Preface

This book has been written for a first course in probability and was developed from lectures given at the University of Illinois during the last five years. Most of the students have been juniors, seniors, and beginning graduates, from the fields of mathematics, engineering and physics. The only formal prerequisite is calculus, but an additional degree of mathematical maturity may be helpful.

In talking about nondiscrete probability spaces, it is difficult to avoid measure-theoretic concepts. However, to develop extensive formal machinery from measure theory before going into probability (as is done in most graduate programs in mathematics) would be inappropriate for the particular audience to whom the book is addressed. Thus I have tried to suggest, when possible, the underlying measure-theoretic ideas, while emphasizing the probabilistic way of thinking, which is likely to be quite novel to anyone studying this subject for the first time.

The major field of application considered in the book is statistics (Chapter 8). In addition, some of the problems suggest connections with the physical sciences. Chapters 1 to 5, and Chapter 8 will serve as the basis for a one-semester or a two-quarter course covering both probability and statistics. If probability alone is to be considered, Chapter 8 may be replaced by Chapter 6 and Chapter 7, as time permits. An asterisk before a section or a problem indicates material that I have normally omitted (without loss of continuity), either because it involves subject matter that many of the students have not been exposed to (for example, complex variables) or because it represents too concentrated a dosage of abstraction.

A word to the instructor about notation. In the most popular terminology, $P\{X \leq x\}$ is written for the probability that the random variable X assumes a value less than or equal to the number x. I tried this once in my class, and I found that as the semester progressed, the capital X tended to become smaller in the students' written work, and the small x larger. The following semester, I switched to the letter R for random variable, and this notation is used throughout the book.

Fairly detailed solutions to some of the problems (and numerical answers to others) are given at the end of the book.

I hope that the book will provide an introduction to more advanced courses in probability and real analysis and that it makes the abstract ideas to be encountered later more meaningful. I also hope that nonmathematics majors who come in contact with probability theory in their own areas find the book useful. A brief list of references, suitable for future study, is given at the end of the book.

I am grateful to the many students and colleagues who have influenced my own understanding of probability theory and thus contributed to this book.

I also thank Mrs. Dee Keel for her superb typing, and the staff of Wiley for its continuing interest and assistance.

Urbana, Illinois, 1969 *Robert B. Ash*

Contents

7. MARKOV CHAINS

8. INTRODUCTION TO STATISTICS

1

Basic Concepts

1.1 INTRODUCTION

The origin of probability theory lies in physical observations associated with games of chance. It was found that if an "unbiased" coin is tossed independently n times, where n is very large, the relative frequency of heads, that is, the ratio of the number of heads to the total number of tosses, is very likely to be very close to 1/2. Similarly, if a card is drawn from a perfectly shuffled deck and then is replaced, the deck is reshuffled, and the process is repeated over and over again, there is (in some sense) convergence of the relative frequency of spades to 1/4.

In the card experiment there are 52 possible outcomes when a single card is drawn. There is no reason to favor one outcome over another (the principle of "insufficient reason" or of "least astonishment"), and so the early workers in probability took as the probability of obtaining a spade the number of favorable outcomes divided by the total number of outcomes, that is, 13/52 or 1/4.

This so-called "classical definition" of probability (the probability of an event is the number of outcomes favorable to the event, divided by the total number of outcomes, where all outcomes are equally likely) is first of all restrictive (it considers only experiments with a finite number of outcomes) and, more seriously, circular (no matter how you look at it, "equally likely" essentially means "equally probable," and thus we are using the concept of probability to define probability itself). Thus we cannot use this idea as the basis of a mathematical theory of probability; however, the early probabilists were not prevented from deriving many valid and useful results.

Similarly, an attempt at a frequency definition of probability will cause trouble. If S_n is the number of occurrences of an event in n independent performances of an experiment, we expect physically that the relative frequency S_n/n should coverge to a limit; however, we cannot assert that the limit exists in a mathematical sense. In the case of the tossing of an unbiased coin, we expect that $S_n/n \to 1/2$, but a conceivable outcome of the process is that the coin will keep coming up heads forever. In other words it is possible that $S_n/n \to 1$, or that $S_n/n \to$ any number between 0 and 1, or that S_n/n has no limit at all.

