Preface

Probability and statistics are fascinating subjects on the interface between mathematics and applied sciences that help us understand and solve practical problems. We believe that you, by learning how stochastic methods come about and why they work, will be able to understand the meaning of statistical statements as well as judge the quality of their content, when facing such problems on your own. Our philosophy is one of *how* and *why*: instead of just presenting stochastic methods as cookbook recipes, we prefer to explain the principles behind them.

In this book you will find the basics of probability theory and statistics. In addition, there are several topics that go somewhat beyond the basics but that ought to be present in an introductory course: simulation, the Poisson process, the law of large numbers, and the central limit theorem. Computers have brought many changes in statistics. In particular, the bootstrap has earned its place. It provides the possibility to derive confidence intervals and perform tests of hypotheses where traditional (normal approximation or large sample) methods are inappropriate. It is a modern useful tool one should learn about, we believe.

Examples and datasets in this book are mostly from real-life situations, at least that is what we looked for in illustrations of the material. Anybody who has inspected datasets with the purpose of using them as elementary examples knows that this is hard: on the one hand, you do not want to boldly state assumptions that are clearly not satisfied; on the other hand, long explanations concerning side issues distract from the main points. We hope that we found a good middle way.

A first course in calculus is needed as a prerequisite for this book. In addition to high-school algebra, some infinite series are used (exponential, geometric). Integration and differentiation are the most important skills, mainly concerning one variable (the exceptions, two dimensional integrals, are encountered in Chapters 9–11). Although the mathematics is kept to a minimum, we strived to be mathematically correct throughout the book. With respect to probability and statistics the book is self-contained.

The book is aimed at undergraduate engineering students, and students from more business-oriented studies (who may gloss over some of the more mathematically oriented parts). At our own university we also use it for students in applied mathematics (where we put a little more emphasis on the math and add topics like combinatorics, conditional expectations, and generating functions). It is designed for a one-semester course: on average two hours in class per chapter, the first for a lecture, the second doing exercises. The material is also well-suited for self-study, as we know from experience.

We have divided attention about evenly between probability and statistics. The very first chapter is a sampler with differently flavored introductory examples, ranging from scientific success stories to a controversial puzzle. Topics that follow are elementary probability theory, simulation, joint distributions, the law of large numbers, the central limit theorem, statistical modeling (informal: why and how we can draw inference from data), data analysis, the bootstrap, estimation, simple linear regression, confidence intervals, and hypothesis testing. Instead of a few chapters with a long list of discrete and continuous distributions, with an enumeration of the important attributes of each, we introduce a few distributions when presenting the concepts and the others where they arise (more) naturally. A list of distributions and their characteristics is found in Appendix A.

With the exception of the first one, chapters in this book consist of three main parts. First, about four sections discussing new material, interspersed with a handful of so-called Quick exercises. Working these—two-or-three-minute—exercises should help to master the material and provide a break from reading to do something more active. On about two dozen occasions you will find indented paragraphs labeled *Remark*, where we felt the need to discuss more mathematical details or background material. These remarks can be skipped without loss of continuity; in most cases they require a bit more mathematical maturity. Whenever persons are introduced in examples we have determined their sex by looking at the chapter number and applying the rule "He is odd, she is even." Solutions to the quick exercises are found in the second to last section of each chapter.

The last section of each chapter is devoted to exercises, on average thirteen per chapter. For about half of the exercises, answers are given in Appendix C, and for half of these, full solutions in Appendix D. Exercises with both a short answer and a full solution are marked with ⊞ and those with only a short answer are marked with ⊡ (when more appropriate, for example, in "Show that ..." exercises, the short answer provides a hint to the key step). Typically, the section starts with some easy exercises and the order of the material in the chapter is more or less respected. More challenging exercises are found at the end.

