

## CHAPTER 1

---

# EXPERIMENTS, SAMPLE SPACES, AND EVENTS

---

### 1.1 INTRODUCTION

The consequences of making a decision today often depend on what will happen in the future, or at least on that limited part of the world and of the future that is relevant to the decision. The main purpose of using statistical methods is to help in making better decisions under uncertainty.

Judging from the failures of weather forecasts, to more spectacular prediction failures, such as bankruptcies of large companies and stock market crashes, it would appear that statistical methods do not perform very well. However, with a possible exception of weather forecasting, these examples are, at best, only partially statistical predictions. Moreover, failures tend to be better remembered than successes. Whatever the case, statistical methods are at present, and are likely to remain indefinitely, our best and most reliable prediction tools.

To analyze a given fragment of reality relevant for the specific purpose at hand, one usually needs to collect some *data*. Data may come from past experiences and observations, or may result from some controlled processes, such as laboratory or field experiments. The data are then used to hypothesize about the laws (often called *mechanisms*) that govern the fragment of reality of interest. In our book we are interested in laws expressed in probabilistic terms: They specify directly, or allow

us to compute, the chances of some events to occur. Knowledge of these chances is, in most cases, the best one can get with regard to prediction and decisions.

Probability theory is a domain of pure mathematics and as such, it has its own conceptual structure. To enable a variety of applications (typically comprising of all areas of human endeavor, ranging from biological, medical, social and physical sciences, to engineering, humanities, business, etc.), such structure must be kept on an abstract level. An application of probability to the particular situation analyzed requires a number of initial steps, in which the elements of the real situation are *interpreted* as abstract concepts of probability theory. Such interpretation is often referred to as building a *probabilistic model* of the situation at hand. How well this is done is crucial to the success of application.

One of the main concepts here is that of an *experiment*—a term used in a sense somewhat broader than usual. It means any process, possibly under partial control, that we may observe and whose behavior in the future is not totally determined because it is influenced, at least in part, by chance.

## 1.2 SAMPLE SPACE

In analyzing an experiment, one is primarily interested in its *outcome*—the concept that is not defined (i.e., a *primitive concept*) but has to be specified in every particular application. This specification may be done in different ways, with the only requirements being that (1) outcomes exclude one another and (2) they exhaust the set of all logical possibilities.

### ■ EXAMPLE 1.1

Consider an experiment consisting of two tosses of a regular die. An outcome is most naturally represented by a pair of numbers that turn up on the upper faces of the die so that they form a pair  $(x, y)$ , with  $x, y = 1, 2, \dots, 6$  (see Table 1.1).

**Table 1.1 Outcomes on a Pair of Dice**

		$y$					
		1	2	3	4	5	6
$x$	1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
	2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
	3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
	4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
	5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
	6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

In the case of an experiment of tossing a die three times, the outcomes will be triplets  $(x, y, z)$ , with  $x, y$ , and  $z$  being integers between 1 and 6.

Since the outcome of an experiment is not known in advance, it is important to determine the set of all possible outcomes. This set, called the *sample space*, forms the conceptual framework for all further considerations of probability.

**Definition 1.2.1** The *sample space*, denoted by  $\mathcal{S}$ , is the set of all outcomes of an experiment. The elements of the sample space are called *elementary* outcomes, or *sample points*.  $\square$

### ■ EXAMPLE 1.2

In Example 1.1 the sample space  $\mathcal{S}$  has  $6^2 = 36$  sample points in the case of two tosses, and  $6^3 = 216$  points in the case of three tosses of a die. The first statement can be verified by direct counting of the elements of the sample space. Similar verification of the second claim, although possible in principle, would be cumbersome. In Chapter 3 we will introduce some methods of determining the sizes of sets without actually counting sample points.

### ■ EXAMPLE 1.3

Suppose that the only available information about the numbers, those that turn up on the upper faces of the die, is their sum. In such a case as outcomes we take 11 possible values of the sum so that

$$\mathcal{S} = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

For instance, all outcomes on the diagonal of Table 1.1—(6, 1), (5, 2), (4, 3), (3, 4), (2, 5), and (1, 6)—are represented by the same value 7.

### ■ EXAMPLE 1.4

If we are interested in the number of accidents that occur at a given intersection within a month, the sample space might be taken as the set  $\mathcal{S} = \{0, 1, 2, \dots\}$  consisting of all nonnegative integers. Realistically, there is a practical limit, say 1000, of the monthly numbers of accidents at this particular intersection. Although one may think that it is simpler to take the sample space  $\mathcal{S} = \{0, 1, 2, \dots, 1000\}$ , it turns out that it is often much simpler to take the infinite sample space if the “practical bound” is not very precise.

Since outcomes can be specified in various ways (as illustrated by Examples 1.1 and 1.3), it follows that the same experiment can be described in terms of different sample spaces  $\mathcal{S}$ . The choice of a sample space depends on the goal of description. Moreover, certain sample spaces for the same experiment lead to easier and simpler analysis. The choice of a “better” sample space requires some skill, which is usually gained through experience. The following two examples illustrate this point.

### ■ EXAMPLE 1.5

Let the experiment consist of recording the lifetime of a piece of equipment, say a light bulb. An outcome here is the time until the bulb burns out. An

outcome typically will be represented by a number  $t \geq 0$  ( $t = 0$  if the bulb is not working at the start), and therefore  $\mathcal{S}$  is the nonnegative part of the real axis. In practice,  $t$  is measured with some precision (in hours, days, etc.), so one might instead take  $\mathcal{S} = \{0, 1, 2, \dots\}$ . Which of these choices is better depends on the type of subsequent analysis.

■ **EXAMPLE 1.6**

Two persons enter a cafeteria and sit at a square table, with one chair on each of its sides. Suppose we are interested in the event “they sit at a corner” (as opposed to sitting across from one another). To construct the sample space, we let  $A$  and  $B$  denote the two persons, and then take as  $\mathcal{S}$  the set of outcomes represented by 12 ideograms in Figure 1.1.

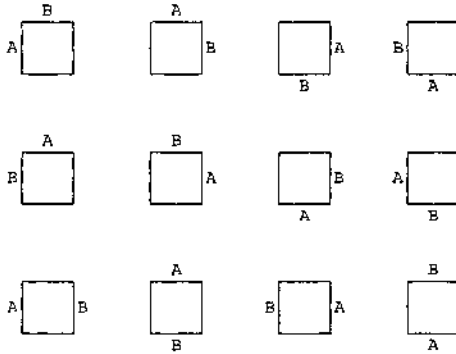


Figure 1.1 Possible seatings of persons  $A$  and  $B$  at a square table

One could argue, however, that such a sample space is unnecessarily large. If we are interested only in the event “they sit at a corner,” then there is no need to label the persons as  $A$  and  $B$ . Accordingly the sample space  $\mathcal{S}$  may be reduced to the set of six outcomes depicted in Figure 1.2.

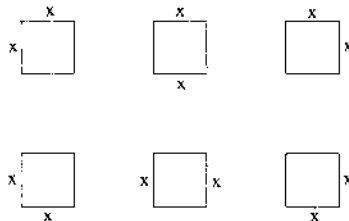
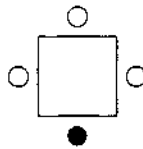


Figure 1.2 Possible seatings of any two persons at a square table

But even this sample space can be simplified. Indeed, one could use the rotational symmetry of the table and argue that once the first person selects

a chair (it does not matter which one), then the sample space consists of just three chairs remaining for the second person (see Figure 1.3).



**Figure 1.3** Possible seatings of one person if the place of the other person is fixed

Sample spaces can be classified according to the number of sample points they contain. *Finite* sample spaces contain finitely many outcomes, and elements of *infinitely countable* sample spaces can be arranged into an infinite sequence; other sample spaces are called *uncountable*.

The next concept to be introduced is that of an *event*. Intuitively, an event is anything about which we can tell whether or not it has occurred, as soon as we know the outcome of the experiment. This leads to the following definition:

**Definition 1.2.2** An *event* is a subset of the sample space  $\mathcal{S}$ . □

#### ■ EXAMPLE 1.7

In Example 1.1, concerning two tosses of a die, an event such as “the sum equals 7” containing six outcomes  $(1, 6)$ ,  $(2, 5)$ ,  $(3, 4)$ ,  $(4, 3)$ ,  $(5, 2)$ , and  $(6, 1)$  is a subset of the sample space  $\mathcal{S}$ . In Example 1.3, the same event consists of one outcome, 7.