In this chapter we introduce the concepts that are to be used in the construction of a mathematical theory of probability. The first ingredient we need is a set Ω, called the *sample space*, representing the collection of possible outcomes of a random experiment. For example, if a coin is tossed once we may take $\Omega = \{H, T\}$, where H corresponds to a head and T to a tail. If the coin is tossed twice, this is a different experiment and we need a different Ω, say $\{HH, HT, TH, TT\}$; in this case one performance of the experiment corresponds to two tosses of the coin.

If a single die is tossed, we may take Ω to consist of six points, say $\Omega = \{1, 2, \ldots, 6\}$. However, another possible sample space consists of two points, corresponding to the outcomes "N is even" and "N is odd," where N is the result of the toss. Thus different sample spaces can be associated with the same experiment. The nature of the particular problem under consideration will dictate which sample space is to be used. If we are interested, for example, in whether or not $N \geq 3$ in a given performance of the experiment, the second sample space, corresponding to "N even" and "N odd," will not be useful to us.

In general, the only physical requirement on Ω is that *a given performance of the experiment must produce a result corresponding to exactly one of the points of* Ω. We have as yet no mathematical requirements on Ω; it is simply a set of points.

Next we come to the notion of *event*. An "event" associated with a random experiment corresponds to a question about the experiment that has a yes or no answer, and this in turn is associated with a subset of the sample space. For example, if a coin is tossed twice and $\Omega = \{HH, HT, TH, TT\}$, "the number of heads is ≤ 1" will be a condition that either occurs or does not occur in a given performance of the experiment. That is, after the experiment is performed, the question "Is the number of heads ≤ 1?" can be answered yes or no. The subset of Ω corresponding to a "yes" answer is $A = \{HT, TH, TT\}$; that is, if the outcome of the experiment is HT, TH, or TT, the answer

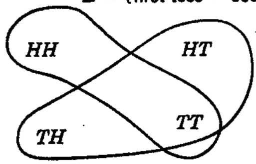

FIGURE 1.1.1 Coin-Tossing Experiment.

to the question "Is the number of heads ≤ 1?" will be "yes," and if the outcome is HH, the answer will be "no." Similarly, the subset of Ω associated with the "event" that the result of the first toss is the same as the result of the second toss is $B = \{HH, TT\}$.

Thus an *event* is defined as a subset of the sample space, that is, a collection of points of the sample space. (We shall qualify this in the next section.)

Events will be denoted by capital letters at the beginning of the English alphabet, such as A, B, C, and so on. An event may be characterized by listing all of its points, or equivalently by describing the conditions under which the event will occur. For example, in the coin-tossing experiment just considered, we write

$$A = \{\text{the number of heads is less than or equal to } 1\}$$

This expression is to be read as "A is the set consisting of those outcomes which satisfy the condition that the number of heads is less than or equal to 1," or, more simply, "A is the event that the number of heads is less than or equal to 1." The event A consists of the points HT, TH, and TT; therefore we write $A = \{HT, TH, TT\}$, which is to be read "A is the event consisting of the points HT, TH, and TT." As another example, if B is the event that the result of the first toss is the same as the result of the second toss, we may describe B by writing $B = \{\text{first toss} = \text{second toss}\}$ or, equivalently, $B = \{HH, TT\}$ (see Figure 1.1.1).

Each point belonging to an event A is said to be *favorable* to A. The event A will occur in a given performance of the experiment if and only if the outcome of the experiment corresponds to one of the points of A. The entire sample space Ω is said to be the *sure* (or *certain*) event; if *must* occur on any given performance of the experiment. On the other hand, the event consisting of none of the points of the sample space, that is, the empty set $\varnothing$, is called the *impossible event*; it can *never* occur in a given performance of the experiment.

1.2 ALGEBRA OF EVENTS (BOOLEAN ALGEBRA)

Before talking about the assignment of probabilities to events, we introduce some operations by which new events are formed from old ones. These operations correspond to the construction of compound sentences by use of the connectives "or," "and," and "not." Let A and B be events in the same sample space. Define the *union* of A and B (denoted by $A \cup B$) as the set consisting of those points belonging to *either A or B or both.* (Unless otherwise specified, the word "or" will have, for us, the inclusive connotation. In other words, the statement "p or q" will always mean "p or q or both.") Define the *intersection* of A and B, written $A \cap B$, as the set of points that belong to *both* A and B. Define the *complement* of A, written A^c, as the set of points which do *not* belong to A.