Much of the material in this book would benefit from illustration with a computer using statistical software. A complete course should also involve computer exercises. Topics like simulation, the law of large numbers, the central limit theorem, and the bootstrap loudly call for this kind of experience. For this purpose, all the datasets discussed in the book are available at http://www.springeronline.com/1-85233-896-2. The same Web site also provides access, for instructors, to a complete set of solutions to the exercises; go to the Springer online catalog or contact textbooks@springer-sbm.com to apply for your password.

Delft, The Netherlands
January 2005

F. M. Dekking
C. Kraaikamp
H. P. Lopuhaä
L. E. Meester

Outcomes, events, and probability

The world around us is full of phenomena we perceive as random or unpredictable. We aim to model these phenomena as *outcomes* of some experiment, where you should think of *experiment* in a very general sense. The outcomes are elements of a *sample space* Ω, and subsets of Ω are called *events.*The events will be assigned a *probability*, a number between 0 and 1 that expresses how likely the event is to occur.

2.1 Sample spaces

Sample spaces are simply sets whose elements describe the outcomes of the experiment in which we are interested.

We start with the most basic experiment: the tossing of a coin. Assuming that we will never see the coin land on its rim, there are two possible outcomes: heads and tails. We therefore take as the sample space associated with this experiment the set $\Omega = \{H, T\}$.

In another experiment we ask the next person we meet on the street in which month her birthday falls. An obvious choice for the sample space is

$$\Omega = \{\text{Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec}\}.$$

In a third experiment we load a scale model for a bridge up to the point where the structure collapses. The outcome is the load at which this occurs. In reality, one can only measure with finite accuracy, e.g., to five decimals, and a sample space with just those numbers would strictly be adequate. However, in principle, the load itself could be any positive number and therefore $\Omega = (0, \infty)$ is the right choice. Even though in reality there may also be an upper limit to what loads are conceivable, it is not necessary or practical to try to limit the outcomes correspondingly.

In a fourth experiment, we find on our doormat three envelopes, sent to us by three different persons, and we look in which order the envelopes lie on top of each other. Coding them 1, 2, and 3, the sample space would be

$$\Omega = \{123, 132, 213, 231, 312, 321\}.$$

QUICK EXERCISE 2.1 If we received mail from four different persons, how many elements would the corresponding sample space have?

In general one might consider the order in which n different objects can be placed. This is called a *permutation* of the n objects. As we have seen, there are 6 possible permutations of 3 objects, and $4 \cdot 6 = 24$ of 4 objects. What happens is that if we add the nth object, then this can be placed in any of n positions in any of the permutations of $n-1$ objects. Therefore there are

$$n \cdot (n-1) \cdot \dots \cdot 3 \cdot 2 \cdot 1 = n!$$

possible permutations of n objects. Here $n!$ is the standard notation for this product and is pronounced "n factorial." It is convenient to define $0! = 1$.

2.2 Events

Subsets of the sample space are called *events*. We say that an event A *occurs* if the outcome of the experiment is an element of the set A. For example, in the birthday experiment we can ask for the outcomes that correspond to a long month, i.e., a month with 31 days. This is the event

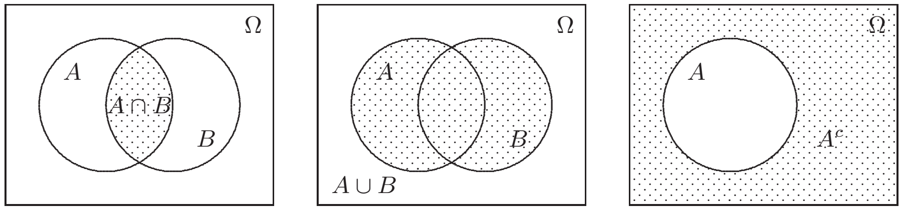
$$L = \{\text{Jan, Mar, May, Jul, Aug, Oct, Dec}\}.$$

Events may be combined according to the usual set operations.

