P(A_1' A_2 A_3).$$

From Proposition 1.7, we find the probabilities on the right hand side to be

$$P(A_1 A_2 A_3') = P((A_1 A_2) A_3') = P(A_1 A_2) - P(A_1 A_2 A_3) = 0.17 - 0.05 = 0.12,$$

$$P(A_1 A_2' A_3) = P((A_1 A_3) A_2') = P(A_1 A_3) - P(A_1 A_2 A_3) = 0.12 - 0.05 = 0.07$$

and

$$P(A_1'A_2A_3) = P((A_2A_3)A_1') = P(A_2A_3) - P(A_1A_2A_3) = 0.09 - 0.05 = 0.04.$$

We thus obtain finally the required probability to be

$$0.12 + 0.07 + 0.04 = 0.23,$$

i.e. there is a 23% chance for a client to be insured against exactly two among the three risks.

EXERCISES

Group A

1. For a specific area, the probability that it is hit by a typhoon during a year is 0.02. For the same period, the probability that the area suffers severe damage due to excessive rain is 0.05, while the probability that both these events occur in a year is 0.001. Find the probability that over a particular year, the area
   (i) is hit by a typhoon but there is no excessive rain;
   (ii) experiences only excessive rain, but is not affected by a typhoon;
   (iii) experiences neither excessive rain nor the occurrence of a typhoon;
   (iv) has excessive rain or is hit by a typhoon.

2. When Sandra drives back home from work, she has to pass through two sets of traffic lights. The probability that she has to stop at the first is 0.35, while the probability that she has to stop at the second one is 0.40. Finally, the probability that she does not stop at a traffic light is 0.50. Calculate the probability that she has to stop
   (i) at both sets of traffic lights;
   (ii) at least at one set of traffic lights;
   (iii) at exactly one set of traffic lights.

3. Let A and B be two events in a sample space Ω. Establish the relationship

$$\begin{aligned} P(A')P(B) - P(A'B) &= P(A)P(B') - P(AB') = P(A'B') - P(A')P(B') \\ &= P(AB) - P(A)P(B). \end{aligned}$$

4. Suppose for the events A and B we have

$$2P(A) = 3P(B) = 4P(AB) \quad \text{and} \quad P(A'B) = 0.05.$$

   (i) Calculate the probabilities $P(A), P(B)$, and $P(AB)$.
   (ii) Find the probabilities of the following events

$$A \cup B, \quad A'B', \quad AB', \quad AB' \cup A'B, \quad A \cup B', \quad A' \cup B.$$

5. Let A and B be two events in a sample space Ω with $P(A) = \alpha$ and $P(B) = \beta$. Confirm the validity for each of the following:

   (i) $P(A'B') = 1 - \alpha - \beta + P(AB)$;

   (ii) $\alpha + \beta - 1 \leq P(AB) \leq \min\{\alpha, \beta\}$;

   (iii) $\max\{\alpha, \beta\} \leq P(A \cup B) \leq \alpha + \beta$.

   Suppose $P(A) = 0.75$ and $P(B) = 0.80$. What is the range of possible values for each of $P(A \cup B)$ and $P(AB)$?

6. Let A and B be two events in a sample space Ω such that $P(A) = 1$ and $P(B) = 1$. Use the results of Exercise 5 to show that

$$P(A \cup B) = 1, \quad P(AB) = 1, \quad P(A'B') = 0.$$

7. Let A and B be two events in a sample space Ω such that $P(A) = 0$ and $P(B) = 0$. Use the results of Exercise 5 to show that

$$P(A \cup B) = 0, \quad P(AB) = 0, \quad P(A'B') = 1.$$

8. Let A_1, A_2, and A_3 be three events in a sample space Ω, which are not necessarily pairwise disjoint.

   (i) Prove that

$$P(A_1 \cup A_2 \cup A_3) \leq P(A_1) + P(A_2) + P(A_3).$$

   (ii) Show that the equality

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3)$$

   holds if and only if

$$P(A_1A_2) + P(A_1A_3) + P(A_2A_3) = P(A_1A_2A_3).$$

Group B

9. With reference to the out-patient visits to a hospital, we consider the following events:

   A: a patient is over 50 years of age;

   B: a patient is under 25 years of age;

   C: the condition of a patient upon arrival is considered to be serious;

   D: a patient needs to stay at the hospital for at least one night.

   We are given the following probabilities associated with the above events:

$$P(A) = 0.52, \quad P(B) = 0.22, \quad P(C) = 0.25, \quad P(D) = 0.24,$$

$$P(AC) = 0.17, \quad P(AD) = 0.12, \quad P(BC) = P(BD) = 0.06,$$

$$P(CD) = 0.18, \quad P(ACD) = 0.08.$$

Express in words what each of the following events represents and find the probability of occurrence for them:

$$A'B', \quad (A \cup B)C, \quad AC' \cup D, \quad (A \cup B)D'.$$

10. Suppose A_1, A_2, and A_3 are three events in a sample space. Let us consider the sums

$$\begin{aligned} S_1 &= P(A_1) + P(A_2) + P(A_3), \\ S_2 &= P(A_1A_2) + P(A_1A_3) + P(A_2A_3), \\ S_3 &= P(A_1A_2A_3). \end{aligned}$$

Express mathematically, in terms of S_1, S_2, and S_3, the probability for the event that

(i) exactly one of the A_i occurs;
(ii) exactly two of the A_i occur;
(iii) none of the A_i occur.

Application: Among the customers of an insurance company, the percentages of those who have home insurance, liability insurance and car insurance are respectively, 35%, 15%, and 40%. Suppose 8% have both home and liability insurances, 25% have both home and car insurances, and 10% have both liability and car insurances. Finally, 3% of the customers have all three types of insurance.

For $i = 0, 1, 2$, calculate the percentage of persons who have exactly i types of insurance with the company.

11. Let A, B, and C be three events such that $C \subseteq B \subseteq A$. Let $a = P(A), b = P(B)$, and $c = P(C)$.

(i) Calculate the probabilities

$$P(A - B), \quad P(A - (B - C)), \quad P((A - B) - C)$$

in terms of a, b, c. What do you observe?

(ii) Obtain a numerical answer to the result in (i) for $a = 1/2, b = 1/4, c = 1/6$.

12. For the events A, B, and C of a sample space, we are given

$$\begin{aligned} &P(ABC) = a, \quad P(A'BC) = b_1, \quad P(AB'C) = b_2, \quad P(ABC') = b_3, \\ &P(AB'C') = c_1, \quad P(A'BC') = c_2, \quad P(A'B'C) = c_3. \end{aligned}$$

Calculate the probabilities of the events A, B, and C in terms of a, b_i, and c_i. Similarly find for the events

$$A - B, \quad B - C, \quad (A - B) - C, \quad A - (B - C), \quad (A \cup BC)' - AC.$$

13. Consider the events $A_1, A_2, \dots, A_n$ defined in a sample space Ω. Show that the probability that none of them appear is equal to

$$P(A_1' A_2' \cdots A_n') = \sum_{r=0}^{n} (-1)^r S_r,$$

where $S_1, S_2, \dots, S_n$ are the sums defined earlier in Proposition 1.10, and $S_0 = 1$.