▶ **Example 1.** Consider the experiment involving the toss of a single die, with $N =$ the result; take a sample space with six points corresponding to $N = 1, 2, 3, 4, 5, 6$. For convenience, label the points of the sample space by the integers 1 through 6.

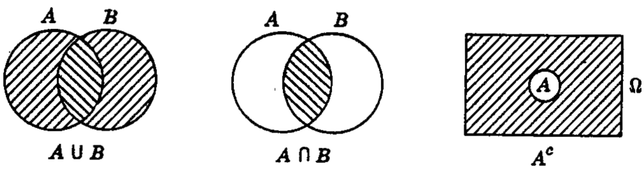

FIGURE 1.2.1 Venn Diagrams.

$$\text{Let } A = \{N \text{ is even}\} \qquad \text{and} \qquad B = \{N \geq 3\}$$

Then

$$A \cup B = \{N \text{ is even or } N \geq 3\} = \{2, 3, 4, 5, 6\}$$
$$A \cap B = \{N \text{ is even and } N \geq 3\} = \{4, 6\}$$
$$A^c = \{N \text{ is not even}\} = \{1, 3, 5\}$$
$$B^c = \{N \text{ is not } \geq 3\} = \{N < 3\} = \{1, 2\} \blacktriangleleft$$

Schematic representations (called *Venn diagrams*) of unions, intersections, and complements are shown in Figure 1.2.1.

Define the *union of n events* $A_1, A_2, \ldots, A_n$ (notation: $A_1 \cup \cdots \cup A_n$, or $\bigcup_{i=1}^n A_i$) as the set consisting of those points which belong to *at least one* of the events $A_1, A_2, \ldots, A_n$. Similarly define the union of an infinite sequence of events $A_1, A_2, \ldots$ as the set of points belonging to at least one of the events $A_1, A_2, \ldots$ (notation: $A_1 \cup A_2 \cup \cdots$, or $\bigcup_{i=1}^\infty A_i$).

Define the *intersection of n events* $A_1, \ldots, A_n$ as the set of points belonging to *all* of the events $A_1, \ldots, A_n$ (notation: $A_1 \cap A_2 \cap \cdots \cap A_n$, or $\bigcap_{i=1}^n A_i$). Similarly define the intersection of an infinite sequence of events as the set of points belonging to all the events in the sequence (notation: $A_1 \cap A_2 \cap \cdots$, or $\bigcap_{i=1}^{\infty} A_i$). In the above example, with $A = \{N \text{ is even}\} = \{2, 4, 6\}$, $B = \{N \geq 3\} = \{3, 4, 5, 6\}$, $C = \{N = 1 \text{ or } N = 5\} = \{1, 5\}$, we have

$$A \cup B \cup C = \Omega, \qquad A \cap B \cap C = \varnothing$$
$$A \cup B^c \cup C = \{2, 4, 6\} \cup \{1, 2\} \cup \{1, 5\} = \{1, 2, 4, 5, 6\}$$
$$(A \cup C) \cap [(A \cap B)^c] = \{1, 2, 4, 5, 6\} \cap \{4, 6\}^c = \{1, 2, 5\}$$

Two events in a sample space are said to be *mutually exclusive* or *disjoint* if A and B have no points in common, that is, if it is impossible that both A and B occur during the *same* performance of the experiment. In symbols, A and B are mutually exclusive if $A \cap B = \varnothing$. In general the events A_1, $A_2, \ldots, A_n$ are said to be mutually exclusive if no two of the events have a point in common; that is, no more than one of the events can occur during the same performance of the experiment. Symbolically, this condition may be written

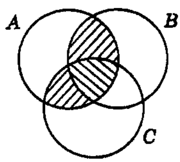

FIGURE 1.2.2 $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

$$A_i \cap A_j = \varnothing \qquad \text{for } i \neq j$$

Similarly, infinitely many events $A_1, A_2, \ldots$ are said to be mutually exclusive if $A_i \cap A_j = \varnothing$ for $i \neq j$.