For example if R is the event that corresponds to the months that have the letter r in their (full) name (so $R = \{\text{Jan}, \text{Feb}, \text{Mar}, \text{Apr}, \text{Sep}, \text{Oct}, \text{Nov}, \text{Dec}\}$), then the long months that contain the letter r are

$$L \cap R = \{\text{Jan}, \text{Mar}, \text{Oct}, \text{Dec}\}.$$

The set $L \cap R$ is called the *intersection* of L and R and occurs if both L and R occur. Similarly, we have the *union* $A \cup B$ of two sets A and B, which occurs if at least one of the events A and B occurs. Another common operation is taking complements. The event $A^c = \{\omega \in \Omega : \omega \notin A\}$ is called the *complement* of A; it occurs if and only if A does *not* occur. The complement of Ω is denoted $\emptyset$, the empty set, which represents the impossible event. Figure 2.1 illustrates these three set operations.



Intersection $A \cap B$ Union $A \cup B$ Complement A^c

Fig. 2.1. Diagrams of intersection, union, and complement.

We call events A and B *disjoint* or *mutually exclusive* if A and B have no outcomes in common; in set terminology: $A \cap B = \emptyset$. For example, the event L "the birthday falls in a long month" and the event {Feb} are disjoint.

Finally, we say that event A *implies* event B if the outcomes of A also lie in B. In set notation: $A \subset B$; see Figure 2.2.

Some people like to use double negations:

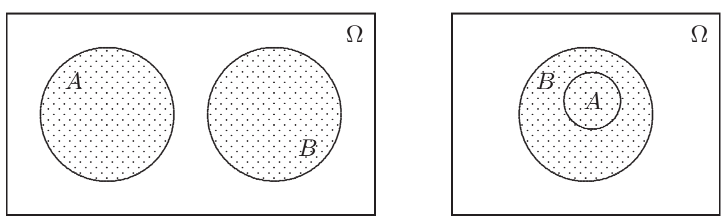
> "It is certainly not true that neither John nor Mary is to blame."

This is equivalent to: "John or Mary is to blame, or both." The following useful rules formalize this mental operation to a manipulation with events.

DEMORGAN'S LAWS. For any two events A and B we have

$$(A \cup B)^c = A^c \cap B^c \text{ and } (A \cap B)^c = A^c \cup B^c.$$

QUICK EXERCISE 2.2 Let J be the event "John is to blame" and M the event "Mary is to blame." Express the two statements above in terms of the events J, J^c, M, and M^c, and check the equivalence of the statements by means of DeMorgan's laws.



Disjoint sets A and B A subset of B

Fig. 2.2. Minimal and maximal intersection of two sets.

2.3 Probability

We want to express how likely it is that an event occurs. To do this we will assign a probability to each event. The assignment of probabilities to events is in general not an easy task, and some of the coming chapters will be dedicated directly or indirectly to this problem. Since *each* event has to be assigned a probability, we speak of a probability *function*. It has to satisfy two basic properties.

DEFINITION. A *probability function* P on a finite sample space Ω assigns to each event A in Ω a number $\mathrm{P}(A)$ in [0,1] such that
(i) $\mathrm{P}(\Omega) = 1$, and
(ii) $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$ if A and B are disjoint.
The number $\mathrm{P}(A)$ is called the probability that A occurs.

Property (i) expresses that the outcome of the experiment is always an element of the sample space, and property (ii) is the additivity property of a probability function. It implies additivity of the probability function over more than two sets; e.g., if A, B, and C are disjoint events, then the two events $A \cup B$ and C are also disjoint, so

$$\mathrm{P}(A \cup B \cup C) = \mathrm{P}(A \cup B) + \mathrm{P}(C) = \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) .$$

We will now look at some examples. When we want to decide whether Peter or Paul has to wash the dishes, we might toss a coin. The fact that we consider this a fair way to decide translates into the opinion that heads and tails are equally likely to occur as the outcome of the coin-tossing experiment. So we put

$$\mathrm{P}(\{H\}) = \mathrm{P}(\{T\}) = \frac{1}{2}.$$

Formally we have to write $\{H\}$ for the set consisting of the single element H, because a probability function is defined on *events*, not on outcomes. From now on we shall drop these brackets.