Application: For $n = 3$ assume that $P(A_i) = 1/3$ for $i = 1, 2, 3$. Moreover, assume that $P(A_i A_j) = 1/9$ for any $i \neq j$, and $P(A_1 A_2 A_3) = 1/27$. Then, show that

$$P(A_1' A_2' A_3') = P(A_1')P(A_2')P(A_3').$$

14. For the events A_1, A_2, and A_3 in the same sample space, it is known that

$$P(A_i) = \frac{1}{2^i}, \quad i = 1, 2, 3,$$

and

$$P(A_i A_{i+1}) = \frac{1}{2^{i+2}}, \quad i = 1, 2.$$

If, in addition, we know that the events A_1 and A_3 are mutually exclusive, calculate the probability that at least one of A_1, A_2, A_3 occur.

1.6 THE CONTINUITY PROPERTY OF PROBABILITY

We know from calculus that a real-valued function is continuous if it preserves limits. More precisely, a real function $f : \mathbb{R} \mapsto \mathbb{R}$ is continuous if and only if for each converging sequence of real numbers $(x_n)_{n \geq 1}$, we have

$$\lim_{n \to \infty} f(x_n) = f(\lim_{n \to \infty} x_n).$$

A similar property holds for the probability as a function. Note, however, that probability is not defined on the set of real numbers, but according to Definition 1.10, it is a set function. In order to consider continuity for such a function, we need to introduce the following concepts:

- A sequence of events $A_1, A_2, \dots$, defined in a sample space Ω, is said to be an **increasing sequence** if

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \subseteq A_n \subseteq A_{n+1} \subseteq \cdots .$$

(see Figure 1.15);

- A sequence of events $A_1, A_2, \dots$, defined in Ω, is said to be a **decreasing sequence** if

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \supseteq A_n \supseteq A_{n+1} \supseteq \cdots .$$

(see Figure 1.16).

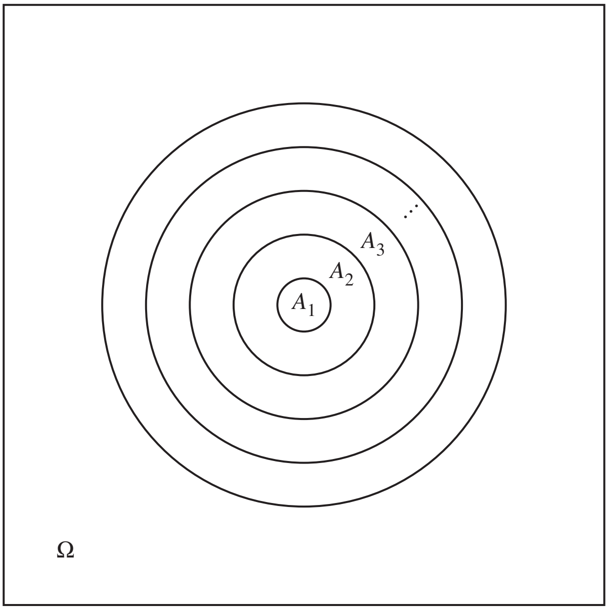


Figure 1.15 An increasing sequence of events.

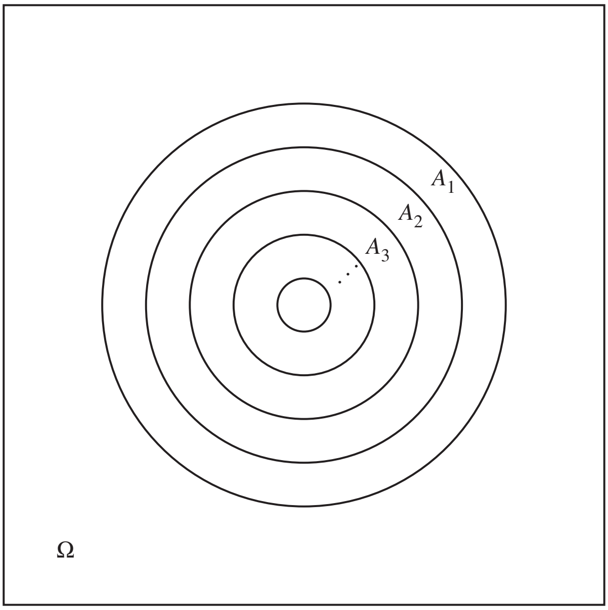


Figure 1.16 A decreasing sequence of events.

It is evident that for an increasing sequence $\{A_n\}_{n\geq 1}$, we have

$$A_n = \bigcup_{i=1}^{n} A_i, \quad n = 1, 2, \ldots .$$

Similarly, for a decreasing sequence $\{A_n\}_{n\geq 1}$, we have

$$A_n = \bigcap_{i=1}^{n} A_i = A_1 A_2 \cdots A_n, \quad n = 1, 2, \ldots .$$

If a sequence of sets is either increasing or decreasing, it is said to be a **monotone sequence**. For such a sequence, we define its **limit**, as follows:

$$\lim_{n\to\infty} A_n = \begin{cases} \bigcup_{i=1}^{\infty} A_i, & \text{if } \{A_n\}_{n\geq 1} \text{ is an increasing sequence} \\ \bigcap_{i=1}^{\infty} A_i & \text{if } \{A_n\}_{n\geq 1} \text{ is a decreasing sequence.} \end{cases}$$

The following example illustrates the above concepts.

Example 1.14 Consider the sample space $\Omega = \mathbb{R}$, the set of real numbers. In it, we define the sequences of events

$$A_n = \left[2 + \frac{1}{n}, 6 - \frac{1}{n}\right], \quad n = 1, 2, 3, \ldots,$$

and

$$B_n = \left[4 - \frac{1}{n}, 4 + \frac{1}{n}\right], \quad n = 1, 2, 3, \ldots .$$

It is easy to see (see also Figure 1.17) that the sequence $\{A_n\}_{n\geq 1}$ is increasing, while the sequence $\{B_n\}_{n\geq 1}$ is decreasing. Further, as $n \to \infty$, the lower endpoint of the set A_n approaches 2, while the upper endpoint converges to 6. Thus,

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n = [2, 6].$$

In an analogous manner, we obtain

$$\lim_{n\to\infty} B_n = \bigcap_{n=1}^{\infty} B_i = \{4\}.$$

We are now ready to prove the following proposition which shows that, when applied to a monotone sequence, probability preserves limits. Apart from its own interest, this is important in proving a number of other important results, as we shall see in later chapters.

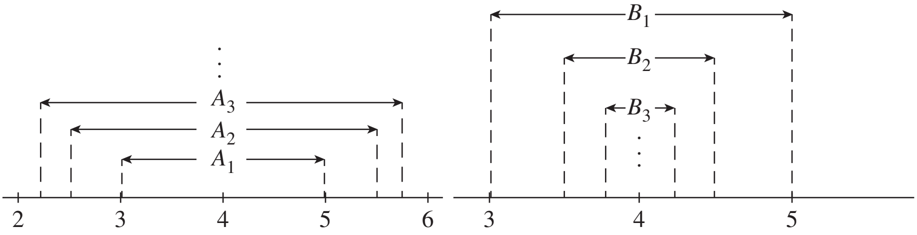


Figure 1.17 The sequences A_n, B_n, $n = 1, 2, \ldots$.