When an experiment is performed, we observe its outcome. In the interpretation developed in this chapter, this means that we observe a point chosen randomly from the sample space. If this point belongs to the subset representing the event  $A$ , we say that *the event  $A$  has occurred*.

We will let events be denoted either by letters  $A, B, C, \dots$ , possibly with identifiers, such as  $A_1, B_k, \dots$ , or by more descriptive means, such as  $\{X = 1\}$  and  $\{a < Z < b\}$ , where  $X$  and  $Z$  are some numerical attributes of the sample points (formally: random variables, to be discussed in Chapter 6). Events can also be described through verbal phrases, such as “two heads in a row occur before the third tail” in the experiment of repeated tosses of a coin.

In all cases considered thus far, we assumed that an outcome (a point in the sample space) can be observed. To put it more precisely, all sample spaces  $\mathcal{S}$  considered so far were constructed in such a way that their points were observable. Thus, for any event  $A$ , we were always able to tell whether it occurred or not.

The following examples show experiments and corresponding sample spaces with sample points that are only partially observable:

### ■ EXAMPLE 1.8 Selection

Candidates for a certain job are characterized by their level  $z$  of skills required for the job. The actual value of  $z$  is not observable, though; what we observe is the candidate's score  $x$  on a certain test. Thus, the sample point in  $\mathcal{S}$  is a pair  $s = (z, x)$ , and only one coordinate of  $s$ ,  $x$ , is observable.

The objective might be to find selection thresholds  $z_0$  and  $x_0$ , such that the rule: "accept all candidates whose score  $x$  exceeds  $x_0$ " would lead to maximizing the (unobservable) number of persons accepted whose true level of skill  $z$  exceeds  $z_0$ . Naturally, to find such a solution, one needs to understand statistical relation between observable  $x$  and unobservable  $z$ .

Another example when the points in the sample space are only partially observable concerns studies of incidence of activities about which one may not wish to respond truthfully, or even to respond at all. These are typically studies related to sexual habits or preferences, abortion, law and tax violation, drug use, and so on.

### ■ EXAMPLE 1.9 Randomized Response

Let  $Q$  generally be the activity analyzed, and assume that the researcher is interested in the frequency of persons who ever participated in activity  $Q$  (for simplicity, we will call them  $Q$ -persons). It ought to be stressed that the objective is *not* to identify the  $Q$ -persons, but only to find the proportion of such persons in the population.

The direct question reduced to something like "Are you a  $Q$ -person?" is not likely to be answered truthfully, if at all. It is therefore necessary to make the respondent safe, guaranteeing that their responses will reveal nothing about them as regards  $Q$ . This can be accomplished as follows: The respondent is given a pair of distinguishable dice, for example, one green and one white. She throws them both at the same time, in such a way that the experimenter does not know the results of the toss (e.g., the dice are in a box and only the respondent looks into the box after it is shaken). The instruction is: If the green die shows an odd face (1, 3, or 5), then respond to the question "Are you a  $Q$ -person?" If the green die shows an even face (2, 4, or 6), then respond to the question "Does the white die show an ace?" The scheme of this response is summarized by the flowchart in Figure 1.4.

The interviewer knows the answer "yes" or "no" but does not know whether it is the answer to the question about  $Q$  or the question about the white die. Here a natural sample space consists of points  $s = (i, x, y)$ , where  $x$  and  $y$  are outcomes on green and white die, respectively, while  $i$  is 1 or 0 depending on whether or not the respondent is a  $Q$ -person. We have  $\phi(s) = \phi(i, x, y) =$  "yes" if  $i = 1$  and  $x = 1, 3, \text{ or } 5$  for any  $y$ , or if  $x = 2, 4, 6$ , and  $y = 1$  for any  $i$ . In all other cases  $\phi(s) =$  "no."

One could wonder what is a possible advantage, if any, of not knowing the question asked and observing only the answer. This does not make sense if we need to know the truth about each individual respondent. However, one should remember that we are only after the overall frequency of  $Q$ -persons.

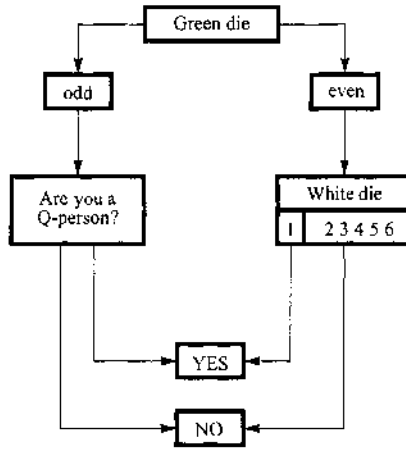


Figure 1.4 Scheme of a randomized response

We are in fact “contaminating” the question by making the respondent answer either a  $Q$ -question or some other auxiliary question. But this is a “controlled contamination”: we know how often (on average) the respondents answer the auxiliary question, and how often their answer is “yes.” Consequently, as we will see in Chapter 12, we can still make an inference about the proportion of  $Q$ -persons from the observed responses.

## PROBLEMS

**1.2.1** List all sample points in sample spaces for the following experiments: (i) We toss a coin. If heads come up, we toss a die. Otherwise, we toss the coin two more times. (ii) A coin is tossed until the total of two tails occurs, but no more than four times (i.e., a coin is tossed until the second tail or fourth toss, whichever comes first).

**1.2.2** Alice, Bob, Carl, and Diana enter the elevator on the first floor of a four-story building. Each of them leaves the elevator on either the second, third, or fourth floor. (i) Find a simple way of describing the sample space (do not list all sample points). (ii) List all sample points such that Carl and Diana leave the elevator on the third floor. (iii) List all sample points if Carl and Diana leave the elevator at the same floor.

**1.2.3** An urn contains five chips, labeled  $1, \dots, 5$ . Three chips are drawn. List all outcomes included in the event “the second largest number drawn was 3.”

**1.2.4** In a game of craps, the player rolls a pair of dice. If he gets a total of 7 or 11, he wins at once; if the total is 2, 3, or 12, he loses at once. Otherwise, the sum, say  $x$ , is his “point,” and he keeps rolling dice until either he rolls another  $x$  (in which case he wins) or he rolls a 7 in which case he loses. Describe the event “the player wins with a point of 5.”

**1.2.5** The experiment consists of placing six balls in three boxes. List all outcomes in the sample space if: (i) The balls are indistinguishable, but the boxes are distinguishable. (*Hint:* There are 28 different placements.) (ii) Neither the balls nor the boxes are distinguishable. (iii) Two balls are white and four are red; the boxes are distinguishable.

**1.2.6** John and Mary plan to meet each other. Each of them is to arrive at the meeting place at some time between 5 p.m. and 6 p.m. John is to wait 20 minutes (or until 6 p.m., whichever comes first), and then leave if Mary does not show up. Mary will wait only 5 minutes (or until 6 p.m., whichever comes first), and then leave if John does not show up. Letting  $x$  and  $y$  denote the arrival times of John and Mary, determine the sample space and describe events (i)–(viii) by drawing pictures, or by appropriate inequalities for  $x$  and  $y$ . If you think that the description is impossible, say so. (i) John arrives before Mary does. (ii) John and Mary meet. (iii) Either Mary comes first or they do not meet. (iv) Mary comes first but they do not meet. (v) John comes very late. (vi) They arrive less than 15 minutes apart and they do not meet. (vii) Mary arrives at 5:15 p.m. and meets John, who is already there. (viii) They almost miss one another.

Problems 1.2.7–1.2.8 concern the possibility of expressing some events, depending on the choice of the sample space.

**1.2.7** Let  $\mathcal{E}$  be the experiment consisting of tossing a coin three times, with H and T standing for heads and tails, respectively.

(i) The following set of outcomes is an incomplete list of the points of the sample space  $\mathcal{S}$  of the experiment  $\mathcal{E}$ : {HHH, HTT, TTT, HHT, TTH, HTH, THH}. Find the missing outcome.

(ii) An alternative sample space  $\mathcal{S}'$  for the same experiment  $\mathcal{E}$  consists of the following four outcomes: no heads (0), one head (1), two heads (2), and three heads (3). Which of the following events can be described as subsets of  $\mathcal{S}$  but not as subsets of  $\mathcal{S}' = \{0, 1, 2, 3\}$ ?

$A_1 =$  More than two heads.

$A_2 =$  Head on the second toss.

$A_3 =$  More tails than heads.