Now it might happen, for example due to an asymmetric distribution of the mass over the coin, that the coin is not completely fair. For example, it might be the case that

$$\mathrm{P}(H) = 0.4999 \text{ and } \mathrm{P}(T) = 0.5001.$$

More generally we can consider experiments with two possible outcomes, say "failure" and "success", which have probabilities $1 - p$ and p to occur, where p is a number between 0 and 1. For example, when our experiment consists of buying a ticket in a lottery with 10 000 tickets and only one prize, where "success" stands for winning the prize, then $p = 10^{-4}$.

How should we assign probabilities in the second experiment, where we ask for the month in which the next person we meet has his or her birthday? In analogy with what we have just done, we put

$$\mathrm{P}(\mathrm{Jan}) = \mathrm{P}(\mathrm{Feb}) = \cdots = \mathrm{P}(\mathrm{Dec}) = \frac{1}{12}.$$

Some of you might object to this and propose that we put, for example,

$$\mathrm{P}(\mathrm{Jan}) = \frac{31}{365} \quad \text{and} \quad \mathrm{P}(\mathrm{Apr}) = \frac{30}{365},$$

because we have long months and short months. But then the very precise among us might remark that this does not yet take care of leap years.

Quick exercise 2.3 If you would take care of the leap years, assuming that one in every four years is a leap year (which again is an approximation to reality!), how would you assign a probability to each month?

In the third experiment (the buckling load of a bridge), where the outcomes are real numbers, it is impossible to assign a positive probability to each outcome (there are just too many outcomes!). We shall come back to this problem in Chapter 5, restricting ourselves in this chapter to finite and countably infinite[1] sample spaces.

In the fourth experiment it makes sense to assign equal probabilities to all six outcomes:

$$\mathrm{P}(123) = \mathrm{P}(132) = \mathrm{P}(213) = \mathrm{P}(231) = \mathrm{P}(312) = \mathrm{P}(321) = \frac{1}{6}.$$

Until now we have only assigned probabilities to the individual outcomes of the experiments. To assign probabilities to events we use the additivity property. For instance, to find the probability $\mathrm{P}(T)$ of the event T that in the three envelopes experiment envelope 2 is on top we note that

$$\mathrm{P}(T) = \mathrm{P}(213) + \mathrm{P}(231) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

In general, additivity of P implies that the probability of an event is obtained by summing the probabilities of the outcomes belonging to the event.

Quick exercise 2.4 Compute $\mathrm{P}(L)$ and $\mathrm{P}(R)$ in the birthday experiment.

Finally we mention a rule that permits us to compute probabilities of events A and B that are *not* disjoint. Note that we can write $A = (A \cap B) \cup (A \cap B^c)$, which is a disjoint union; hence

$$\mathrm{P}(A) = \mathrm{P}(A \cap B) + \mathrm{P}(A \cap B^c)\,.$$

If we split $A \cup B$ in the same way with B and B^c, we obtain the events $(A \cup B) \cap B$, which is simply B and $(A \cup B) \cap B^c$, which is nothing but $A \cap B^c$.

[1] This means: although infinite, we can still count them one by one; $\Omega = \{\omega_1, \omega_2, \ldots\}$. The interval [0,1] of real numbers is an example of an uncountable sample space.

Thus

$$P(A \cup B) = P(B) + P(A \cap B^c)\,.$$

Eliminating $P(A \cap B^c)$ from these two equations we obtain the following rule.