Proposition 1.11 *For any monotone sequence of events* $\{A_n\}_{n\geq 1}$ *of a sample space* Ω, *we have*

$$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n).$$

Proof: Suppose first that the sequence $\{A_n\}_{n\geq 1}$ is increasing. Then, the sets

$$B_1 = A_1, \quad B_2 = A_2 - A_1, \quad B_3 = A_3 - A_2, \ldots, \quad B_n = A_n - A_{n-1}, \ldots$$

are pairwise disjoint, and that they further satisfy the following equalities:

$$\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i = A_n, \quad n = 1, 2, \ldots,$$

and

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i.$$

In words, each set B_n contains those elements of A_n which are not contained in any of the A_i, for $i = 1, 2, \ldots, n-1$. In view of Property *P3* of Definition 1.10, we now obtain

$$P(\lim_{n\to\infty} A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n).$$

However, for the sequence of partial sums in the series on the right hand side, we observe that

$$\sum_{i=1}^{n} P(B_i) = P\left(\bigcup_{i=1}^{n} B_i\right) = P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_n), \quad n = 1, 2, \ldots,$$

and, taking the limit as $n \to \infty$, we deduce

$$\sum_{n=1}^{\infty} P(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P(A_n).$$

This shows that the result of the proposition is true for an increasing sequence. Assume now that the sequence $\{A_n\}_{n\geq 1}$ is decreasing. Then, the sequence of their complements, $\{A_i'\}$, is increasing, so that the previous result applies to them, namely,

$$\lim_{n\to\infty} P(A_n') = P(\lim_{n\to\infty} A_n').$$

But,

$$\lim_{n\to\infty} P(A'_n) = \lim_{n\to\infty} (1 - P(A_n))$$

and so

$$P\left(\lim_{n\to\infty} A'_n\right) = P\left(\bigcap_{i=1}^{\infty} A'_i\right) = P\left(\left(\bigcup_{i=1}^{\infty} A_i\right)'\right) = 1 - P\left(\bigcup_{i=1}^{\infty} A_i\right) = 1 - P\left(\lim_{n\to\infty} A_n\right).$$

It now follows from the above that

$$1 - \lim_{n\to\infty} P(A_n) = 1 - P\left(\lim_{n\to\infty} A_n\right),$$

which completes the proof. □

Example 1.15 Suppose we have a population (1st generation) whose individuals may produce descendants, thus forming the 2nd generation, and so on. We are given that the probability this population becomes extinct by the n-th generation (since all individuals of that generation die before they give birth to any descendants), equals $\exp[-(3n+1)/(4n)]$. What is the probability that the population will never become extinct?

SOLUTION Let us denote by A_n, for $n = 1, 2, \ldots$, the events

A_n: the population becomes extinct by the n-th generation of individuals.

Then, it is apparent that the sequence of events $\{A_n\}_{n\geq 1}$ is increasing, that is

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \subseteq A_n \subseteq A_{n+1} \subseteq \cdots .$$

The event that the population becomes extinct eventually is described by the union

$$\bigcup_{n=1}^{\infty} A_n = \lim_{n\to\infty} A_n,$$

and the probability of this event is given by

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \exp\left(-\frac{3n+1}{4n}\right) = e^{-3/4}.$$

As a consequence, the required probability equals

$$1 - P\left(\bigcup_{n=1}^{\infty} A_n\right) = 1 - e^{-3/4} = 0.52763.$$

EXERCISES

Group A

1. Let the sample space for an experiment be the set of positive integers, and consider the sequence of events $\{A_n\}_{n\geq 1}$ defined in it, where A_n is given by

$$A_n = \{n, n+1, n+2, \dots\}, \quad \text{for } n = 1, 2, \dots .$$

Show that $\{A_n\}_{n\geq 1}$ is a monotone sequence and find its limit as $n \to \infty$.

2. This exercise shows that an arbitrary sequence of events in a sample space can be written as the union of increasing events. Let $\{A_n\}_{n\geq 1}$ be a sequence of events in a sample space (not necessarily monotone), and define a new sequence $\{C_n\}_{n\geq 1}$ by $C_n = \cup_{i=1}^n A_i$, $n = 1, 2, \dots$.

   (i) Show that $\{C_n\}_{n\geq 1}$ is an increasing sequence of events in Ω.

   (ii) Verify that

$$\lim_{n\to\infty} C_n = \bigcup_{i=1}^{\infty} A_i.$$

   (iii) Show that

$$\lim_{n\to\infty} P(C_n) = P\left(\bigcup_{i=1}^{\infty} A_i\right).$$

3. Let the sample space Ω for an experiment be the set of real numbers, a be a real number, and let further the probabilities of the events (in this case, the open intervals)

$$A_n = \left(a - \frac{n}{2}, a + \frac{n}{2}\right), \quad n = 1, 2, \dots,$$

be given by

$$P(A_n) = \frac{1}{n}, \quad n = 1, 2, \dots .$$

Calculate the probability of the elementary event $\{a\}$.

4. Let the sample space Ω for an experiment be the set of real numbers. In this space, we define the events

$$A_n = \left[2 + \frac{1}{2^n}, 4 - \frac{1}{2^n}\right], \quad n = 1, 2, \dots .$$

Assuming that for $n = 1, 2, \dots$, the probability of the event A_n is

$$P(A_n) = \frac{2^n - 1}{2^{n+1}},$$

calculate the probability of the open interval $(2, 4)$.

Group B

5. Let $\{A_n\}$ be a decreasing sequence of events on a sample space, for which we have

$$P(A_1) = q, \quad P(A_n - A_{n+1}) = q^n p, \qquad n = 1, 2, \dots ,$$

where $0 < q < 1$ and $p = 1 - q$.

(i) Show that $P(A_n) = q^n$ for each $n = 1, 2, \dots$.

(ii) Calculate the probability $P(A)$ for the limiting event

$$A = \lim_{n\to\infty} A_n.$$

6. (**Borel–Cantelli lemma**). Let $\{A_n\}$ be a sequence of events in a sample space Ω, such that the sequence of real numbers

$$a_n = \sum_{i=1}^{n} P(A_i), \quad n = 1, 2, \dots ,$$

converges to a real number. Prove that

$$P\left(\bigcap_{r=1}^{\infty} \bigcup_{n=r}^{\infty} A_n\right) = 0,$$

that is, the probability that infinitely many, among the events $A_1, A_2, \dots$, occur is zero. (*Hint*: Apply Proposition 1.11 to the sequence of events $\{B_r\}_{r\geq 1}$, where $B_r = \cup_{n=r}^{\infty} A_n$ for $r = 1, 2, \dots$.)

1.7 BASIC CONCEPTS AND FORMULAS

Sample space Ω	The set of all possible outcomes which may appear in a realization of a chance experiment
An event A in a sample space Ω	A subset of the sample space, $A \subseteq \Omega$
Intersection AB between two events	The event which occurs when both A and B occur
Union $A \cup B$ between two events	The event which occurs when at least one of the events A and B occur
Disjoint events A and B	Two events which cannot happen together, $AB = \emptyset$
Complement A' of an event A	The event which occurs if and only if A does not occur ($A \cup A' = \Omega$ and A, A' disjoint)
Difference $A - B$ between two events	The event which occurs when A does but B does not occur

Relative frequency f_A	n_A/n, where n is the total number of (identical) repetitions of an experiment and n_A is the number of occurrences of A in these repetitions
Frequentist definition of probability	$P(A) = \lim_{n\to\infty} \frac{n_A}{n}$
Axiomatic definition of probability	$P1.$ $P(A) \geq 0$ for any event A on the sample space Ω; $P2.$ $P(\Omega) = 1$; $P3.$ If $A_1, A_2, \ldots$ are pairwise disjoint events, $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$
Properties of probability	$P(\emptyset) = 0$, $P(\Omega) = 1$; $0 \leq P(A) \leq 1$ for any event A; $P(A') = 1 - P(A)$; $P(A) \leq P(B)$ for any events $A \subseteq B$; $P(A - B) = P(AB') = P(A) - P(AB)$; $P(A) = \sum_{i=1}^{\infty} P(a_i)$, for $A = \{a_1, a_2, \ldots\}$; $P(A \cup B) = P(A) + P(B) - P(AB)$, so that if A and B are disjoint, $P(A \cup B) = P(A) + P(B)$.
Continuity property of probability	$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n)$, for any monotone sequence of events $\{A_n\}_{n\geq 1}$.