$A_4 =$  At least one tail, with head on the last toss.

$A_5 =$  At least two faces the same.

$A_6 =$  Head and tail alternate.

(iii) Still another sample space  $\mathcal{S}''$  for the experiment  $\mathcal{E}$  consists of the four outcomes (0, 0), (0, 1), (1, 0), and (1, 1). The first coordinate is 1 if the first two tosses show the same face and 0 otherwise; the second coordinate is 1 if the last two tosses show the same face, and 0 otherwise. For instance, if we observe HHT, the outcome is (1, 0). List the outcomes of  $\mathcal{S}$  that belong to the event  $A = \{(1, 1), (0, 1)\}$  of  $\mathcal{S}''$ . (iv) Which of the following events can be represented as subsets of  $\mathcal{S}$ , but cannot be represented as subsets of  $\mathcal{S}''$ ?

$B_1 =$  First and third tosses show the same face.

- $B_2 =$  Heads on all tosses.  
 $B_3 =$  All faces the same.  
 $B_4 =$  Each face appears at least once.  
 $B_5 =$  More heads than tails.

**1.2.8** Let  $\mathcal{E}$  be the experiment consisting of tossing a die twice. Let  $\mathcal{S}$  be the sample space with sample points  $(i, j)$ ,  $i, j = 1, 2, \dots, 6$ , with  $i$  and  $j$  being the numbers of dots that appear in the first and second toss, respectively.

(i) Let  $\mathcal{S}'$  be the sample space for the experiment  $\mathcal{E}$  consisting of all possible sums  $i + j$  so that  $\mathcal{S}' = \{2, 3, \dots, 12\}$ . Which of the following events can be defined as subsets of  $\mathcal{S}$  but not of  $\mathcal{S}'$ ?

- $A_1 =$  One face odd, the other even.  
 $A_2 =$  Both faces even.  
 $A_3 =$  Faces different.  
 $A_4 =$  Result on the first toss less than the result on the second.  
 $A_5 =$  Product greater than 10.  
 $A_6 =$  Product greater than 30.

(ii) Let  $\mathcal{S}''$  be the sample space for the experiment  $\mathcal{E}$  consisting of all possible absolute values of the difference  $|i - j|$  so that  $\mathcal{S}'' = \{0, 1, 2, 3, 4, 5\}$ . Which of the following events can be defined as subsets of  $\mathcal{S}$  but not of  $\mathcal{S}''$ ?

- $B_1 =$  One face shows twice as many dots as the other,  
 $B_2 =$  Faces the same,  
 $B_3 =$  One face shows six times as many dots as the other,  
 $B_4 =$  One face odd, the other even,  
 $B_5 =$  The ratio of the numbers of dots on the faces is different from 1.

**1.2.9** Referring to Example 1.9, suppose that we modify it as follows: The respondent tosses a green die (with the outcome unknown to the interviewer). If the outcome is odd, he responds to the Q-question; otherwise, he responds to the question "Were you born in April?" Again, the interviewer observes only the answer "yes" or "no." Apart from the obvious difference in frequency of the answer "yes" to the auxiliary question (on the average one in 12 instead of one in 6), are there any essential differences between this scheme and the scheme in Example 1.9? Explain your answer.

### 1.3 ALGEBRA OF EVENTS

Next we introduce some concepts that will allow us to form composite events out of simpler ones. We begin with the relations of *inclusion* and *equality*.

**Definition 1.3.1** The event  $A$  is *contained* in the event  $B$ , or  $B$  *contains*  $A$ , if every sample point of  $A$  is also a sample point of  $B$ . Whenever this is true, we will write  $A \subset B$ , or equivalently,  $B \supset A$ . □

An alternative terminology here is that  $A$  *implies* (or *entails*)  $B$ .

**Definition 1.3.2** Two events  $A$  and  $B$  are said to be *equal*,  $A = B$ , if  $A \subset B$  and  $B \subset A$ .  $\square$

It follows that two events are equal if they consist of exactly the same sample points.

■ **EXAMPLE 1.10**

Consider two tosses of a coin, and the corresponding sample space  $\mathcal{S}$  consisting of four outcomes: HH, HT, TH, and TT. The event  $A =$  “heads in the first toss”  $= \{HH, HT\}$  is contained in the event  $B =$  “at least one head”  $= \{HH, HT, TH\}$ . The events “the results alternate” and “at least one head and one tail” imply one another, and hence are equal.

**Definition 1.3.3** The set containing no elements is called the *empty set* and is denoted by  $\emptyset$ . The event corresponding to  $\emptyset$  is called a *null (impossible) event*.  $\square$

■ **EXAMPLE 1.11** \*

<sup>1</sup> The reader may wonder whether it is correct to use the definite article in the definition above and speak of “*the empty set*,” since it would appear that there may be many different empty sets. For instance, the set of all kings of the United States and the set of all real numbers  $x$  such that  $x^2 + 1 = 0$  are both empty, but one consists of people and the other of numbers, so they cannot be equal. This is not so, however, as is shown by the following formal argument (to appreciate this argument, one needs some training in logic). Suppose that  $\emptyset_1$  and  $\emptyset_2$  are two empty sets. To prove that they are equal, one needs to prove that  $\emptyset_1 \subset \emptyset_2$  and  $\emptyset_2 \subset \emptyset_1$ . Formally, the first inclusion is the implication: “if  $s$  belongs to  $\emptyset_1$ , then  $s$  belongs to  $\emptyset_2$ .” This implication is true, because its premise is false: there is no  $s$  that belongs to  $\emptyset_1$ . The same holds for the second implication, so  $\emptyset_1 = \emptyset_2$ .

We now give the definitions of three principal operations on events: *complementation, union, and intersection*.

**Definition 1.3.4** The set that contains all sample points that are not in the event  $A$  will be called the *complement* of  $A$  and denoted  $A^c$ , to be read also as “not  $A$ .”  $\square$

**Definition 1.3.5** The set that contains all sample points belonging either to  $A$  or to  $B$  (so possibly to both of them) is called the *union* of  $A$  and  $B$  and denoted  $A \cup B$ , to be read as “ $A$  or  $B$ .”  $\square$

**Definition 1.3.6** The set that contains all sample points belonging to both  $A$  and  $B$  is called the *intersection* of  $A$  and  $B$ , and denoted  $A \cap B$ , to be read as “ $A$  and  $B$ .”  $\square$

An alternative notation for a complement is  $A'$  or  $\overline{A}$ , whereas in the case of an intersection one often writes  $AB$  instead of  $A \cap B$ .

<sup>1</sup> Asterisks denote more advanced material, as explained in the Preface.

The operations above have the following interpretations in terms of occurrences of events:

1. Event  $A^c$  occurs if event  $A$  does not occur.
2. Event  $A \cup B$  occurs when either  $A$  or  $B$  or both events occur.
3. Event  $A \cap B$  occurs when both  $A$  and  $B$  occur.

■ **EXAMPLE 1.12**

Consider the experiment of tossing a coin three times, with the sample space consisting of outcomes described as HHH, HHT, and so on. Let  $A$  be the event “heads and tails alternate,” and let  $B$  be “heads on the last toss.” The event  $A^c$  occurs if either heads or tails occur at least twice in a row so that  $A^c = \{HHH, HHT, THH, HTT, TTT, TTH\}$  while  $B^c$  is “tails on the last toss,” hence  $B^c = \{HHT, THT, HTT, TTT\}$ . The union  $A \cup B$  is the event “either the results alternate or it is heads on the last toss,” meaning  $A \cup B = \{HTH, THT, HHH, THH, TTH\}$ . Observe that while  $A$  has two outcomes and  $B$  has four outcomes, their union has only five outcomes, since the outcome HTH appears in both events. This common part is the intersection  $A \cap B$ .

Some formulas can be simplified by introducing the operation of the *difference* of two events.

**Definition 1.3.7** The *difference*  $A \setminus B$  of events  $A$  and  $B$  contains all sample points that belong to  $A$  but not to  $B$

$$A \setminus B = A \cap B^c.$$

The *symmetric difference*,  $A \div B$ , contains sample points that belong to  $A$  or to  $B$ , but not to both of them:

$$A \div B = (A \cap B^c) \cup (A^c \cap B) = (A \cup B) \setminus (A \cap B). \quad \square$$

■ **EXAMPLE 1.13**

In Example 1.12, the difference  $B^c \setminus A$  is described as “at least two identical outcomes in a row and tails on the last toss,” which means the event  $\{HHT, HTT, TTT\}$ .