THE PROBABILITY OF A UNION. For any two events A and B we have

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

From the additivity property we can also find a way to compute probabilities of complements of events: from $A \cup A^c = \Omega$, we deduce that

$$P(A^c) = 1 - P(A)\,.$$

2.4 Products of sample spaces

Basic to statistics is that one usually does not consider *one* experiment, but that the same experiment is performed several times. For example, suppose we throw a coin two times. What is the sample space associated with this new experiment? It is clear that it should be the set

$$\Omega = \{H, T\} \times \{H, T\} = \{(H, H), (H, T), (T, H), (T, T)\}.$$

If in the original experiment we had a fair coin, i.e., $P(H) = P(T)$, then in this new experiment all 4 outcomes again have equal probabilities:

$$P((H, H)) = P((H, T)) = P((T, H)) = P((T, T)) = \frac{1}{4}.$$

Somewhat more generally, if we consider two experiments with sample spaces Ω_1 and Ω_2 then the combined experiment has as its sample space the set

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}.$$

If Ω_1 has r elements and Ω_2 has s elements, then $\Omega_1 \times \Omega_2$ has rs elements. Now suppose that in the first, the second, and the combined experiment all outcomes are equally likely to occur. Then the outcomes in the first experiment have probability $1/r$ to occur, those of the second experiment $1/s$, and those of the combined experiment probability $1/rs$. Motivated by the fact that $1/rs = (1/r) \times (1/s)$, we will assign probability $p_i p_j$ to the outcome (ω_i, ω_j) in the combined experiment, in the case that ω_i has probability p_i and ω_j has probability p_j to occur. One should realize that this is by no means the only way to assign probabilities to the outcomes of a combined experiment. The preceding choice corresponds to the situation where the two experiments do not influence each other in any way. What we mean by this influence will be explained in more detail in the next chapter.

Quick exercise 2.5 Consider the sample space $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ of some experiment, where outcome a_i has probability p_i for $i = 1, \ldots, 6$. We perform this experiment twice in such a way that the associated probabilities are

$$\mathrm{P}((a_i, a_i)) = p_i, \quad \text{and} \quad \mathrm{P}((a_i, a_j)) = 0 \text{ if } i \neq j, \quad \text{for } i, j = 1, \ldots, 6.$$

Check that P is a probability function on the sample space $\Omega = \{a_1, \ldots, a_6\} \times \{a_1, \ldots, a_6\}$ of the combined experiment. What is the relationship between the first experiment and the second experiment that is determined by this probability function?

We started this section with the experiment of throwing a coin twice. If we want to learn more about the randomness associated with a particular experiment, then we should repeat it more often, say n times. For example, if we perform an experiment with outcomes 1 (success) and 0 (failure) five times, and we consider the event A "exactly one experiment was a success," then this event is given by the set

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

in $\Omega = \{0,1\} \times \{0,1\} \times \{0,1\} \times \{0,1\} \times \{0,1\}$. Moreover, if success has probability p and failure probability $1-p$, then

$$\mathrm{P}(A) = 5 \cdot (1-p)^4 \cdot p,$$

since there are five outcomes in the event A, each having probability $(1-p)^4 \cdot p$.

Quick exercise 2.6 What is the probability of the event B "exactly two experiments were successful"?

In general, when we perform an experiment n times, then the corresponding sample space is

$$\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n,$$

where Ω_i for $i = 1, \ldots, n$ is a copy of the sample space of the original experiment. Moreover, we assign probabilities to the outcomes $(\omega_1, \ldots, \omega_n)$ in the standard way described earlier, i.e.,

$$\mathrm{P}((\omega_1, \omega_2, \ldots, \omega_n)) = p_1 \cdot p_2 \cdots p_n,$$

if each ω_i has probability p_i.

2.5 An infinite sample space

We end this chapter with an example of an experiment with infinitely many outcomes. We toss a coin repeatedly until the first head turns up. The outcome of the experiment is the number of tosses it takes to have this first occurrence of a head. Our sample space is the space of all positive natural numbers

$$\Omega = \{1, 2, 3, \ldots\}.$$

What is the probability function P for this experiment?