1.8 COMPUTATIONAL EXERCISES

The die outcomes, as well as the other entries in Table 1.1, were created using the algebraic package Mathematica. Specifically, with the following sequence of commands, we can simulate 100 repetitions for the experiment of throwing two dice, and then calculate the frequencies and relative frequencies for the events of interest.

```
n=100
a=Table[Random[Integer,{1,6}],{n}];
b=Table[Random[Integer,{1,6}],{n}];
s1=0;s2=0;s3=0;
Print["Throw  Outcome    Frequency for  Relative Frequency for"];
Print["nr                  A   B   C       A       B        C"];
Print["---------------------------------------------------------"]
Do[
If[(a[[i]]==6 || b[[i]]==6),s1=s1+1];
If [ a[[i]]==4 , s2=s2+1];
If [ a[[i]]+b[[i]]==7, s3=s3+1];
f1=N[Round[10000*s1/i] /10000,3];
```

```
f2=N[Round[10000*s2/i] /10000,3];
f3=N[Round[10000*s3/i] /10000,3];

Print [i,"        ",a[[i]],"  ",b[[i]],"         ",s1,"   ",s2,"
",s3,"      ",f1,"         ",f2,"         ",f3]
,{i,1,n}]
```

Working in a similar way, use Mathematica (or any other appropriate software, such as Maple, R or a programming language) to obtain simulated repetitions of the experiments described in Exercises 1–7 below, and to find the probabilities of the events $A, B, C, \ldots$ defined in these exercises.

1. We toss a coin four times:

   A: the first and the third outcomes are “Heads”;

   B: exactly two heads appear;

   C: the number of heads is 3 or more.

2. We roll a die twice:

   A: the first outcome is 5 and the second one is greater than 2;

   B: the first outcome is at least 4 and the second outcome is an even integer;

   C: the sum of the two outcomes is 5;

   D: the sum of the two outcomes is at least 9;

   E: the difference between the two outcomes is at most 3.

3. We throw a die until a six appears for the first time:

   A: the experiment is completed after exactly 3 throws;

   B: the experiment is completed after at least 3 throws;

   C: the experiment is completed after at most 3 throws.

4. We throw a die until we get three consecutive identical outcomes (e.g. 111, 222 and so on).

   A: the experiment finishes after more than 20 throws;

   B: the experiment terminates at the 12th throw;

   C: the experiment finishes at or before the 20th throw.

5. Each of two boxes contain 20 red balls and 30 black balls. We select a ball from the first box randomly and place it into the second one. Then, we select a ball from the second box and place it into the first one.

   A: both balls selected were red;

   B: the ball selected from the second box was black;

   C: at the end of the above experiment, each box contains the same number of balls from each color as in the beginning.

6. Each of five boxes contain 15 red balls and 25 black balls. We select a ball from the first box randomly and place it into the second one. Then, we select a ball from the second box and place it into the third one, and so on, until finally we select a ball from the 5th box:

   A: the ball selected from the 5th box is red;

   B: the ball selected from the 5th box has the same color with that selected from the 4th box;

   C: all five balls selected are of the same color.

7. We play the following game 50 times. We throw a die and, if the outcome is $1, 2, 3$ or 4 we win, otherwise we lose:

   A: we win at least 40 times;

   B: we win at most 25 times.

1.9 SELF-ASSESSMENT EXERCISES

1.9.1 True–False Questions

1. When throwing a die, the event "the outcome is a multiple of 3" is an elementary outcome.

2. For two events A and B in a sample space, the probability that at least one of them occurs is $P(A \cup B)$.

3. The sample space for an experiment is the closed interval $[0, 4]$ of the real line. Then, Ω is a finite sample space.

4. If A, B are disjoint events in a sample space and we know that A occurs, then B does not occur.

5. If an event A satisfies the condition $P(A) = 1 - 3P(A')$, then $P(A) = 1/4$.

6. For any events A and B in a sample space, we have $(A \cup B)' = A'B'$.

7. Let A be an event in a sample space such that $P(A) = 1$, and B be another event on that space. Then, $P(A \cup B') = 0$.

8. Let A, B, and C be three events in a sample space. If $A \subseteq B$ and the events A and C are disjoint, then the events B and C are also disjoint.

9. Assume that A, B, and C are three events on a sample space. If the events A and C are disjoint, and the events A and B are disjoint, then B and C are also disjoint events.

10. We toss a coin 300 times and observe that in 160 of these tosses the outcome is "Heads." Based on this, the relative frequency of the event "Heads occur in a single toss of a coin" is 160.

11. For any events A, B, and C in the same sample space, we have

$$A(BC) = (AB) \cup (AC).$$

12. If $P(A) = 1/4$ and $P(A') = 5P(B) - 1$, then the probability of the event B is also $1/4$.

13. Let A and B be two events in a sample space Ω such that $A \subseteq B$, and C be another event in Ω. If the events B and C are disjoint, then A and C will also be disjoint events.

14. Let A and B be two events in a sample space Ω with $A \subseteq B$. Then, $A' \cup B = \Omega$.

15. If for the events A and B we know that $P(A) - P(B) = 1/3$, then $P(A - B) \leq 1/3$.

16. If two events A and B, defined in the same sample space Ω, are mutually exclusive, then $A' \cup B = \Omega$.

17. Let $A_1, A_2, \ldots$ be a sequence of events defined in a sample space Ω, and let a new sequence $\{B_n\}_{n \geq 1}$ be defined by

$$B_n = A_1 A_2 \cdots A_n, \quad n = 1, 2, \ldots .$$

Then, $\{B_n\}_{n \geq 1}$ is a decreasing sequence of events in Ω.

18. Let $A_1, A_2, \ldots$ be a sequence of events defined in a sample space Ω, and let a new sequence $\{C_n\}_{n \geq 1}$ be defined by

$$C_n = \bigcup_{i=n+1}^{\infty} A_i, \quad n = 1, 2, \ldots .$$

Then, $\{C_n\}_{n \geq 1}$ is a decreasing sequence of events in Ω.

19. Let $A_1, A_2, \ldots$ be a monotone sequence of events defined in a sample space Ω. If it is known that

$$P(A_i) = \frac{1}{5}\left(\frac{2}{3}\right)^i, \quad \text{for } i = 1, 2, \ldots ,$$

then the probability of the event that at least one of the A_i's occur is equal to zero.