Next we have the following important concept:

**Definition 1.3.8** If  $A \cap B = \emptyset$ , then the events  $A$  and  $B$  are called *disjoint*, or *mutually exclusive*. □

■ **EXAMPLE 1.14**

Based on Example 1.12 we know that the following two events are disjoint:  $C$  = “more heads than tails” and the intersection  $A \cap B^c$  = “the results alternate, ending with tails.”

Example 1.14 shows that to determine whether or not events are disjoint, it is not necessary to list the outcomes in both events and check whether there exist common outcomes. Apart from the fact that such listing is not feasible when sample spaces are large, it is often simpler to employ some logical reasoning, for instance, that one of the events is contained in the complement of the other (i.e., if one of them occurs, the other does not). In the case above, if the results alternate and end with tails, then the outcome must be THT. Since there are more tails than heads,  $C$  does not occur.

The definitions of union and intersection can be extended to the case of a finite and even infinite number of events (discussed in the Section 1.4). Thus

$$A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^n A_i \quad (1.1)$$

is the event that contains the sample points belonging to  $A_1$  or  $A_2$  or ... or  $A_n$ . Consequently, (1.1) is the event “at least one  $A_i$  occurs.” Similarly

$$A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^n A_i \quad (1.2)$$

is the event that contains the sample points belonging to  $A_1$  and  $A_2$  and ... and  $A_n$ . Consequently, the event (1.2) is “all  $A_i$ 's occur.”

■ **EXAMPLE 1.15**

Suppose that  $n$  shots are fired at a target, and let  $A_i$ ,  $i = 1, 2, \dots, n$  denote the event “the target is hit on the  $i$ th shot.” Then the union  $A_1 \cup \cdots \cup A_n$  is the event “the target is hit” (at least once). Its complement  $(A_1 \cup \cdots \cup A_n)^c$  is the event “the target is missed” (on every shot), which is the same as the intersection  $A_1^c \cap \cdots \cap A_n^c$ .

A perceptive reader may note that the unions  $A_1 \cup \cdots \cup A_n$  and intersections  $A_1 \cap \cdots \cap A_n$  do not require an extension of the definition of union and intersection for two events. Indeed, we could consider unions such as

$$A_1 \cup (A_2 \cup (\cdots (A_{n-2} \cup (A_{n-1} \cup A_n)) \cdots)),$$

where the union of only two events is formed in each set of parentheses. The property of associativity (below) shows that parentheses can be omitted so that the expression  $A_1 \cup \cdots \cup A_n$  is unambiguous. The same argument applies to intersections.

The operations on events defined in this section obey some laws. The most important ones are listed below.

**Idempotence:**

$$A \cup A = A, \quad A \cap A = A.$$

**Double Complementation:**

$$(A^c)^c = A.$$

**Absorption:**

$$A \cup B = B \text{ iff } A \cap B = A \text{ iff } A \subset B. \quad (1.3)$$

In particular,

$$A \cup \emptyset = A, \quad A \cup S = S, \quad A \cap \emptyset = \emptyset, \quad A \cap S = A,$$

which in view of (1.3) means that  $\emptyset \subset A \subset S$ .

**Commutativity:**

$$A \cup B = B \cup A, \quad A \cap B = B \cap A.$$

**Associativity:**

$$\begin{aligned} A \cup (B \cup C) &= (A \cup B) \cup C, \\ A \cap (B \cap C) &= (A \cap B) \cap C. \end{aligned}$$

**Distributivity:**

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C), \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C). \end{aligned}$$

**De Morgan's Laws:**

$$\begin{aligned} (A_1 \cup \dots \cup A_n)^c &= A_1^c \cap \dots \cap A_n^c, \\ (A_1 \cap \dots \cap A_n)^c &= A_1^c \cup \dots \cup A_n^c. \end{aligned} \quad (1.4)$$

When studying mutual relations between composite events in the same sample space, it is often helpful to use Venn diagrams, where the sample space  $S$  is represented by a rectangle, while its subsets represent events (see Figure 1.5).

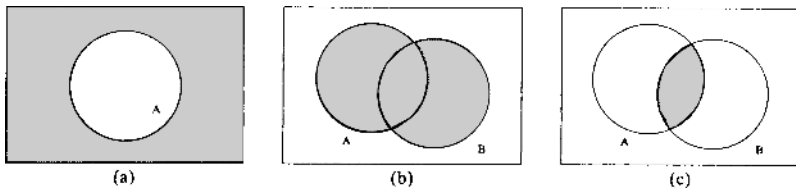


Figure 1.5 Complement, union and intersection

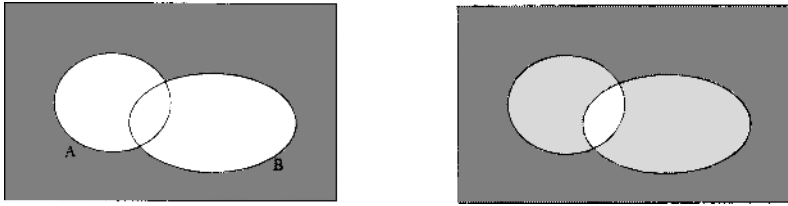


Figure 1.6 The first De Morgan's law

The complement of event  $A$  is represented in Figure 1.5(a), the union and intersection of the events  $A$  and  $B$  are represented in Figure 1.5(b) and (c), respectively.

Venn diagrams can also be used to check the validity of formulas. For example, consider the first De Morgan's law (1.4) for the case of two events:

$$(A \cup B)^c = A^c \cap B^c. \quad (1.5)$$

Venn diagrams made separately for the left-hand side and the right-hand side of (1.5) (see Figure 1.6) indicate that both regions are the same. Although a picture does not constitute a proof, it may provide convincing evidence that the statement is true, and sometimes may even suggest a method of proving the statement.

## PROBLEMS

For the problems below, remember that a statement (expressed as a sentence or formula) is true if it is true under *all* circumstances, and it is false if there is at least one case where it does not hold.

**1.3.1** Answer true or false. Justify your answer.

- (i) If  $A$  and  $B$  are distinct events (i.e.,  $A \neq B$ ) such that  $A$  and  $B^c$  are disjoint, then  $A^c$  and  $B$  are also disjoint.
- (ii) If  $A$  and  $B$  are disjoint, then  $A^c$  and  $B^c$  are also disjoint.
- (iii) If  $A$  and  $B$  are disjoint, and also  $B$  and  $C$  are disjoint, then  $A$  and  $C$  are disjoint.
- (iv) If  $A$  and  $B$  are both contained in  $C$ , then  $C^c \subset A^c \cap B^c$ .
- (v) If  $A$  is contained in  $B$ ,  $C$  is contained in  $D$ , and  $B$  is disjoint from  $D$ , then  $A$  is disjoint from  $C$ .
- (vi) If  $A \cup B^c = B^c$ , then  $B \subset A^c$ .

**1.3.2** In the statements below  $A, B, C$  and  $D$ , are events. Find those statements or formulas that are true.

- (i) If  $A \cap B = A \cap C$  then  $B = C$ .
- (ii)  $A \cup (A \cap B) \cup (A \cap B^c) = A$ .
- (iii)  $A \cup (A \cap B) \cup (A \cap B^c) = B$ .
- (iv) If  $A \setminus B = C$ , then  $A = B \cup C$ .
- (v)  $(A \cup B) \cap (C \cup D) = (A \cap C) \cup (A \cap D) \cup (B \cap C) \cup (B \cap D)$ .
- (vi)  $(A \cap B) \cup (C \cap D) = (A \cup C) \cap (A \cup D) \cap (B \cup C) \cap (B \cup D)$ .

(vii)  $(A^c \cup B^c \cup C^c)^c = A^c \cap B^c \cap C^c$ .

(viii) If  $A \subset B$ , and  $B \cap C = \emptyset$ , then  $C^c \cap A \cap B^c = \emptyset$ .

(ix) If  $A \cap B$ ,  $A \cap C$  and  $B \cap C$  are not empty, then  $A \cap B \cap C$  is not empty.

(x) Show that  $(A \div B) \div C = A \div (B \div C)$ .

**1.3.3** Find  $X$  if: (i)  $A \div X = \emptyset$ . (ii)  $A \div X = A$ . (iii)  $A \div X = S$ . (iv)  $A \div X = B$ .