Suppose the coin has probability p of falling on heads and probability $1-p$ to fall on tails, where $0 < p < 1$. We determine the probability $\mathrm{P}(n)$ for each n. Clearly $\mathrm{P}(1) = p$, the probability that we have a head right away. The event $\{2\}$ corresponds to the outcome (T, H) in $\{H, T\} \times \{H, T\}$, so we should have

$$\mathrm{P}(2) = (1-p)p.$$

Similarly, the event $\{n\}$ corresponds to the outcome $(T, T, \ldots, T, T, H)$ in the space $\{H, T\} \times \cdots \times \{H, T\}$. Hence we should have, in general,

$$\mathrm{P}(n) = (1-p)^{n-1}p, \qquad n = 1, 2, 3, \ldots.$$

Does this define a probability function on $\Omega = \{1, 2, 3, \ldots\}$? Then we should at least have $\mathrm{P}(\Omega) = 1$. It is not directly clear how to calculate $\mathrm{P}(\Omega)$: since the sample space is no longer finite we have to amend the definition of a probability function.

> DEFINITION. A *probability function* on an infinite (or finite) sample space Ω assigns to each event A in Ω a number $\mathrm{P}(A)$ in $[0, 1]$ such that
> (i) $\mathrm{P}(\Omega) = 1$, and
> (ii) $\mathrm{P}(A_1 \cup A_2 \cup A_3 \cup \cdots) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) + \cdots$
> if $A_1, A_2, A_3, \ldots$ are disjoint events.

Note that this new additivity property is an extension of the previous one because if we choose $A_3 = A_4 = \cdots = \emptyset$, then

$$\begin{aligned}\mathrm{P}(A_1 \cup A_2) &= \mathrm{P}(A_1 \cup A_2 \cup \emptyset \cup \emptyset \cup \cdots) \\ &= \mathrm{P}(A_1) + \mathrm{P}(A_2) + 0 + 0 + \cdots = \mathrm{P}(A_1) + \mathrm{P}(A_2)\,.\end{aligned}$$

Now we can compute the probability of Ω:

$$\begin{aligned}\mathrm{P}(\Omega) &= \mathrm{P}(1) + \mathrm{P}(2) + \cdots + \mathrm{P}(n) + \cdots \\ &= p + (1-p)p + \cdots (1-p)^{n-1}p + \cdots \\ &= p[1 + (1-p) + \cdots (1-p)^{n-1} + \cdots].\end{aligned}$$

The sum $1 + (1-p) + \cdots + (1-p)^{n-1} + \cdots$ is an example of a *geometric series*. It is well known that when $|1-p| < 1$,

$$1 + (1-p) + \cdots + (1-p)^{n-1} + \cdots = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

Therefore we do indeed have $\mathrm{P}(\Omega) = p \cdot \dfrac{1}{p} = 1$.

QUICK EXERCISE 2.7 Suppose an experiment in a laboratory is repeated every day of the week until it is successful, the probability of success being p. The first experiment is started on a Monday. What is the probability that the series ends on the next Sunday?

2.6 Solutions to the quick exercises

2.1 The sample space is $\Omega = \{1234, 1243, 1324, 1342, \ldots, 4321\}$. The best way to count its elements is by noting that for *each* of the 6 outcomes of the three-envelope experiment we can put a fourth envelope in any of 4 positions. Hence Ω has $4 \cdot 6 = 24$ elements.

2.2 The statement "It is certainly not true that neither John nor Mary is to blame" corresponds to the event $(J^c \cap M^c)^c$. The statement "John or Mary is to blame, or both" corresponds to the event $J \cup M$. Equivalence now follows from DeMorgan's laws.

2.3 In four years we have $365 \times 3 + 366 = 1461$ days. Hence long months each have a probability $4 \times 31/1461 = 124/1461$, and short months a probability $120/1461$ to occur. Moreover, $\{\text{Feb}\}$ has probability $113/1461$.

2.4 Since there are 7 long months and 8 months with an "r" in their name, we have $\mathrm{P}(L) = 7/12$ and $\mathrm{P}(R) = 8/12$.