1.9.2 Multiple Choice Questions

1. The event that exactly one of the three events A, B, and C in a sample space Ω occurs is

(a) $(ABC)'$ (b) $(A \cup B \cup C)'$

(c) $AB'C'$ (d) $(AB'C') \cup (A'BC') \cup (A'B'C)$

(e) $AB \cup C$

2. For the events A and B in a sample space Ω, we know that

$$P(A) = (1 + \lambda)/3, \quad P(B) = 1 - \lambda^2$$

for some real number λ. The admissible range of values for λ is

(a) $\lambda \geq 0$ (b) $0 \leq \lambda \leq 1$ (c) $-1 \leq \lambda \leq 1$

(d) $0 \leq \lambda \leq 2/3$ (e) $-1/3 \leq \lambda \leq 2/3$.

3. Let A, B, and C be three events in a sample space Ω, such that $P(A) = 2P(B)$ and $P(B) = 2P(C)$. If in addition we have $A = (B \cup C)'$, then the probability of the event A is

   (a) $1/7$ (b) $2/7$ (c) $4/7$ (d) $1/2$ (e) $3/5$.

4. Marc tosses a coin until "Heads" appear for the second time. He is interested in the number of "Tails" which appear before the second appearance of "Heads." Then, a suitable sample space for this experiment is

   (a) $\{0, 1, 2, \dots\}$ (b) $\{0, 1, 2\}$ (c) $\{1, 2\}$ (d) $\{2, 3, \dots\}$ (e) $\{1, 2, \dots\}$.

5. In a single throw of a die, consider the events $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$. The event $(B - A)'$ is equal to

   (a) $\{4\}$ (b) $\{2, 4, 5, 6\}$ (c) $\{1, 3\}$

   (d) $\{4, 6\}$ (e) $\{1, 2, 3, 5\}$.

6. For the events A and B in a sample space Ω, we know that $P(A \cup B) = 1/6$. Then, the value of $P(A') - P(B)$ is

   (a) always equal to $5/6$ (b) equal to $5/6$ provided that $AB = \emptyset$

   (c) always equal to $1/6$ (d) equal to $1/6$ provided that $AB = \emptyset$

   (e) none of the above.

7. Maria is waiting at a bus stop for her friend Sarah so that they meet and go to a concert together. Maria wants to know how long she will have to wait until the next bus arrives at the bus stop and whether Sarah will be on it. A suitable sample space for this experiment is (a "0" below indicates that Sarah will not be on the next bus, while a "1" indicates she will be; further, t stands for Maria's waiting time until the arrival of the next bus):

   (a) $\Omega_1 = \{0, 1\}$ (b) $\Omega_2 = \{(1, t) : t \geq 0\}$

   (c) $\Omega_3 = \{(i, t) : i = 0, 1 \text{ and } t \geq 0\}$ (d) $\Omega_4 = \{0, 1\} \cup [0, \infty)$

   (e) $\Omega_5 = \{0, 1\} \cap [0, \infty)$.

8. Which of the following statements is correct with reference to the sample spaces Ω_i, for $i = 1, 2, 3, 4, 5$, defined in the last problem?

   (a) Ω_1 and Ω_5 are the only sample spaces which are finite

   (b) Ω_2 is a discrete sample space, but Ω_4 is not

   (c) Ω_4 is a discrete sample space, but Ω_2 is not

   (d) the set Ω_5 has infinitely many elements

   (e) Ω_3 is an infinitely countable sample space.

9. Let A, B, and C be three events in a sample space Ω, such that $A \cup B \cup C = A$. Then, the following is always true:

   (a) $B \cup C = A$ (b) $B \cup C \subseteq A$ (c) $A \subseteq B \cup C$

   (d) $BC = A$ (e) $A \subseteq BC$.

10. We throw a die until a six appears for the first time, at which point the experiment stops. Let A_i be the event that the outcome of the ith throw is a six. The event "a six appears for the first time in the 3rd throw" can be expressed in terms of the sets A_i as

    (a) $A_1A_2A_3$ (b) $A_1A_2A_3'$ (c) $A_1'A_2'A_3$

    (d) $A_1 \cup A_2 \cup A_3$ (e) $A_1 \cup A_2 \cup A_3'$.

11. We toss a coin successively. Let B_i be the event that the outcome of the ith toss is Heads. Then, the event "Heads occur for the first time at, or after, the second toss" can be written, in terms of the sets B_i, as follows:

    (a) B_1 (b) B_1B_2 (c) $B_1'B_2$ (d) $B_1'B_2B_3B_4\cdots$ (e) B_1'.

12. For the disjoint events A and B in a sample space Ω, we know that

    $$P(A \cup B) = 5/6, \quad 4P(A) + P(B') = 1.$$

    Then, the probabilities of the events A and B are, respectively, equal to

    (a) $P(A) = 1/6, \quad P(B) = 2/3$ (b) $P(A) = 2/3, \quad P(B) = 1/6$

    (c) $P(A) = 1/3, \quad P(B) = 1/2$ (d) $P(A) = 1/2, \quad P(B) = 1/3$

    (e) $P(A) = 1/18, \quad P(B) = 7/9$.

13. At a large University class, there are 140 male students, 40 of whom own a car, and 160 female students, 20 of whom own a car. Let A be the event that "a randomly selected student is female" while B represents the event "a student owns a car." Using the relative frequency as an estimate for the probability of an event, the probability of the event $A'B'$ equals

    (a) 7/15 (b) 1/15 (c) 7/8 (d) 5/7 (e) 1/3.

14. When Carol visits the local supermarket she buys a pack of crisps with probability 0.3, a chocolate with probability 0.4 and her favorite fruit juice with probability 0.6. The probability that she buys both crisps and chocolate is 0.2, that she buys both crisps and the fruit juice is 0.45, and that she buys chocolate and the fruit juice is 0.15. The probability that she will buy at least one of these three products in her next visit is

    (a) equal to 0.3 (b) equal to 0.7 (c) equal to 0.5

    (d) at least 0.5 (e) at most 0.5.

1.10 REVIEW PROBLEMS

1. In a major athletics competition, an athlete has three attempts to clear a certain height. If he succeeds in his first or second attempt, he makes no other attempts on this height.

   (i) Write down a suitable sample space Ω for this experiment and identify which elements of that space are included in the event
   E: the athlete clears the height in one of his three attempts.

   (ii) Let E_i be the event that "the athlete makes a total of i attempts in this height," for $i = 1, 2, 3$. Identify the elements of each event E_i.

   (iii) Express in terms of the events E_1, E_2, and E_3 each of the following events:

   A: the athlete fails in his first attempt;

   B: the athlete fails in his first two attempts;

   C: the athlete clears the height in his second attempt.

   Can you do the same for the event E in Part (i)? If yes, explain how; otherwise, explain why not.

2. Lena, Nick and Tom, who work for the same company, when they arrive for work one morning, meet at the ground floor elevator of the company building, which has three floors above the ground floor. Write down a sample space for the experiment which describes at which floor each person is going to. Then, identify the following events as suitable subsets of this sample space:

   A_1: Lena will leave the elevator on the first floor.

   A_2: Nick and Tom will leave the elevator on the second floor.

   A_3: none of the three will get off on the first floor.

   A_4: Lena and Nick will not go beyond the second floor.

   A_5: Nick is the only person to leave the elevator on the first floor.

   A_6: all three are going to different floors of the building.

   A_7: at least two persons are going to the same floor of the building.

3. The claims arriving at an insurance company are classified by the company as either Large (L) or Small (S) according to their size. The company wants to study the number of claims arriving prior to the first large claim.

