

---

# Chapter

# 2

## CONDITIONAL PROBABILITY

---

- 2.1 The Definition of Conditional Probability
- 2.2 Independent Events
- 2.3 Bayes' Theorem

- 2.4 The Gambler's Ruin Problem
- 2.5 Supplementary Exercises

### 2.1 The Definition of Conditional Probability

*A major use of probability in statistical inference is the updating of probabilities when certain events are observed. The updated probability of event  $A$  after we learn that event  $B$  has occurred is the conditional probability of  $A$  given  $B$ .*

**Example**  
**2.1.1**

**Lottery Ticket.** Consider a state lottery game in which six numbers are drawn without replacement from a bin containing the numbers 1–30. Each player tries to match the set of six numbers that will be drawn without regard to the order in which the numbers are drawn. Suppose that you hold a ticket in such a lottery with the numbers 1, 14, 15, 20, 23, and 27. You turn on your television to watch the drawing but all you see is one number, 15, being drawn when the power suddenly goes off in your house. You don't even know whether 15 was the first, last, or some in-between draw. However, now that you know that 15 appears in the winning draw, the probability that your ticket is a winner must be higher than it was before you saw the draw. How do you calculate the revised probability? ◀

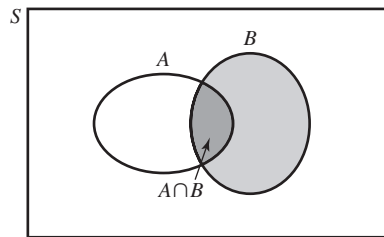
Example 2.1.1 is typical of the following situation. An experiment is performed for which the sample space  $S$  is given (or can be constructed easily) and the probabilities are available for all of the events of interest. We then learn that some event  $B$  has occurred, and we want to know how the probability of another event  $A$  changes after we learn that  $B$  has occurred. In Example 2.1.1, the event that we have learned is  $B = \{\text{one of the numbers drawn is 15}\}$ . We are certainly interested in the probability of

$$A = \{\text{the numbers 1, 14, 15, 20, 23, and 27 are drawn}\},$$

and possibly other events.

If we know that the event  $B$  has occurred, then we know that the outcome of the experiment is one of those included in  $B$ . Hence, to evaluate the probability that  $A$  will occur, we must consider the set of those outcomes in  $B$  that also result in the occurrence of  $A$ . As sketched in Fig. 2.1, this set is precisely the set  $A \cap B$ . It is therefore natural to calculate the revised probability of  $A$  according to the following definition.

**Figure 2.1** The outcomes in the event  $B$  that also belong to the event  $A$ .



**Definition**  
**2.1.1**

**Conditional Probability.** Suppose that we learn that an event  $B$  has occurred and that we wish to compute the probability of another event  $A$  taking into account that we know that  $B$  has occurred. The new probability of  $A$  is called the *conditional probability of the event  $A$  given that the event  $B$  has occurred* and is denoted  $\Pr(A|B)$ . If  $\Pr(B) > 0$ , we compute this probability as

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}. \quad (2.1.1)$$

The conditional probability  $\Pr(A|B)$  is not defined if  $\Pr(B) = 0$ .

For convenience, the notation in Definition 2.1.1 is read simply as the conditional probability of  $A$  given  $B$ . Eq. (2.1.1) indicates that  $\Pr(A|B)$  is computed as the proportion of the total probability  $\Pr(B)$  that is represented by  $\Pr(A \cap B)$ , intuitively the proportion of  $B$  that is also part of  $A$ .

**Example**  
**2.1.2**

**Lottery Ticket.** In Example 2.1.1, you learned that the event

$$B = \{\text{one of the numbers drawn is 15}\}$$

has occurred. You want to calculate the probability of the event  $A$  that your ticket is a winner. Both events  $A$  and  $B$  are expressible in the sample space that consists of the  $\binom{30}{6} = 30!/(6!24!)$  possible combinations of 30 items taken six at a time, namely, the unordered draws of six numbers from 1–30. The event  $B$  consists of combinations that include 15. Since there are 29 remaining numbers from which to choose the other five in the winning draw, there are  $\binom{29}{5}$  outcomes in  $B$ . It follows that

$$\Pr(B) = \frac{\binom{29}{5}}{\binom{30}{6}} = \frac{29!24!6!}{30!5!24!} = 0.2.$$

The event  $A$  that your ticket is a winner consists of a single outcome that is also in  $B$ , so  $A \cap B = A$ , and

$$\Pr(A \cap B) = \Pr(A) = \frac{1}{\binom{30}{6}} = \frac{6!24!}{30!} = 1.68 \times 10^{-6}.$$

It follows that the conditional probability of  $A$  given  $B$  is

$$\Pr(A|B) = \frac{\frac{6!24!}{30!}}{0.2} = 8.4 \times 10^{-6}.$$

This is five times as large as  $\Pr(A)$  before you learned that  $B$  had occurred. ◀