2.5 Checking that P is a probability function Ω amounts to verifying that $0 \leq \mathrm{P}((a_i, a_j)) \leq 1$ for all i and j and noting that

$$\mathrm{P}(\Omega) = \sum_{i,j=1}^{6} \mathrm{P}((a_i, a_j)) = \sum_{i=1}^{6} \mathrm{P}((a_i, a_i)) = \sum_{i=1}^{6} p_i = 1.$$

The two experiments are *totally* coupled: one has outcome a_i if and only if the other has outcome a_i.

2.6 Now there are 10 outcomes in B (for example (0,1,0,1,0)), each having probability $(1-p)^3 p^2$. Hence $\mathrm{P}(B) = 10(1-p)^3 p^2$.

2.7 This happens if and only if the experiment fails on Monday,..., Saturday, and is a success on Sunday. This has probability $p(1-p)^6$ to happen.

2.7 Exercises

2.1 ⊡ Let A and B be two events in a sample space for which $\mathrm{P}(A) = 2/3$, $\mathrm{P}(B) = 1/6$, and $\mathrm{P}(A \cap B) = 1/9$. What is $\mathrm{P}(A \cup B)$?

2.2 Let E and F be two events for which one knows that the probability that at least one of them occurs is $3/4$. What is the probability that neither E nor F occurs? *Hint:* use one of DeMorgan's laws: $E^c \cap F^c = (E \cup F)^c$.

2.3 Let C and D be two events for which one knows that $\mathrm{P}(C) = 0.3, \mathrm{P}(D) = 0.4$, and $\mathrm{P}(C \cap D) = 0.2$. What is $\mathrm{P}(C^c \cap D)$?

2.4 ⊡ We consider events A, B, and C, which can occur in some experiment. Is it true that the probability that *only* A occurs (and not B or C) is equal to $\mathrm{P}(A \cup B \cup C) - \mathrm{P}(B) - \mathrm{P}(C) + \mathrm{P}(B \cap C)$?

2.5 The event $A \cap B^c$ that A occurs but not B is sometimes denoted as $A \setminus B$. Here $\setminus$ is the set-theoretic minus sign. Show that $\mathrm{P}(A \setminus B) = \mathrm{P}(A) - \mathrm{P}(B)$ if B implies A, i.e., if $B \subset A$.

2.6 When $\mathrm{P}(A) = 1/3$, $\mathrm{P}(B) = 1/2$, and $\mathrm{P}(A \cup B) = 3/4$, what is

a. $\mathrm{P}(A \cap B)$?
b. $\mathrm{P}(A^c \cup B^c)$?

2.7 ⊡ Let A and B be two events. Suppose that $\mathrm{P}(A) = 0.4$, $\mathrm{P}(B) = 0.5$, and $\mathrm{P}(A \cap B) = 0.1$. Find the probability that A or B occurs, but not both.

2.8 ⊞ Suppose the events D_1 and D_2 represent disasters, which are rare: $\mathrm{P}(D_1) \leq 10^{-6}$ *and* $\mathrm{P}(D_2) \leq 10^{-6}$. What can you say about the probability that at least one of the disasters occurs? What about the probability that they *both* occur?

2.9 We toss a coin three times. For this experiment we choose the sample space

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

where T stands for tails and H for heads.

a. Write down the set of outcomes corresponding to each of the following events:

A : "we throw tails exactly two times."
B : "we throw tails at least two times."
C : "tails did not appear *before* a head appeared."
D : "the first throw results in tails."

b. Write down the set of outcomes corresponding to each of the following events: A^c, $A \cup (C \cap D)$, and $A \cap D^c$.

2.10 In some sample space we consider two events A and B. Let C be the event that A or B occurs, but not both. Express C in terms of A and B, using only the basic operations "union," "intersection," and "complement."

2.11 ⊡ An experiment has only two outcomes. The first has probability p to occur, the second probability p^2. What is p?