   (i) Give an appropriate sample space Ω which can be used for the above study. Is Ω a finite sample space, a countably infinite or a continuous one?

   (ii) Write down explicitly each of the following events:

   A: the first large claim is the ith claim to arrive at the company, for $i = 1, 3, 5, 10$;

   B: the number of claims until the arrival of the first large claim is an even integer;

   C: the number of claims until the arrival of the first large claim is an odd integer;

D: the number of claims until the arrival of the first large claim is at most 5;

E: the number of claims until the arrival of the first large claim is greater than 5.

Which of these subsets of the sample space Ω are finite sets and which ones are countably infinite?

4. Let A and B be events in a sample space Ω for which it is known that $P(A) \geq a$ and $P(B) \geq b$, where a and b are given real numbers.

   (i) Show that
   $$P(AB) \geq a + b - 1.$$

   (ii) If $a + b > 1$, what do you conclude about the events A and B?

5. Assume that A, B, and C are three events in a sample space Ω. Show that the following relations hold:

   (i) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A'BC) - P(AB'C)$
   $-P(ABC') - 2P(ABC)$;

   (ii) $P(ABC) \geq P(A) + P(B) + P(C) - 2$;

   (iii) $P(A'BC) + P(AB'C) + P(ABC') \leq 2P(A' \cup B' \cup C')$;

   (iv) $P(AB'C') + P(A'BC') + P(A'B'C) \leq 2P(A \cup B \cup C)$.

6. Assume that the probability of each elementary event $\{i\}$ defined in the sample space $\Omega = \{1, 2, 3, \dots\}$ is given by
   $$P(\{i\}) = 5^{i-1}/7^i, \quad i = 1, 2, 3, \dots .$$
   Let us define the events
   $$A_n = \{n, n+1, n+2, n+3, n+4\}, \quad n = 1, 2, 3, \dots .$$

   (i) After establishing that the relationship
   $$P(A_n) = \left(\frac{5}{7}\right)^{n-1} P(A_1), \quad n = 1, 2, 3, \dots ,$$
   holds, verify that $\lim_{n\to\infty} P(A_n) = 0$.

   (ii) Can you use the result of Proposition 1.11 to conclude that
   $$P\left(\lim_{n\to\infty} A_n\right) = 0 \quad ?$$

7. In the experiment of throwing a die twice, consider the following events:

   A: the first outcome is 6;

   B: the second outcome is 4;

   C: the sum of the two outcomes is 9;

   D: the first outcome is greater than the second.

Provide a suitable sample space for this experiment and then find which elements of this sample space each of the events below contains:

$$AB, AB', (AB') \cup D, BD, BCD, BC'D, A \cup (C'D'),$$
$$ABCD, ABC'D, A \cup B \cup C \cup D.$$

8. For the experiment of throwing a die twice, we consider again the events A, B, C, and D from the last exercise and denote by E the event "exactly two among the events A, B, C, and D occur":

   (i) Express the event E in terms of the events A, B, C and D.

   (ii) Find which elements of the sample space are contained in E.

9. On a particular day, a restaurant has a special three-course menu with the following choices:

Course	Choices
Starter	Caesar salad or shrimp cocktail or stuffed mushrooms
Main course	Roast beef or chicken or fish or vegetarian lasagna
Dessert	Ice cream or profiterole or fresh strawberries

   Poppy, who is visiting this restaurant with her friends, is to choose one course from each category above.

   (i) How many outcomes does the sample space for this experiment have?

   (ii) Let A be the event that she chooses stuffed mushrooms for a starter. How many outcomes are in the event A?

   (iii) Suppose B is the event she decides to have vegetarian lasagna as her main course. List the outcomes of the sample space which the event B contains.

   (iv) With A and B as defined above, how many outcomes are there in the event AB?

   (v) How many outcomes are there in the event $A \cup B$?

10. We consider the events $A_1, A_2, \ldots, A_n$ of a sample space Ω and, from these events, we form n new events $B_1, B_2, \ldots, B_n$ defined as follows: $B_1 = A_1$, while for $i = 2, 3, \ldots, n$,

$$B_i = A_i - \left(\bigcup_{j=1}^{i-1} A_j \right).$$

   (i) Verify that the events $B_1, B_2, \ldots, B_n$ are pairwise disjoint and that they satisfy the relations

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i$$

   and

$$P(B_i) \leq P(A_i), \quad i = 1, 2, \ldots, n.$$

(ii) Express the probability of the event

$$A = \bigcup_{i=1}^{n} A_i$$

in terms of the probabilities of the events B_i, $i = 1, 2, \dots, n$;

(iii) Show that the following inequality holds:

$$P\left(\bigcup_{i=1}^{n} A_n\right) \le \sum_{i=1}^{n} P(A_i).$$

This is known as the **subadditive property** of probability (or as **Boole's inequality**), and it is also true for an infinite collection of events $A_1, A_2, \dots$.

11. Let $A_1, A_2, \dots, A_n$ be an arbitrary collection of n events in a sample space Ω. Show that

$$P(A_1 A_2 \cdots A_n) \ge 1 - \sum_{i=1}^{n} P(A_i') = \sum_{i=1}^{n} P(A_i) - (n-1).$$

This is known as **Bonferroni's inequality**.

(*Hint*: Apply Boole's inequality from the last exercise to the events A_i', $i = 1, 2, \dots, n$.)

12. Let Ω be a sample space and suppose we have defined a set function, $P(\cdot)$, which satisfies the properties $P1$–$P3$ of Definition 1.10, on that space. Examine whether each of the set functions defined below can be used as a probability on that space; here, A is an arbitrary event on Ω.

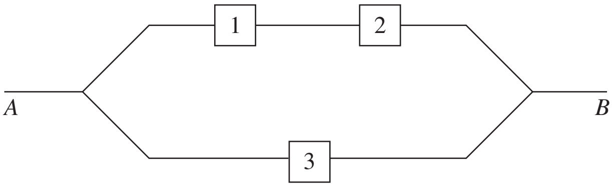
(i) $P_1(A) = P(A)/2$
(ii) $P_2(A) = 1 - P(A)$
(iii) $P_3(A) = |P(A)|$
(iv) $P_4(A) = [P(A)]^2$
(v) $P_5(A) = (P(A) + 1)/2$
(vi) $P_6(A) = (1 - P(A))/2$.

13. A water network has three connections, C_1, C_2, C_3, as shown in Figure 1.18. For each connection, at the places marked $1, 2, 3$, some switches have been put and at a particular instant, any switch can be either ON (thus allowing water to pass through) or OFF. A connection is considered to be working if water can pass from point A to point B.

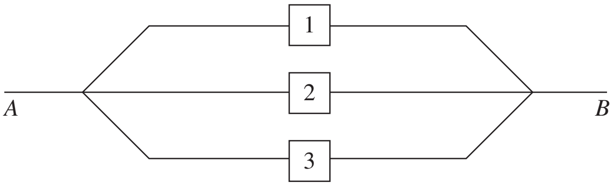
(i) Give a suitable sample space to describe the possible switch positions in the network.

(ii) For each connection, identify the events that describe the working status of the connection.