In some ways the algebra of events is similar to the algebra of real numbers, with union corresponding to addition and intersection to multiplication. For example, the commutative and associative properties hold.

$$A \cup B = B \cup A, \qquad A \cup (B \cup C) = (A \cup B) \cup C$$
$$A \cap B = B \cap A, \qquad A \cap (B \cap C) = (A \cap B) \cap C \qquad (1.2.1)$$

Furthermore, we can prove that for events A, B, and C in the same sample space we have

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \qquad (1.2.2)$$

There are several ways to establish this; for example, we may verify that the sets of both the left and right sides of the equality above are represented by the area in the Venn diagram of Figure 1.2.2.

Another approach is to use the definitions of union and intersection to show that the sets in question have precisely the same members; that is, we show that any point which belongs to the set on the left necessarily belongs to the set on the right, and conversely. To do this, we proceed as follows.

$$\begin{aligned} x \in A \cap (B \cup C) &\Rightarrow x \in A \quad \text{and} \quad x \in B \cup C \\ &\Rightarrow x \in A \quad \text{and} \quad (x \in B \text{ or } x \in C) \end{aligned}$$

(The symbol $\Rightarrow$ means "implies," and $\Leftrightarrow$ means "implies and is implied by.")

CASE 1. $x \in B$. Then $x \in A$ and $x \in B$, so $x \in A \cap B$, so $x \in (A \cap B) \cup (A \cap C)$.

CASE 2. $x \in C$. Then $x \in A$ and $x \in C$, so $x \in A \cap C$, so $x \in (A \cap B) \cup (A \cap C)$.

Thus $x \in A \cap (B \cup C) \Rightarrow x \in (A \cap B) \cup (A \cap C)$; that is, $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. (The symbol $\subset$ is read "is a subset of"; we say that $A_1 \subset A_2$ provided that $x \in A_1 \Rightarrow x \in A_2$; see Figure 1.2.3. Notice that, according to this definition, a set A is a subset of itself: $A \subset A$.)

Conversely: Let $x \in (A \cap B) \cup (A \cap C)$. Then $x \in A \cap B$ or $x \in A \cap C$.

CASE 1. $x \in A \cap B$. Then $x \in B$, so $x \in B \cup C$, so $x \in A \cap (B \cup C)$.

CASE 2. $x \in A \cap C$. Then $x \in C$, so $x \in B \cup C$, so $x \in A \cap (B \cup C)$.

Thus $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$; hence

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

As another example we show that

$$(A_1 \cup A_2 \cup \cdots \cup A_n)^c = A_1^c \cap A_2^c \cap \cdots \cap A_n^c \qquad (1.2.3)$$

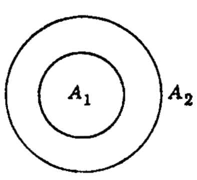

FIGURE 1.2.3 $A_1 \subset A_2$.

The steps are as follows.

$$
\begin{aligned}
x \in (A_1 \cup \cdots \cup A_n)^c &\Leftrightarrow x \notin A_1 \cup \cdots \cup A_n \\
&\Leftrightarrow \text{it is not the case that } x \text{ belongs to at least one of the } A_i \\
&\Leftrightarrow x \in \text{none of the } A_i \\
&\Leftrightarrow x \in A_i^c \quad \text{for all } i \\
&\Leftrightarrow x \in A_1^c \cap \cdots \cap A_n^c
\end{aligned}
$$

An identical argument shows that

$$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c \tag{1.2.4}$$

and similarly

$$\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c \qquad \text{i.e. } (A_1 \cap \cdots \cap A_n)^c = A_1^c \cup \cdots \cup A_n^c \tag{1.2.5}$$

Also

$$\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c \tag{1.2.6}$$

The identities (1.2.3)–(1.2.6) are called the *DeMorgan laws.*

In many ways the algebra of events differs from the algebra of real numbers, as some of the identities below indicate.

$$
\begin{aligned}
A \cup A &= A & A \cup A^c &= \Omega \\
A \cap A &= A & A \cap A^c &= \varnothing \\
A \cap \Omega &= A & A \cup \varnothing &= A \\
A \cup \Omega &= \Omega & A \cap \varnothing &= \varnothing
\end{aligned} \tag{1.2.7}
$$