**1.3.4** In a group of 1000 students of a certain college, 60 take French, 417 take calculus, and 509 take statistics. Moreover, 20 take French and calculus, 17 take French and statistics, and 147 take statistics and calculus. However, 196 students do not take any of these three subjects. Determine the number of students who take French, calculus, and statistics.

**1.3.5** Let  $A$ ,  $B$ , and  $C$  be three events. Match, where possible, events  $D_1$  through  $D_{10}$  with events  $E_1$  through  $E_{11}$ . Matching means that the events are exactly the same; that is, if one occurs, so must the other and conversely (see the Definition 1.3.2). (*Hint*: Draw a Venn diagram for each event  $D_1, \dots, D_{10}$ , do the same for events  $E_1, \dots, E_{11}$ , and then compare the diagrams.)

Among events  $A$ ,  $B$ ,  $C$ :

$D_1$  = two or more occur.

$D_2$  = exactly one occurs.

$D_3$  = only  $A$  occurs.

$D_4$  = all occur.

$D_5$  = none occurs.

$D_6$  = at most one occurs.

$D_7$  = at least one occurs.

$D_8$  = exactly two occur.

$D_9$  = no more than two occur.

$D_{10}$  =  $B$  occurs.

$$E_1 = A \cup B \cup C.$$

$$E_2 = (A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C).$$

$$E_3 = (A \cap B)^c \cap (A \cap C)^c \cap (B \cap C)^c,$$

$$E_4 = (A \cup B \cup C)^c.$$

$$E_5 = A^c \cap B^c \cap C^c.$$

$$E_6 = A \cap B \cap C.$$

$$E_7 = B.$$

$$E_8 = A \cap B^c \cap C^c.$$

$$E_9 = (A \cap B \cap C^c) \cup (A \cap B^c \cap C) \cup (A^c \cap B \cap C).$$

$$E_{10} = (A \cap B \cap C)^c.$$

$$E_{11} = (A \cap B) \cup (A \cap C) \cup (B \cap C).$$

**1.3.6** A standard deck of cards is dealt among players  $N$ ,  $S$ ,  $E$ , and  $W$ . Let  $N_k$ ,  $k = 1, 2, 3, 4$  be the event " $N$  has at least  $k$  aces," and let  $S_k$ ,  $E_k$  and  $W_k$  be defined similarly. For each of the events below, determine the number of aces that  $N$  has.

(i)  $N_1 \cap S_1 \cap E_1 \cap W_1$ . (ii)  $E_2 \cap (W_2 \cup S_2)$ . (iii)  $N_3 \setminus N_4$ . (iv)  $S_3 \cap W_1$ . (v)  $S_1^c \cap W_1^c \cap E_1^c$ . (vi)  $N_2 \cap E_2$ .

**1.3.7** Five burglars,  $A, B, C, D,$  and  $E,$  divide the loot, consisting of 5 identical gold bars and 4 identical diamonds. Let  $A_{jk}$  be the event that  $A$  got *at least*  $j$  gold bars and *at most*  $k$  diamonds. Let  $B_{jk}, C_{jk}$  denote analogous events for burglars  $B, C$  (e.g.,  $B_{21}$  is the event that  $B$  got 2, 3, 4, or 5 gold bars and 0 or 1 diamond). Determine the number  $x$  of gold bars and the number  $y$  of diamonds received by  $E$  if the following events occur (if determination of  $x$  and/or  $y$  is impossible, give the range of values): **(i)**  $(A_{20} \cup B_{20} \cup C_{20}) \cap D_{30}$ . **(ii)**  $E_{12}^c$ . **(iii)**  $A_{23} \cap B_{13} \cap C_{13} \cap D_{13}$ . **(iv)**  $A_{23} \cup B_{13} \cup C_{13} \cup D_{13}$ .

**1.3.8** Let  $A^{nc}$  be defined inductively by  $A^{0c} = A, A^{(n+1)c} = (A^{nc})^c$ . Find  $A^{mc} \cap A^{nc}$  and  $A^{mc} \cup A^{nc}$  for  $m, n > 0$ .

### 1.4 INFINITE OPERATIONS ON EVENTS

As already mentioned, the operations of union and intersection can be extended to infinitely many events. Let  $A_1, A_2, \dots$  be an infinite sequence of events. Then

$$A_1 \cup A_2 \cup \dots = \bigcup_{i=1}^{\infty} A_i \quad \text{and} \quad A_1 \cap A_2 \cap \dots = \bigcap_{i=1}^{\infty} A_i$$

are events “at least one  $A_i$  occurs” and “all  $A_i$ ’s occur.”

If at least one event  $A_i$  occurs, then there is one that occurs first. This remark leads to the following useful decomposition of a union of events into a union of *disjoint* events:

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_1^c \cap A_2) \cup (A_1^c \cap A_2^c \cap A_3) \cup \dots, \tag{1.6}$$

where  $A_1^c \cap \dots \cap A_{k-1}^c \cap A_k$  is the event “ $A_k$  is the first event in the sequence that occurs.”

For an infinite sequence  $A_1, A_2, \dots$  one can define two events:

$$\limsup A_n = \bigcap_{k=1}^{\infty} \bigcup_{i=k}^{\infty} A_i \tag{1.7}$$

and

$$\liminf A_n = \bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} A_i, \tag{1.8}$$

these being, respectively, the event that “infinitely many  $A_i$ ’s occur” and the event that “all except finitely many  $A_i$ ’s occur.” Here the inner union in the event (1.7) is the event “at least one event  $A_i$  with  $i \geq k$  will occur”; call this event  $B_k$ . The intersection over  $k$  means that the event  $B_k$  occurs for every  $k$ . No matter how large  $k$  we take, there will be at least one event  $A_i$  with  $i \geq k$  that will occur. But this is possible only if infinitely many  $A_i$ ’s occur.

For the event  $\liminf A_n$  the argument is similar. The intersection  $A_k \cap A_{k+1} \cap \dots = C_k$  occurs if all events  $A_i$  with  $i \geq k$  occur. The union  $C_1 \cup C_2 \cup \dots$  means

that at least one of the events  $C_k$  will occur, and that means that all  $A_i$  will occur, except possibly finitely many.

If all events (except possibly finitely many) occur, then infinitely many of them must occur, so that  $\limsup A_n \supset \liminf A_n$ . If  $\limsup A_n \subset \liminf A_n$ , then (see the definition of equality of events) we say that the sequence  $\{A_n\}$  converges, and  $\limsup A_n = \liminf A_n$ .

The most important class of convergent sequences of events consists of *monotone* sequences, when  $A_1 \subset A_2 \subset \dots$  (increasing sequence) or when  $A_1 \supset A_2 \supset \dots$  (decreasing sequence). We have the following theorem:

**Theorem 1.4.1** *If the sequence  $A_1, A_2, \dots$  is increasing, then*

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

and in case of a decreasing sequence, we have

$$\lim A_n = \bigcap_{n=1}^{\infty} A_n.$$

*Proof.* If the sequence is increasing, then the inner union  $(\bigcup_{i=1}^{\infty} A_i)$  in  $\limsup A_n$  remains the same independently of  $k$  so that  $\limsup A_n = \bigcup_{i=1}^{\infty} A_i$ . On the other hand, the inner intersection in  $\liminf A_n$  equals  $A_k$  so that  $\liminf A_n = \bigcup_{k=1}^{\infty} A_k$ , which is the same as  $\limsup A_n$ , as was to be shown. A similar argument holds for decreasing sequences.  $\square$

The following two examples illustrate the concept of convergence of events.

■ **EXAMPLE 1.16**

Let  $B(r)$  and  $C(r)$  be the sets of points on the plane  $(x, y)$  satisfying the conditions  $x^2 + y^2 < r^2$  and  $x^2 + y^2 \leq r^2$ , respectively. If  $A_n = B(1 + 1/n)$ , then  $\{A_n\}$  is a decreasing sequence, and therefore  $\lim A_n = \bigcap_{n=1}^{\infty} B(1 + 1/n)$ . Since  $x^2 + y^2 < (1 + 1/n)^2$  for all  $n$  if and only if  $x^2 + y^2 \leq 1$ , we have  $\bigcap_{n=1}^{\infty} B(1 + 1/n) = C(1)$ . On the other hand, if  $A_n = C(1 - 1/n)$ , then  $\{A_n\}$  is an increasing sequence, and  $\lim A_n = \bigcup_{n=1}^{\infty} C(1 - 1/n) = B(1)$ . We leave a justification of the last equality to the reader.