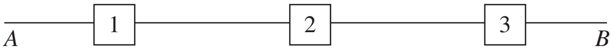
(iii) Let us denote by p_i, $i = 1, 2, 3$, the probability that the switch i is in the ON position and by p_{ij} for $i \ne j$ the probability that both switches i and j are ON ($i = 1, 2, 3$ and $j = 1, 2, 3$). Finally, p_{123} denotes the probability that all three switches are in the ON position. Express the probabilities of the events found in Part (ii) in terms of $p_1, p_2, p_3, p_{12}, p_{13}, p_{23}$, and p_{123}.



Connection C_1



Connection C_2



Connection C_3

Figure 1.18 Water network with three connections.

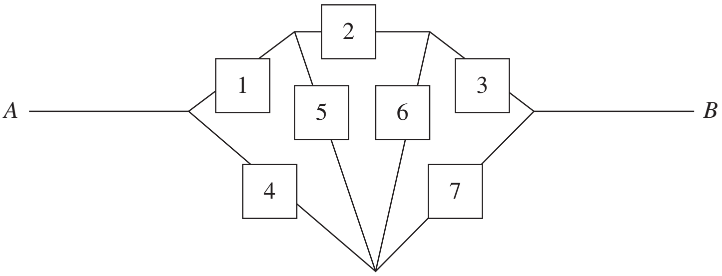
1.11 APPLICATIONS

1.11.1 System Reliability

Reliability engineering is a field that deals with the study of the performance of units/components or systems, i.e. sets of interacting or interdependent components that have been assembled as an integrated whole to carry out a specific task.

In general, the **reliability** of a system (or a unit) is defined as the ability of the system to perform and maintain the function(s) it has been specifically designed for. In probability theory, the term reliability is traditionally used to describe the probability that a system is capable of completing its expected function during an interval of time. Reliability evaluation is of crucial interest in reliability engineering, since it can be used effectively during the design stage so that its performance is optimized. The theoretical results of reliability theory have a lot of applications in a wide spectrum of areas such as airplane industry, nuclear accelerators safety, telecommunications networks, fluid transportation systems, and so on. It is worth mentioning that the failure of mechanical devices such as pumps, network relay stations, ships, airplane engines, etc. is similar in many ways to the death of biological organisms. In this case, the death of the organism should be interpreted as failure of the unit/system and the reliability corresponds to the death rate.

As an illustration, let us consider the following pipeline system, used for transferring water from point A to point B (arcs 1-7 represent seven pump stations). This system consists of 7 units (the 7 pump stations) and each of them can be either working (permits flow of the water through it) or not working:



The system as a whole is working if water flow from point A to point B is feasible through working pumps.

It is clear that if components 1, 2, and 3 work, the system will be functioning no matter what the other components' states are (working or not working). If we denote by $I = \{1, 2, \dots, 7\}$ the set of all components of the system, then the subset $P_1 = \{1, 2, 3\}$ will be called a **path set** of the system in the sense that it describes a path of working components that guarantees the functioning of the system. On the other hand, there exists specific subsets of I with the following property: If all components of the subset are down, the system will not work no matter what the states of the rest of the components are. As an example, we mention $C_1 = \{1, 4\}$ or $C_2 = \{3, 4, 5, 6\}$. These sets are usually referred as **cut sets** of the system.

Should one be able to spot out all the path sets (or all the cut sets) of a system, then he/she could describe its operation (or breakdown) as a union of events and subsequently calculate its reliability by applying the inclusion–exclusion formula (see Proposition 1.10).

Now, we shall present an illustrative example of reliability evaluation for a specific structure.

Let n components be linearly connected in such a way that the system fails if and only if at least k consecutive components fail. This type of structure is usually called **consecutive-*k*-out-of-*n*: *F* system** and has a lot of applications in several areas such as network telecommunication systems and, fluid transportation networks. Let us consider, for example, a sequence of n microwave stations designed to transmit information from place A to place B. Assume that the microwave stations are equidistantly spaced between places A and B and each station is able to transmit information a distance up to k microwave stations. Then, it can be readily verified that the system fails if and only if at least k consecutive microwave stations fail.

From the definition, it is clear that, if we denote by $I = \{1, 2, \dots, n\}$ the set of system components, the cut sets of the system can be described as follows.

$$C_i = \{i, i+1, \dots, i+k-1\}, \quad i = 1, 2, \dots, n-k+1.$$

Denoting by A_i, $i = 1, 2, \dots, n-k+1$, the event

$$A_i : \text{all the components contained in } C_i \text{ are not working,}$$

the probability that the system does not work equals $P\left(\bigcup_{i=1}^{n-k+1} A_i\right)$ and consequently the reliability R of the system can be expressed as

$$
\begin{aligned}
R &= P(\text{systems works}) = 1 - P(\text{system does not work}) \\
&= 1 - P\left(\bigcup_{i=1}^{n-k+1} A_i\right).
\end{aligned}
$$

As an example, let us consider the special case of $n = 5$ and $k = 3$. Such a system consists of 5 components and fails if and only if 3 consecutive components fail. Therefore,

$$C_1 = \{1, 2, 3\}, \quad C_2 = \{2, 3, 4\}, \quad C_3 = \{3, 4, 5\}$$

and A_1, A_2, and A_3 are the events

$$
\begin{aligned}
A_1 &: \text{ components } 1, 2, 3 \text{ do not work,} \\
A_2 &: \text{ components } 2, 3, 4 \text{ do not work,} \\
A_3 &: \text{ components } 3, 4, 5 \text{ do not work,}
\end{aligned}
$$

while the system reliability is given by

$$R = 1 - P(A_1 \cup A_2 \cup A_3).$$

It should be stressed that, besides the three cut sets given above, one might verify that there exist some additional cut sets that are supersets of C_1, C_2, and C_3, such as $C_4 = \{1, 2, 3, 4\}$, $C_5 = \{2, 3, 4, 5\}$, etc. However, these cut sets are not necessary for our analysis since the inclusion of the associated events in the union appearing in the RHS of the last expression does not affect the value of the resulting probability.

Exploiting the inclusion–exclusion formula, we get

$$R = 1 - S_1 + S_2 - S_3,$$

where

$$
\begin{aligned}
S_1 &= P(A_1) + P(A_2) + P(A_3), \\
S_2 &= P(A_1A_2) + P(A_2A_3) + P(A_1A_3), \\
S_3 &= P(A_1A_2A_3).
\end{aligned}
$$

Let us denote by q_i the probability that component i is not working, $q_{i,j}$ the probability that both components i and j are not working ($i \neq j$), $q_{i,j,k}$ the probability that all three components i, j, k are not working ($i \neq j \neq k$) and so on. Then, it is clear that

$$S_1 = q_{123} + q_{234} + q_{345}.$$

Since the event A_1A_2 occurs if and only if all components $1, 2, 3, 4$ are down, A_2A_3 occurs if and only if all components $2, 3, 4, 5$ are down, and A_1A_3 occurs if and only if all components $1, 2, 3, 4, 5$ are down, we conclude that

$$S_2 = q_{1234} + q_{2345} + q_{12345}.$$

Likewise, the event $A_1A_2A_3$ occurs if and only if all components $1, 2, 3, 4, 5$ are down, and so $S_3 = q_{12345}$. Therefore, the system reliability is given by

$$R = 1 - (q_{123} + q_{234} + q_{345}) + q_{1234} + q_{2345}. \tag{1.8}$$

Let us next consider the case when the components of the system are of the same quality, i.e. each one of them has the same probability, say p $(0 < p < 1)$, to be functioning (working) at a specific time point. If, in addition, we assume that the components work independently of each other, as we shall see later on in Chapter 3, the quantities appearing in (1.8) can be expressed as $q_{ijk} = (1-p)^3$ for all $i \neq j \neq k$, and $q_{ijk\ell} = (1-p)^4$ for $i \neq j \neq k \neq \ell$. In this case, the reliability of the system in (1.8) becomes

$$R = 1 - 3(1-p)^3 + 2(1-q)^4 = p + 3p^2 - 5p^3 + 2p^4. \tag{1.9}$$

In Figure 1.19, we present a plot of R as a function of p. Clearly, $r = R(p)$ is an increasing function of p (for $0 \leq p \leq 1$) attaining the value 0 for $p = 0$ and 1 for $p = 1$. In Table 1.2, we provide the values of $R = R(p)$ for several values of p.