Another method of verifying relations among events involves algebraic manipulation, using the identities already derived. Four examples are given below; in working out the identities, it may be helpful to write $A \cup B$ as $A + B$ and $A \cap B$ as AB.

$$1. \quad A \cup (A \cap B) = A \tag{1.2.8}$$

PROOF.

$$A + AB = A\Omega + AB = A(\Omega + B) = A\Omega = A$$

$$2. \quad (A \cup B) \cap (A \cup C) = A \cup (B \cap C) \tag{1.2.9}$$

PROOF.

$$\begin{aligned}(A+B)(A+C) &= (A+B)A + (A+B)C \\ &= AA + AB + AC + BC \qquad \text{(note } AB = BA\text{)} \\ &= A(\Omega + B + C) + BC \\ &= A\Omega + BC \\ &= A + BC\end{aligned}$$

$$3. \quad A \cup [(A \cap B)^c] = \Omega \tag{1.2.10}$$

PROOF.

$$A + (AB)^c = A + A^c + B^c = \Omega + B^c = \Omega$$

$$4. \quad (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B) = A \cup B \tag{1.2.11}$$

PROOF.

$$\begin{aligned}AB^c + AB + A^cB &= AB^c + AB + AB + A^cB \qquad \text{[see (1.2.7)]} \\ &= A(B^c + B) + (A + A^c)B \\ &= A\Omega + \Omega B \\ &= A + B\end{aligned}$$

(see Figure 1.2.4).

As another example, let Ω be the set of nonnegative real numbers. Let

$$A_n = \left[0, 1 - \frac{1}{n}\right] = \left\{x \in \Omega: 0 \le x \le 1 - \frac{1}{n}\right\} \qquad n = 1, 2, \ldots$$

(This will be another common way of describing an event. It is to be read: "A_n is the set consisting of those points x in Ω such that $0 \le x \le 1 - 1/n$." If there is no confusion about what space Ω we are considering, we shall simply write $A_n = \{x: 0 \le x \le 1 - 1/n\}$.) Then

$$\bigcup_{n=1}^{\infty} A_n = [0, 1) = \{x: 0 \le x < 1\}$$

$$\bigcap_{n=1}^{\infty} A_n = \{0\}$$

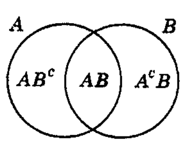

FIGURE 1.2.4 Venn Diagram Illustrating $(A \cap B^c) \cup (A \cap B) \cup (A^c \cap B) = A \cup B$.

As an illustration of the DeMorgan laws,

$$\left(\bigcup_{n=1}^{\infty} A_n\right)^c = [0, 1)^c = [1, \infty) = \{x: x \geq 1\}$$

$$\bigcap_{n=1}^{\infty} A_n^c = \bigcap_{n=1}^{\infty} \left(1 - \frac{1}{n}, \infty\right) = [1, \infty)$$

(Notice that $x > 1 - 1/n$ for all $n = 1, 2, \ldots \Leftrightarrow x \geq 1$.) Also

$$\left(\bigcap_{n=1}^{\infty} A_n\right)^c = \{0\}^c = (0, \infty) = \{x: x > 0\}$$

$$\bigcup_{n=1}^{\infty} A_n^c = \bigcup_{n=1}^{\infty} \left(1 - \frac{1}{n}, \infty\right) = (0, \infty)$$

PROBLEMS

1. An experiment involves choosing an integer N between 0 and 9 (the sample space consists of the integers from 0 to 9, inclusive). Let $A = \{N \leq 5\}$, $B = \{3 \leq N \leq 7\}$, $C = \{N$ is even and $N > 0\}$. List the points that belong to the following events.

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

2. Let A, B, and C be *arbitrary* events in the same sample space. Let D_1 be the event that at least two of the events A, B, C occur; that is, D_1 is the set of points common to at least two of the sets A, B, C.
   Let $D_2 = \{$exactly two of the events A, B, C occur$\}$
   $D_3 = \{$at least one of the events A, B, C occur$\}$
   $D_4 = \{$exactly one of the events A, B, C occur$\}$
   $D_5 = \{$not more than two of the events A, B, C occur$\}$
   Each of the events D_1 through D_5 can be expressed in terms of A, B, and C by using unions, intersections, and complements. For example, $D_3 = A \cup B \cup C$. Find suitable expressions for D_1, D_2, D_4, and D_5.