■ **EXAMPLE 1.17**

Let  $A_n = B(1 + 1/n)$  for  $n$  odd and  $A_n = B(1/3 - 1/2n)$  for  $n$  even. The sequence  $\{A_n\}$  is now  $B(2), B(1/12), B(4/3), B(5/24), \dots$ , so it is not monotone. We have here  $\limsup A_n = C(1)$ , since every point  $(x, y)$  with  $x^2 + y^2 \leq 1$  belongs to infinitely many  $A_n$ . On the other hand,  $\liminf A_n = B(1/3)$ . For  $x^2 + y^2 < 1/9$ , we have  $x^2 + y^2 < (1/3 - 1/2n)^2$  if  $n$  is large enough (and also  $x^2 + y^2 < 1 + 1/n$  for all  $n$ ). However, if  $x^2 + y^2 \geq 1/3$ , then  $(x, y)$  does not belong to any  $A_n$  with even  $n$ . Thus  $\limsup A_n \neq \liminf A_n$ , and the sequence  $\{A_n\}$  does not converge.

Infinite operations on events play a very important role in the development of the theory, especially in determining limiting probabilities.

The definitions below will prepare the ground for the considerations in the following chapters. In Chapter 2 we will introduce probability as a number assigned to an event. Formally, we will be considering numerical functions defined on events, that is, on subsets of the sample space  $\mathcal{S}$ . As long as  $\mathcal{S}$  is finite or countably infinite, we can take the class of all subsets of  $\mathcal{S}$  as the domain of definition of probability. In case of infinite but *not* countable  $\mathcal{S}$  (e.g., where  $\mathcal{S}$  is an interval, the real line, or a plane) it may not be possible to define probability on the class of *all* subsets of  $\mathcal{S}$ . Although the explanation lies beyond the scope of this book, we will show how the difficulties can be avoided by suitable restriction of the class of subsets of  $\mathcal{S}$  that are taken as events. We begin with the concept of *closure* under some operation.

**Definition 1.4.1** We say that the class  $\mathcal{A}$  of subsets of  $\mathcal{S}$  is *closed* under a given operation if the sets resulting from performing this operation on elements of  $\mathcal{A}$  are also elements of  $\mathcal{A}$ .  $\square$

Complementation  $A^c$ , finite union  $A_1 \cup \cdots \cup A_n$ , infinite union  $A_1 \cup A_2 \cup \cdots$ , limits of sequences  $\lim A_n$ , are few examples of such operations.

■ **EXAMPLE 1.18**

Let  $\mathcal{S} = \{0, 1, 2, \dots\}$ , and let  $\mathcal{A}$  consist of all subsets of  $\mathcal{S}$  that are finite.  $\mathcal{A}$  is closed under finite unions and all intersections, finite or not. Indeed, if  $A_1, \dots, A_n$  are finite sets, then  $A = A_1 \cup \cdots \cup A_n$  is also finite. Similarly, if  $A_1, A_2, \dots$  are finite, then  $\bigcap_i A_i \subset A_1$ , and hence  $\bigcap_i A_i$  is also finite. However,  $\mathcal{A}$  is *not closed* under complementation: if  $A$  is finite ( $A \in \mathcal{A}$ ), then  $A^c$  is not finite, and hence  $A^c \notin \mathcal{A}$ . On the other hand, if  $\mathcal{A}$  is the class of all subsets of  $\mathcal{S}$  that contain some fixed element, say 0, then  $\mathcal{A}$  is closed under all intersections and unions, but it is not closed under complementation.

■ **EXAMPLE 1.19**

Let  $\mathcal{S}$  be the real line, and let  $\mathcal{A}$  be the class of all intervals closed on the right and open on the left—meaning intervals of the form  $(a, b] = \{x : a < x \leq b\}$ . Assume that we allow here  $b \leq a$ , in which case  $(a, b]$  is empty. Then  $\mathcal{A}$  is closed under the operation of intersection with  $(a, b] \cap (c, d] = (\max(a, c), \min(b, d)]$ .

The following three concepts play a central role in the construction of the probability theory:

**Definition 1.4.2** A nonempty class  $\mathcal{A}$  of subsets of  $\mathcal{S}$  that is closed under complementation and all finite operations (i.e., finite union, finite intersection) is called a *field*. If  $\mathcal{A}$  is closed under complementation and all countable operations, it is called

a  $\sigma$ -field. Finally, if  $\mathcal{A}$  is closed under monotone passage to the limit,<sup>2</sup> it is called a *monotone class*.  $\square$

Let us observe that Definition 1.4.2 can be formulated in a more efficient way. For  $\mathcal{A}$  to be a field, it suffices to require that if  $A, B \in \mathcal{A}$  then  $A^c \in \mathcal{A}$  and  $A \cap B \in \mathcal{A}$  (or  $A^c \in \mathcal{A}$  and  $A \cup B \in \mathcal{A}$ ). Any of these two conditions implies (by induction and De Morgan's laws) the closure of  $\mathcal{A}$  under all finite operations. Consequently, for  $\mathcal{A}$  to be a  $\sigma$ -field, it suffices to require that whenever  $A_1, A_2, \dots \in \mathcal{A}$ , then  $A_i^c \in \mathcal{A}$  and  $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$  (or  $A_i^c \in \mathcal{A}$  and  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ ); this follows again from De Morgan's laws.<sup>3</sup>

It is important to realize that closure under countable operations is stronger than closure under any finite operations. This means that there exist classes of sets that are fields but not  $\sigma$ -fields. This is illustrated by the following example:

■ **EXAMPLE 1.20**

Let  $S = \{1, 2, 3, \dots\}$ , and let  $\mathcal{A}$  be the class of all subsets  $A$  of  $S$  such that either  $A$  or  $A^c$  is finite. Then  $\mathcal{A}$  is a field but not a  $\sigma$ -field. First, if  $A \in \mathcal{A}$ , then  $A^c \in \mathcal{A}$  because the definition of  $\mathcal{A}$  is symmetric with respect to complementation. Next, if  $A$  and  $B$  are both in  $\mathcal{A}$ , so is their union. If  $A$  and  $B$  are both finite, then  $A \cup B$  is finite and hence belongs to  $\mathcal{A}$ . On the other hand, if either  $A^c$  or  $B^c$  (or both) are finite, then  $(A \cup B)^c = A^c \cap B^c$  is also finite because it is contained in  $A^c$  and also in  $B^c$ .

Thus  $\mathcal{A}$  is a field. However,  $\mathcal{A}$  is not a  $\sigma$ -field. Let  $A_n$  be the set consisting only of the element  $n$  (i.e.,  $A_n = \{n\}$ ). Clearly,  $A_n \in \mathcal{A}$ . Take now  $\bigcup_{n=1}^{\infty} A_{2n} = \{2, 4, 6, \dots\}$ . This is a countable union of sets in  $\mathcal{A}$  that is not in  $\mathcal{A}$  (since the set of all even numbers is not finite, nor does it have a finite complement).

Typically it is easy to determine that a class of sets is a field, while direct verification that it is a  $\sigma$ -field can be difficult. On the other hand, it is occasionally easy to verify that a class of sets is a monotone class. Thus the following theorem is sometimes useful:

**Theorem 1.4.2** *A  $\sigma$ -field is a monotone class. Conversely, a field that is a monotone class is a  $\sigma$ -field.*

*Proof.* To prove this theorem, assume first that  $\mathcal{A}$  is a  $\sigma$ -field, and let  $A_1, A_2, \dots$  be a monotone sequence of elements of  $\mathcal{A}$ . If  $A_1 \subset A_2 \subset \dots$ , then  $\lim A_n = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ , whereas if  $A_1 \supset A_2 \supset \dots$ , then  $\lim A_n = \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$ . So  $\mathcal{A}$  is a monotone class. On the other hand, let  $\mathcal{A}$  be a monotone class and a field, and let  $A_1, A_2, \dots$  be an arbitrary sequence of elements of  $\mathcal{A}$ . Put  $B_n = A_1 \cup \dots \cup A_n$ . Then since  $\mathcal{A}$  is a field, and also  $B_1 \subset B_2 \subset \dots$ ,  $B_n \in \mathcal{A}$  for every  $n$ . Further, since  $\mathcal{A}$

<sup>2</sup>In view of the fact proved earlier that all monotone sequences converge, this condition means that (a) if  $A_1 \subset A_2 \subset \dots$  is an increasing sequence of sets in  $\mathcal{A}$ , then  $\bigcup_i A_i \in \mathcal{A}$ , and (b) if  $A_1 \supset A_2 \supset \dots$  is a decreasing sequence of sets in  $\mathcal{A}$ , then  $\bigcap_i A_i \in \mathcal{A}$ .