Based on the above results, one may easily answer the question: What is the quality p of the components that guarantees a specific reliability level, say R_0, for the system? More explicitly, we see that we have to solve the inequality

$$p + 3p^2 - 5p^3 + 2p^4 \geq R_0,$$

or make use of Table 1.2 and look for the smallest value of p that satisfies the inequality $R(p) \geq R_0$. Therefore, if we wish to maintain a system reliability (at least) 95%, we should use components with functioning probability $p = 0.73$; instead, if we are satisfied with a system reliability 90% we can use components with $p = 0.65$, and so on.

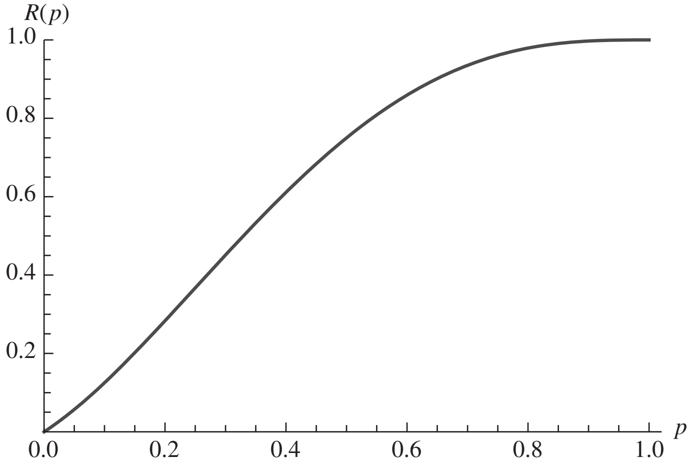


Figure 1.19 Plot of the system reliability, $R = R(p)$, as a function of p.

p	$R(p)$	p	$R(p)$	p	$R(p)$	p	$R(p)$
0.01	0.0103	0.26	0.3841	0.51	0.7623	0.76	0.9652
0.02	0.0212	0.27	0.4009	0.52	0.7744	0.77	0.9691
0.03	0.0326	0.28	0.4177	0.53	0.7861	0.78	0.9727
0.04	0.0445	0.29	0.4345	0.54	0.7975	0.79	0.9761
0.05	0.0569	0.30	0.4512	0.55	0.8086	0.80	0.9792
0.06	0.0697	0.31	0.4678	0.56	0.8194	0.81	0.9820
0.07	0.0830	0.32	0.4843	0.57	0.8299	0.82	0.9846
0.08	0.0967	0.33	0.5007	0.58	0.8400	0.83	0.9869
0.09	0.1108	0.34	0.5170	0.59	0.8498	0.84	0.9890
0.10	0.1252	0.35	0.5331	0.60	0.8592	0.85	0.9909
0.11	0.1399	0.36	0.5491	0.61	0.8683	0.86	0.9925
0.12	0.1550	0.37	0.5649	0.62	0.8771	0.87	0.9940
0.13	0.1703	0.38	0.5805	0.63	0.8855	0.88	0.9952
0.14	0.1858	0.39	0.5960	0.64	0.8936	0.89	0.9963
0.15	0.2016	0.40	0.6112	0.65	0.9014	0.90	0.9972
0.16	0.2176	0.41	0.6262	0.66	0.9088	0.91	0.9979
0.17	0.2338	0.42	0.6410	0.67	0.9159	0.92	0.9985
0.18	0.2501	0.43	0.6555	0.68	0.9227	0.93	0.9990
0.19	0.2666	0.44	0.6698	0.69	0.9291	0.94	0.9994
0.20	0.2832	0.45	0.6839	0.70	0.9352	0.95	0.9996
0.21	0.2999	0.46	0.6977	0.71	0.9410	0.96	0.9998
0.22	0.3166	0.47	0.7112	0.72	0.9464	0.97	0.9999
0.23	0.3335	0.48	0.7244	0.73	0.9516	0.98	1.0000
0.24	0.3503	0.49	0.7374	0.74	0.9564	0.99	1.0000
0.25	0.3672	0.50	0.7500	0.75	0.9609	1.00	1.0000

Table 1.2 The system reliability for several values of functioning probability p.

In closing, we mention that the analysis of the system could also be carried out by working with path sets instead of cut sets. It is not difficult to verify that the path sets of the system are

$$P_1 = \{3\}, \quad P_2 = \{1,4\}, \quad P_3 = \{2,5\}, \quad P_4 = \{2,4\}$$

(we have again excluded the path sets that contain – as a subset – any of P_1, P_2, P_3, P_4) and, introducing the events,

$$B_i : \text{all components in the path set } i \text{ are working},$$

for $i = 1, 2, 3, 4$, we may express the system reliability as

$$R = P(B_1 \cup B_2 \cup B_3 \cup B_4) = S_1 - S_2 + S_3 - S_4,$$

where

$$\begin{aligned}
S_1 &= P(B_1) + P(B_2) + P(B_3) + P(B_4),\\
S_2 &= P(B_1B_2) + P(B_1B_3) + P(B_1B_4) + P(B_2B_3) + P(B_2B_4) + P(B_3B_4),\\
S_3 &= P(B_1B_2B_3) + P(B_1B_2B_4) + P(B_1B_3B_4) + P(B_2B_3B_4),\\
S_4 &= P(B_1B_2B_3B_4).
\end{aligned}$$

The probabilities involved in S_1, S_2, S_3, S_4 can be expressed in terms of

p_i : component i is working,

p_{ij} : both components i and j are working $(i \neq j)$,

p_{ijk} : all three components $i, j,$ and k are working $(i \neq j \neq k)$,

and so on. For example,

$$P(B_1) = p_3, \quad P(B_1B_2) = p_{134}, \quad P(B_3B_4) = p_{245}, \text{ and } P(B_2B_3B_4) = p_{1245}.$$

In the special case of independent components with equal working probability p, we can make use of the formulas (the proof of which can be easily verified from the theory presented in Chapter 3)

$$p_i = p, \quad p_{ij} = p^2, \quad p_{ijk} = p^3, \quad p_{ijk\ell} = p^4, \quad p_{12345} = p^5,$$

and arrive at the expression (1.9) once again.

KEY TERMS

axiom of countable additivity
certain (sure) event
complement of an event
continuity property of probability
continuous sample space
difference of events
discrete sample space
disjoint (or mutually exclusive) events
event
finite additivity property
finite sample space
impossible event
intersection of events
law of large numbers
monotonically decreasing sequence of events
monotonically increasing sequence of events
nondeterministic experiment
random (or stochastic or nondeterministic) experiment
relative frequency
sample space
statistical regularity
tree diagram
union of events