3. A public opinion poll (circa 1850) consisted of the following three questions:
   (a) Are you a registered Whig?
   (b) Do you approve of President Fillmore's performance in office?
   (c) Do you favor the Electoral College system?
   A group of 1000 people is polled. Assume that the answer to each question must be either "yes" or "no." It is found that:
   550 people answer "yes" to the third question and 450 answer "no."
   325 people answer "yes" exactly twice; that is, their responses contain two "yeses" and one "no."

100 people answer "yes" to all three questions.
125 registered Whigs approve of Fillmore's performance.
How many of those who favor the Electoral College system do not approve of Fillmore's performance, and in addition are not registered Whigs? HINT: Draw a Venn diagram.

4. If A and B are events in a sample space, define $A - B$ as the set of points which belong to A but not to B; that is, $A - B = A \cap B^c$. Establish the following.
   (a) $A \cap (B - C) = (A \cap B) - (A \cap C)$
   (b) $A - (B \cup C) = (A - B) - C$
   Is is true that $(A - B) \cup C = (A \cup C) - B$?
5. Let Ω be the reals. Establish the following.

$$(a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right] = \bigcup_{n=1}^{\infty} \left[a + \frac{1}{n}, b\right)$$

$$[a, b] = \bigcap_{n=1}^{\infty} \left[a, b + \frac{1}{n}\right) = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b\right]$$

6. If A and B are disjoint events, are A^c and B^c disjoint? Are $A \cap C$ and $B \cap C$ disjoint? What about $A \cup C$ and $B \cup C$?
7. If $A_n \subset A_{n-1} \subset \cdots \subset A_1$, show that $\bigcap_{i=1}^n A_i = A_n$, $\bigcup_{i=1}^n A_i = A_1$.
8. Suppose that $A_1, A_2, \ldots$ is a sequence of subsets of Ω, and we know that for each n, $\bigcap_{i=1}^n A_i$ is not empty. Is it true that $\bigcap_{i=1}^\infty A_i$ is not empty? (A related question about real numbers: if, for each n, we have $\sum_{i=1}^n a_i < b$, is it true that $\sum_{i=1}^\infty a_i < b$?)
9. If $A, B_1, B_2, \ldots$ are arbitrary events, show that

$$A \cap \left(\bigcup_i B_i\right) = \bigcup_i (A \cap B_i)$$

This is the distributive law with infinitely many factors.

1.3 PROBABILITY

We now consider the assignment of probabilities to events. A technical complication arises here. It may not always be possible to regard all subsets of Ω as events. We may discard or fail to measure some of the information in the outcome corresponding to the point $\omega \in \Omega$, so that for a given subset A of Ω, it may not be possible to give a yes or no answer to the question "Is $\omega \in A$?" For example, if the experiment involves tossing a coin five times, we may record the results of only the first three tosses, so that $A = \{$at least four heads$\}$ will not be "measurable"; that is, membership of $\omega \in A$ cannot be determined from the given information about ω.

In a given problem there will be a particular class of subsets of Ω called the "class of events." For reasons of mathematical consistency, we require that the event class $\mathscr{F}$ form a *sigma field*, which is a collection of subsets of Ω satisfying the following three requirements.

$$\Omega \in \mathscr{F} \tag{1.3.1}$$

$$A_1, A_2, \ldots \in \mathscr{F} \quad \text{implies} \quad \bigcup_{n=1}^{\infty} A_n \in \mathscr{F} \tag{1.3.2}$$

That is, $\mathscr{F}$ is closed under finite or countable union.

$$A \in \mathscr{F} \quad \text{implies} \quad A^c \in \mathscr{F} \tag{1.3.3}$$

That is, $\mathscr{F}$ is closed under complementation.