<sup>3</sup>For various relations among classes of sets defined through closure properties under operations, for example, see Chow and Teicher (1997) and Chung (2001).

is a monotone class,  $\lim B_n \in \mathcal{A}$ . However,  $\lim B_n = \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$ , so  $\mathcal{A}$  is a  $\sigma$ -field, as asserted.  $\square$

The last in this series of concepts is that of the minimal field (or  $\sigma$ -field, or monotone class) containing a given set or collection of sets. We begin with some examples.

#### ■ EXAMPLE 1.21

Let  $S$  be any set. On one extreme, the class consisting of two sets,  $\emptyset$  and  $S$ , is closed under any operation so that  $\mathcal{A} = \{\emptyset, S\}$  is a field, a  $\sigma$ -field, and a monotone class. On the other extreme, the class of *all* subsets of  $S$  is also closed under any operations, finite or not, and hence is a field, a  $\sigma$ -field, and a monotone class. These two classes of subsets of  $S$  form the smallest and the largest fields ( $\sigma$ -field, monotone class).

For any event  $A$  it is easy to check that the class  $\mathcal{A}$ , consisting of the four events  $\{\emptyset, A, A^c, S\}$ , is closed under any operations: unions, intersections, and complements of members of  $\mathcal{A}$  are again members of  $\mathcal{A}$ . This class is an example of a field ( $\sigma$ -field, monotone class) that contains the events  $A$  and  $A^c$ , and it is the smallest such field ( $\sigma$ -field, monotone class).

On the other hand, the class  $\mathcal{A}$ , consisting of events  $\{\emptyset, A, S\}$ , is a monotone class, but neither a field nor  $\sigma$ -field. If  $A$  and  $B$  are two events, then the smallest field  $\mathcal{A}$  containing  $A$  and  $B$  must contain also the sets  $A^c, B^c$ , the intersections  $A \cap B, A \cap B^c, A^c \cap B, A^c \cap B^c$ , as well as their unions  $A \cup B, A \cup B^c, A^c \cup B$ , and  $A^c \cup B^c$ . The closure property implies that unions such as  $(A \cap B) \cup (A \cup B^c)$ , must also belong to  $\mathcal{A}$ .

We are ready to present the final step in our construction.

**Theorem 1.4.3** *For any nonempty class  $\mathcal{K}$  of subsets of  $S$  there exists a unique smallest field ( $\sigma$ -field, monotone class) containing all sets in  $\mathcal{K}$ . It is called the field ( $\sigma$ -field, monotone class) generated by  $\mathcal{K}$ .*

*Proof.* We will prove the assertion for fields. Observe first that if  $\mathcal{A}_1$  and  $\mathcal{A}_2$  are fields, then their intersection  $\mathcal{A}_1 \cap \mathcal{A}_2$  (i.e., the class of sets that belong to both  $\mathcal{A}_1$  and  $\mathcal{A}_2$ ) is also a field. For instance, if  $A, B \in \mathcal{A}_i$  ( $i = 1, 2$ ), then  $A \cup B \in \mathcal{A}_i$  because each  $\mathcal{A}_i$  is a field, and consequently  $A \cup B \in \mathcal{A}_1 \cap \mathcal{A}_2$ . A similar argument holds for intersections and complements.

Note that if  $\mathcal{A}_1$  and  $\mathcal{A}_2$  contain the class  $\mathcal{K}$ , then the intersection  $\mathcal{A}_1 \cap \mathcal{A}_2$  also contains  $\mathcal{K}$ . The foregoing property extends to any intersection of fields containing  $\mathcal{K}$  (not only the intersections of two such fields).

Now let  $\mathcal{C}$  be the intersection of *all* fields containing  $\mathcal{K}$ . We claim that  $\mathcal{C}$  is the minimal unique field containing  $\mathcal{K}$ . We have to show that (1)  $\mathcal{C}$  exists, (2)  $\mathcal{C}$  is a field containing  $\mathcal{K}$ , (3)  $\mathcal{C}$  is unique, and (4)  $\mathcal{C}$  is minimal.

For property (1) it is enough to show that there exists at least one field containing  $\mathcal{K}$ . We may take here the class of all subsets of  $S$ : it is a field (as well as a  $\sigma$ -field and monotone class), and it contains all sets in  $\mathcal{K}$ . Property (2) follows from the fact that the intersection of fields containing  $\mathcal{K}$  is a field containing  $\mathcal{K}$ . Property (3) (i.e.,

uniqueness of  $\mathcal{C}$ ) follows from the fact that the result of the operation of intersection is unique.

Finally, suppose that there exists a field  $\mathcal{C}'$  containing  $\mathcal{K}$  such that  $\mathcal{C}' \subset \mathcal{C}$ . Then  $\mathcal{C}'$  must appear as one of the factors in the intersection defining  $\mathcal{C}$  so that  $\mathcal{C} \subset \mathcal{C}'$ . Consequently  $\mathcal{C}' = \mathcal{C}$ . This completes the proof for the case of fields. The proofs for  $\sigma$ -fields and monotone classes are exactly the same, since an intersection of  $\sigma$ -fields (or monotone classes) containing  $\mathcal{K}$  is again a  $\sigma$ -field (monotone class) containing  $\mathcal{K}$ .  $\square$

One may find it disturbing that Theorem 1.4.3 asserts the existence and uniqueness of some objects without giving a clue as to how to find them in practical situations. In fact the nonconstructive character of the theorem, combined with its generality, is instead a great help. As we will see in Chapter 2, the natural objects of our interest (the domains of definition of probability) will be  $\sigma$ -fields of events. Beyond the trivial situations of finite or countably infinite sample spaces  $\mathcal{S}$ , where one can always consider the maximal  $\sigma$ -field consisting of all subsets of  $\mathcal{S}$ , one is forced to restrict consideration to classes of events that form  $\sigma$ -fields generated by some "simple" events. The events in these  $\sigma$ -fields are typically of a very rich structure, and one seldom has useful criteria for distinguishing events (elements of the  $\sigma$ -field in question) from "non-events," that is, subsets of  $\mathcal{S}$  to which probabilities are not assigned. However, as shown by the two examples below, the smallest  $\sigma$ -field generated by some class is richer than the smallest field generated by the same class.

#### ■ EXAMPLE 1.22

A point moves randomly on the plane, and its location is recorded at some time  $t$ . The outcome of this experiment is the pair  $(x, y)$  of coordinates of the observed location of the point (e.g., imagine here the location of a particle of dust in a liquid, tossed about by random hits from molecules of the medium, and performing Brownian motion; or imagine a location of a previously marked bird at the time of its capture in a bird migration study or the ages of both husband and wife at the time one of them dies).

In any study of this kind (regardless its ultimate purpose), the "natural" sample space  $\mathcal{S}$  is a plane or part of the plane, (the positive quadrant, etc.). The "simple" events here are of the form  $a < x \leq b, c < y \leq d$ , that is, rectangles with sides parallel to the axes. The reason for distinguishing these events as "simple" is that, as will be explained in later chapters, it is often easy to assign probabilities to these events. The reason for the particular configuration of strict and non-strict inequalities (i.e., north and east side included, south and west side excluded) will also become apparent from the analysis below. To simplify the language, we will call such events Rectangles, and use a capital letter to signify the specific assumption about which sides are included and which are not. Naturally we will allow for infinite Rectangles, such as  $\{a < x \leq \infty, -\infty < y \leq b\}$ .

It is easy to determine the field generated by all Rectangles: These are events that result from finite operations on Rectangles. Clearly, the comple-

ment of a Rectangle is a union of at most eight disjoint (infinite) Rectangles (see Figure 1.7), whereas the intersection of Rectangles is again a Rectangle (or is empty).