2.12 ⊞ In the UEFA Euro 2004 playoffs draw 10 national football teams were matched in pairs. A lot of people complained that "the draw was not fair," because each strong team had been matched with a weak team (this is commercially the most interesting). It was claimed that such a matching is extremely unlikely. We will compute the probability of this "dream draw" in this exercise. In the spirit of the three-envelope example of Section 2.1 we put the names of the 5 strong teams in envelopes labeled $1, 2, 3, 4$, and 5 and of the 5 weak teams in envelopes labeled $6, 7, 8, 9$, and 10. We shuffle the 10 envelopes and then match the envelope on top with the next envelope, the third envelope with the fourth envelope, and so on. One particular way a "dream draw" occurs is when the five envelopes labeled $1, 2, 3, 4, 5$ are in the odd numbered positions (in any order!) and the others are in the even numbered positions. This way corresponds to the situation where the first match of each strong team is a home match. Since for each pair there are two possibilities for the home match, the total number of possibilities for the "dream draw" is $2^5 = 32$ times as large.

a. An outcome of this experiment is a sequence like $4, 9, 3, 7, 5, 10, 1, 8, 2, 6$ of labels of envelopes. What is the probability of an outcome?

b. How many outcomes are there in the event "the five envelopes labeled $1, 2, 3, 4, 5$ are in the odd positions—in any order, and the envelopes labeled $6, 7, 8, 9, 10$ are in the even positions—in any order"?

c. What is the probability of a "dream draw"?

2.13 In some experiment first an arbitrary choice is made out of four possibilities, and then an arbitrary choice is made out of the remaining three possibilities. One way to describe this is with a product of two sample spaces $\{a, b, c, d\}$:

$$\Omega = \{a, b, c, d\} \times \{a, b, c, d\}.$$

a. Make a 4×4 table in which you write the probabilities of the outcomes.

b. Describe the event "c is one of the chosen possibilities" and determine its probability.

2.14 ⊞ Consider the Monty Hall "experiment" described in Section 1.3. The door behind which the car is parked we label a, the other two b and c. As the sample space we choose a product space

$$\Omega = \{a, b, c\} \times \{a, b, c\}.$$

Here the first entry gives the choice of the candidate, and the second entry the choice of the quizmaster.

a. Make a 3×3 table in which you write the probabilities of the outcomes. *N.B.* You should realize that the candidate does *not know* that the car is in a, but the quizmaster will never open the door labeled a because he *knows* that the car is there. You may assume that the quizmaster makes an arbitrary choice between the doors labeled b and c, when the candidate chooses door a.

b. Consider the situation of a "no switching" candidate who will stick to his or her choice. What is the event "the candidate wins the car," and what is its probability?

c. Consider the situation of a "switching" candidate who will not stick to her choice. What is now the event "the candidate wins the car," and what is its probability?

2.15 The rule $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$ from Section 2.3 is often useful to compute the probability of the union of two events. What would be the corresponding rule for three events A, B, and C? It should start with

$$\mathrm{P}(A \cup B \cup C) = \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) - \cdots .$$

Hint: you could use the sum rule suitably, or you could make a diagram as in Figure 2.1.

2.16 ⊞ Three events E, F, and G cannot occur simultaneously. Further it is known that $\mathrm{P}(E \cap F) = \mathrm{P}(F \cap G) = \mathrm{P}(E \cap G) = 1/3$. Can you determine $\mathrm{P}(E)$?
Hint: if you try to use the formula of Exercise 2.15 then it seems that you do not have enough information; make a diagram instead.

2.17 A post office has two counters where customers can buy stamps, etc. If you are interested in the number of customers in the two queues that will form for the counters, what would you take as sample space?

2.18 In a laboratory, two experiments are repeated every day of the week in different rooms until at least one is successful, the probability of success being p for each experiment. Supposing that the experiments in different rooms and on different days are performed independently of each other, what is the probability that the laboratory scores its first successful experiment on day n?

2.19 ⊡ We repeatedly toss a coin. A head has probability p, and a tail probability $1 - p$ to occur, where $0 < p < 1$. The outcome of the experiment we are interested in is the number of tosses it takes until a head occurs for the *second* time.

a. What would you choose as the sample space?

b. What is the probability that it takes 5 tosses?