Notice that if $A_1, A_2, \ldots \in \mathscr{F}$, then $A_1^c, A_2^c, \ldots \in \mathscr{F}$ by (1.3.3); hence $\bigcup_{n=1}^{\infty} A_n^c \in \mathscr{F}$ by (1.3.2). By the DeMorgan laws, $\bigcap_{n=1}^{\infty} A_n = (\bigcup_{n=1}^{\infty} A_n^c)^c$; hence, by (1.3.3), $\bigcap_{n=1}^{\infty} A_n \in \mathscr{F}$. Thus $\mathscr{F}$ is closed under finite or countable intersection. Also, by (1.3.1) and (1.3.3), the empty set $\varnothing$ belongs to $\mathscr{F}$.

Thus, for example, if the question "Did A_n occur?" has a definite answer for $n = 1, 2, \ldots$, so do the questions "Did at least one of the A_n occur?" and "Did all the A_n occur?"

Note also that if we apply the algebraic operations of Section 1.2 to sets in $\mathscr{F}$, the new sets we obtain still belong to $\mathscr{F}$.

In many cases we shall be able to take $\mathscr{F}$ = the collection of *all* subsets of Ω, so that every subset of Ω is an event. Problems in which $\mathscr{F}$ cannot be chosen in this way generally arise in uncountably infinite sample spaces; for example, Ω = the reals. We shall return to this subject in Chapter 2.

We are now ready to talk about the assignment of probabilities to events. If $A \in \mathscr{F}$, the probability $P(A)$ should somehow reflect the long-run relative frequency of A in a large number of independent repetitions of the experiment. Thus $P(A)$ should be a number between 0 and 1, and $P(\Omega)$ should be 1.

Now if A and B are disjoint events, the number of occurrences of $A \cup B$ in n performances of the experiment is obtained by adding the number of occurrences of A to the number of occurrences of B. Thus we should have

$$P(A \cup B) = P(A) + P(B) \qquad \text{if } A \text{ and } B \text{ are disjoint}$$

and, similarly,

$$P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) \qquad \text{if } A_1, \ldots, A_n \text{ are disjoint}$$

For mathematical convenience we require that

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

when we have a *countably infinite* family of disjoint events $A_1, A_2, \ldots$.

The assumption of countable rather than simply finite additivity has not been convincingly justified physically or philosophically; however, it leads to a much richer mathematical theory.

A function that assigns a number $P(A)$ to each set A in the sigma field $\mathscr{F}$ is called a *probability measure* on $\mathscr{F}$, provided that the following conditions are satisfied.

$$P(A) \geq 0 \qquad \text{for every } A \in \mathscr{F} \tag{1.3.4}$$

$$P(\Omega) = 1 \tag{1.3.5}$$

If $A_1, A_2, \ldots$ are disjoint sets in $\mathscr{F}$, then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots \tag{1.3.6}$$

We may now give the underlying mathematical framework for probability theory.

DEFINITION. A *probability space* is a triple $(\Omega, \mathscr{F}, P)$, where Ω is a set, $\mathscr{F}$ a sigma field of subsets of Ω, and P a probability measure on $\mathscr{F}$.

We shall not, at this point, embark on a general study of probability measures. However, we shall establish four facts from the definition. (All sets in the arguments to follow are assumed to belong to $\mathscr{F}$.)

$$1. \quad P(\varnothing) = 0 \tag{1.3.7}$$

PROOF. $A \cup \varnothing = A$; hence $P(A \cup \varnothing) = P(A)$. But A and $\varnothing$ are disjoint and so $P(A \cup \varnothing) = P(A) + P(\varnothing)$. Thus $P(A) = P(A) + P(\varnothing)$; consequently $P(\varnothing) = 0$.

$$2. \quad P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1.3.8}$$

PROOF. $A = (A \cap B) \cup (A \cap B^c)$, and these sets are disjoint (see Figure 1.2.4). Thus $P(A) = P(A \cap B) + P(A \cap B^c)$. Similarly $P(B) = P(A \cap B) + P(A^c \cap B)$. Thus $P(A) + P(B) - P(A \cap B) = P(A \cap B) + P(A \cap B^c) + P(A^c \cap B) = P(A \cup B)$. Intuitively, if we add the outcomes in A to those in B, we have counted those in $A \cap B$ twice; subtracting the outcomes in $A \cap B$ yields the outcomes in $A \cup B$.