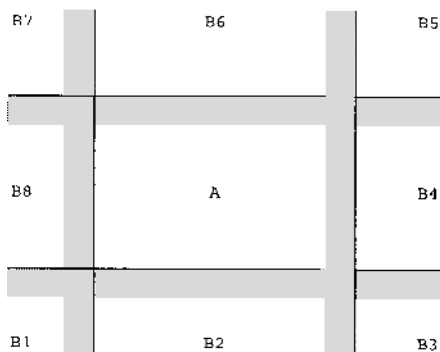


Figure 1.7 Complement of a Rectangle

Since unions are reduced to intersections of complements by De Morgan's laws, every element of the smallest field containing all Rectangles is the union of a finite number of disjoint Rectangles. On the other hand, there exist events that do not belong to this field of events. As a simple example, one might be interested in the event that the point  $(x, y)$  lies within distance  $r$  from some fixed point (from the initial location of the particle, the point of release of the bird, etc.). This event is a circle on the plane, and hence a subset of  $\mathcal{S}$ , which is not decomposable into a finite number of Rectangles. On the other hand, a circle does belong to the  $\sigma$ -field spanned by Rectangles: it is representable as a countable union of Rectangles, or equivalently, as an infinite intersection of sets built up of Rectangles.

Similarly in this example there are other events, which are not in the field generated by Rectangles and which could be considered, such as triangles, rectangles with sides not parallel to the axes, and ellipses.

### ■ EXAMPLE 1.23

Take an experiment consisting of tossing a coin infinitely many times. The "natural" sample space  $\mathcal{S}$  is the space of all infinite sequences  $x = (\xi_1, \xi_2, \dots)$ , where  $\xi_i = 0$  or 1 (or any other two distinct symbols representing heads and tails). The "simple" events here are of the form "heads on the  $n$ th toss," that is, sets of all infinite sequences  $x = (\xi_1, \xi_2, \dots)$  with the  $n$ th coordinate  $\xi_n$  satisfying  $\xi_n = 0$ . The events in the field generated by the simple events are of the form "heads on tosses  $k_1, \dots, k_n$  and tails on tosses  $r_1, \dots, r_m$ ," with both  $m$  and  $n$  finite and the outcomes of all other tosses remaining unspecified.

An event that does not belong to this field, but does belong to the  $\sigma$ -field generated by the simple events, is the event that "as the number of tosses increases, the frequency of heads approaches a limit." Clearly, to determine

whether or not this event occurs, it does not suffice to know any finite number of coordinates  $\xi_n$ .

To generalize this example, replace the outcome of the coin tosses by the result of some experiment repeated infinitely many times. This way the coordinate  $\xi_n$  carries more information than it does for the outcome of  $n$ th coin toss. The “simple” events are now of the form of sets of sequences  $x = (\xi_1, \xi_1, \dots)$  with  $\xi_i \in A_i$  for  $i = 1, \dots, n$ , while the  $\xi_i$ ’s for  $i > n$  are unconstrained. Here  $A_1, \dots, A_n$  are events that occur at the first  $n$  times of observations. The “simple” events described above, of an obvious interest and importance both in applications and in building the theory, are called “cylinder” events. The smallest  $\sigma$ -field containing all cylinder events comprises all events that may be of interest, including those that are obtained through limits of sequences of cylinder events.

**PROBLEMS**

**1.4.1** Let  $B_1, B_2, \dots$  be a countable partition of  $S$ ; that is,  $B_i \cap B_j = \emptyset$  for all  $i \neq j$ , and  $\bigcup_i B_i = S$ . Let  $A_n = B_n \cup B_{n+1} \cup \dots$ . Find  $\lim A_n$ .

**1.4.2** Assume that John will live forever. He plays a certain game each day. Let  $A_i$  be the event that he wins the game on the  $i$ th day.

(i) Let  $B$  be the event that John will win every game starting on January 1, 2015. Label the following statements as true or false: (a)  $B = \liminf A_n$ . (b)  $B \subset \liminf A_n$ . (c)  $B \supset \limsup A_n$ . (d)  $B = \limsup A_n$ .

(ii) Assume now that John starts playing on a Monday. Match the following events  $C_1$  through  $C_9$  with events  $D_1$  through  $D_{11}$ :

- $C_1 =$  John loses infinitely many games.
- $C_2 =$  When John loses on a Thursday, he wins on the following Sunday.
- $C_3 =$  John never wins on three consecutive days.
- $C_4 =$  John wins every Wednesday.
- $C_5 =$  John wins on infinitely many Wednesdays.
- $C_6 =$  John wins on a Wednesday.
- $C_7 =$  John never wins on a weekend.
- $C_8 =$  John wins infinitely many games and loses infinitely many games.
- $C_9 =$  If John wins on some day, he never loses on the next day.

$$D_1 = \bigcap_{j=0}^{\infty} [A_{7j+4} \cup A_{7(j+1)}].$$

$$D_2 = \bigcap_{j=0}^{\infty} A_{7j+3}.$$

$$D_3 = \bigcap_{j=0}^{\infty} [A_{7j+6}^c \cap A_{7(j+1)}^c].$$

$$D_4 = \bigcup_{n=0}^{\infty} \left\{ \left[ \bigcap_{i=1}^n A_i^c \right] \cap \left[ \bigcap_{k=n+1}^{\infty} A_k \right] \right\}.$$

$$\begin{aligned}
 D_5 &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_{7k+3}. \\
 D_6 &= \bigcap_{i=1}^{\infty} [A_i^c \cup A_{i+1}^c \cup A_{i+2}^c]. \\
 D_7 &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k^c. \\
 D_8 &= \bigcup_{j=0}^{\infty} A_{7j+3}. \\
 D_9 &= \left\{ \bigcup_{j=0}^{\infty} [A_{7j+6} \cup A_{7(j+1)}] \right\}^c. \\
 D_{10} &= \bigcap_{i=0}^{\infty} [A_i \cap A_{i+1} \cap A_{i+2}]^c. \\
 D_{11} &= \left[ \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k^c \right] \cap \left[ \bigcap_{j=1}^{\infty} \bigcup_{m=j}^{\infty} A_m \right].
 \end{aligned}$$

**1.4.3** Let  $A_1, \dots, A_n$  be distinct subsets of  $\mathcal{S}$ . (i) Find the maximum number of sets (including  $\mathcal{S}$  and  $\emptyset$ ) of the smallest field containing  $A_1, \dots, A_n$ . (ii) Find the maximum number of sets in this field if  $A_{n-1} \subset A_n$ . (iii) Answer (ii) if  $A_1 \subset A_2 \subset \dots \subset A_n$ . (iv) Answer (ii) if  $A_1 = \dots = A_n = \emptyset$ . (v) Answer (i)–(iv) for a  $\sigma$ -field.

**1.4.4** For  $0 < \alpha < 1$ , let  $I(\alpha) = \{x : 1 - \alpha < x < 1 + \alpha\}$ . Consider a sequence  $\alpha_1, \alpha_2, \dots$  of numbers satisfying  $0 < \alpha_n < 1$  for all  $n$ , and let  $A_n = I(\alpha_n)$ . (i) Find  $\limsup A_n$  and  $\liminf A_n$ . (ii) Find conditions, expressed in terms of  $\alpha_n$ , under which  $\lim A_n$  exists, and find this limit. (iii) Define  $J(\alpha) = \{x : 1 - \alpha \leq x \leq 1 + \alpha\}$  and  $B_n = J(\alpha_n)$ . Answer questions (i) and (ii) for sequence  $\{B_n\}$ .

**1.4.5** Let  $\mathcal{S} = \{0, 1, 2, \dots\}$  be the set of all integers. For  $A \subset \mathcal{S}$ , let  $f_n(A)$  be the number of elements in the intersection  $A \cap \{0, 1, \dots, n\}$ . Let  $\mathcal{A}$  be the class of all sets  $A$  for which the limit

$$q(A) = \lim_{n \rightarrow \infty} \frac{f_n(A)}{n}$$

exists. Show that  $\mathcal{A}$  is not a field. [Hint: Let  $A_1 = \{1, 3, 5, \dots\}$  and  $A_2 = \{\text{all odd integers between } 2^{2n} \text{ and } 2^{2n+1} \text{ and all even integers between } 2^{2n+1} \text{ and } 2^{2n+2} \text{ for } n = 0, 1, \dots\}$ . Show that both  $A_1$  and  $A_2$  are in  $\mathcal{A}$  but  $A_1 \cap A_2 \notin \mathcal{A}$ .]

**1.4.6** Let  $\mathcal{S} = (-\infty, +\infty)$ . Show that the class of all finite unions of intervals of the form  $[a, b]$ ,  $(a, b)$ ,  $[a, b)$ , and  $(a, b]$ , with possibly infinite  $a$  or  $b$  (intervals of the form  $[a, \infty)$ , etc.) forms a field.