3. If $B \subset A$, then $P(B) \leq P(A)$; in fact,

$$P(A - B) = P(A) - P(B) \tag{1.3.9}$$

where $A - B$ is the set of points that belong to A but not to B.

PROOF. $P(A) = P(B) + P(A - B)$, since $B \subset A$ (see Figure 1.3.1), and the result follows because $P(A - B) \geq 0$. Intuitively, if the occurrence of B

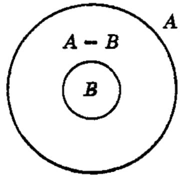

FIGURE 1.3.1

always implies the occurrence of A, A must occur at least as often as B in any sequence of performances of the experiment.

$$4.\quad P(A_1 \cup A_2 \cup \cdots) \leq P(A_1) + P(A_2) + \cdots \tag{1.3.10}$$

That is, the probability that at least one of a finite or countably infinite collection of events will occur is less than or equal to the sum of the probabilities; note that, for the case of two events, this follows from $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$.

PROOF. We make use of the fact that any union may be written as a disjoint union, as follows.

$$A_1 \cup A_2 \cup \cdots = A_1 \cup (A_1^c \cap A_2) \cup (A_1^c \cap A_2^c \cap A_3) \cup \cdots \cup (A_1^c \cap A_2^c \cap \cdots \cap A_{n-1}^c \cap A_n) \cup \cdots \tag{1.3.11}$$

To see this, observe that if x belongs to the set on the right then $x \in A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n$ for some n; hence $x \in A_n$. Thus x belongs to the set on the left. Conversely, if x belongs to the set on the left, then $x \in A_n$ for some n. Let n_0 be the smallest such n. Then $x \in A_1^c \cap \cdots \cap A_{n_0-1}^c \cap A_{n_0}$, and so x belongs to the set on the right. Thus

$$P(A_1 \cup A_2 \cup \cdots) = \sum_{n=1}^{\infty} P(A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n) \leq \sum_{n=1}^{\infty} P(A_n)$$

using (1.3.9); notice that

$$A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n \subset A_n.$$

REMARKS. The basic difficulty with the classical and frequency definitions of probability is that their approach is to try somehow to *prove* mathematically that, for example, the probability of picking a heart from a perfectly shuffled deck is 1/4, or that the probability of an unbiased coin coming up heads is 1/2. This cannot be done. All we can say is that if a card is picked at random and then replaced, and the process is repeated over and over again, the result that the ratio of hearts to total number of drawings will be close to 1/4 is in accord with our intuition and our physical experience. For this reason we should *assign* a probability 1/4 to the event of obtaining a heart, and similarly we should *assign* a probability 1/52 to each possible outcome of the experiment. The only reason for doing this is that the consequences agree with our experience. If you decide that some mysterious factor caused the ace of spades to be more likely than any other card, you could incorporate this factor by assigning a higher probability to the ace of spades. The mathematical development of the theory would not be affected; however, the conclusions you might draw from this assumption would be at variance with experimental results.

One can never really use mathematics to *prove* a specific physical fact. For example, we cannot prove mathematically that there is a physical quantity called "force." What we can do is postulate a mathematical entity called "force" that satisfies a certain differential equation. We can build up a collection of mathematical results that, when interpreted properly, provide a reasonable description of certain physical phenomena (reasonable until another mathematical theory is constructed that provides a better description). Similarly, in probability theory we are faced with situations in which our intuition or some physical experiments we have carried out suggest certain results. Intuition and experience lead us to an *assignment* of probabilities to events. As far as the mathematics is concerned, any assignment of probabilities will do, subject to the rules of mathematical consistency. However, our hope is to develop mathematical results that, when interpreted and related to physical experience, will help to make precise such notions as "the ratio of the number of heads to the total number of observations in a very large number of independent tosses of an unbiased coin is very likely to be very close to 1/2."

We emphasize that the insights gained by the early workers in probability are not to be discarded, but instead cast in a more precise form.

PROBLEMS

1. Write down some examples of sigma fields other than the collection of all subsets of a given set Ω.
2. Give an example to show that $P(A - B)$ need not equal $P(A) - P(B)$ if B is not a subset of A.