Definition 2.1.1 for the conditional probability  $\Pr(A|B)$  is worded in terms of the subjective interpretation of probability in Sec. 1.2. Eq. (2.1.1) also has a simple meaning in terms of the frequency interpretation of probability. According to the

frequency interpretation, if an experimental process is repeated a large number of times, then the proportion of repetitions in which the event  $B$  will occur is approximately  $\Pr(B)$  and the proportion of repetitions in which both the event  $A$  and the event  $B$  will occur is approximately  $\Pr(A \cap B)$ . Therefore, among those repetitions in which the event  $B$  occurs, the proportion of repetitions in which the event  $A$  will also occur is approximately equal to

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

**Example  
2.1.3**

**Rolling Dice.** Suppose that two dice were rolled and it was observed that the sum  $T$  of the two numbers was odd. We shall determine the probability that  $T$  was less than 8.

If we let  $A$  be the event that  $T < 8$  and let  $B$  be the event that  $T$  is odd, then  $A \cap B$  is the event that  $T$  is 3, 5, or 7. From the probabilities for two dice given at the end of Sec. 1.6, we can evaluate  $\Pr(A \cap B)$  and  $\Pr(B)$  as follows:

$$\begin{aligned}\Pr(A \cap B) &= \frac{2}{36} + \frac{4}{36} + \frac{6}{36} = \frac{12}{36} = \frac{1}{3}, \\ \Pr(B) &= \frac{2}{36} + \frac{4}{36} + \frac{6}{36} + \frac{4}{36} + \frac{2}{36} = \frac{18}{36} = \frac{1}{2}.\end{aligned}$$

Hence,

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{2}{3}.$$

**Example  
2.1.4**

**A Clinical Trial.** It is very common for patients with episodes of depression to have a recurrence within two to three years. Prien et al. (1984) studied three treatments for depression: imipramine, lithium carbonate, and a combination. As is traditional in such studies (called *clinical trials*), there was also a group of patients who received a placebo. (A placebo is a treatment that is supposed to be neither helpful nor harmful. Some patients are given a placebo so that they will not know that they did not receive one of the other treatments. None of the other patients knew which treatment or placebo they received, either.) In this example, we shall consider 150 patients who entered the study after an episode of depression that was classified as “unipolar” (meaning that there was no manic disorder). They were divided into the four groups (three treatments plus placebo) and followed to see how many had recurrences of depression. Table 2.1 summarizes the results. If a patient were selected at random from this study and it were found that the patient received the placebo treatment, what is the conditional probability that the patient had a relapse? Let  $B$  be the event that the patient received the placebo, and let  $A$  be the event that

**Table 2.1** Results of the clinical depression study in Example 2.1.4

Response	Treatment group			Placebo	Total
	Imipramine	Lithium	Combination		
Relapse	18	13	22	24	77
No relapse	22	25	16	10	73
Total	40	38	38	34	150

the patient had a relapse. We can calculate  $\Pr(B) = 34/150$  and  $\Pr(A \cap B) = 24/150$  directly from the table. Then  $\Pr(A|B) = 24/34 = 0.706$ . On the other hand, if the randomly selected patient is found to have received lithium (call this event  $C$ ) then  $\Pr(C) = 38/150$ ,  $\Pr(A \cap C) = 13/150$ , and  $\Pr(A|C) = 13/38 = 0.342$ . Knowing which treatment a patient received seems to make a difference to the probability of relapse. In Chapter 10, we shall study methods for being more precise about how much of a difference it makes. ◀

**Example**  
**2.1.5**

**Rolling Dice Repeatedly.** Suppose that two dice are to be rolled repeatedly and the sum  $T$  of the two numbers is to be observed for each roll. We shall determine the probability  $p$  that the value  $T = 7$  will be observed before the value  $T = 8$  is observed.

The desired probability  $p$  could be calculated directly as follows: We could assume that the sample space  $S$  contains all sequences of outcomes that terminate as soon as either the sum  $T = 7$  or the sum  $T = 8$  is obtained. Then we could find the sum of the probabilities of all the sequences that terminate when the value  $T = 7$  is obtained.

However, there is a simpler approach in this example. We can consider the simple experiment in which two dice are rolled. If we repeat the experiment until either the sum  $T = 7$  or the sum  $T = 8$  is obtained, the effect is to restrict the outcome of the experiment to one of these two values. Hence, the problem can be restated as follows: Given that the outcome of the experiment is either  $T = 7$  or  $T = 8$ , determine the probability  $p$  that the outcome is actually  $T = 7$ .

If we let  $A$  be the event that  $T = 7$  and let  $B$  be the event that the value of  $T$  is either 7 or 8, then  $A \cap B = A$  and

$$p = \Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A)}{\Pr(B)}.$$

From the probabilities for two dice given in Example 1.6.5,  $\Pr(A) = 6/36$  and  $\Pr(B) = (6/36) + (5/36) = 11/36$ . Hence,  $p = 6/11$ . ◀

## The Multiplication Rule for Conditional Probabilities

In some experiments, certain conditional probabilities are relatively easy to assign directly. In these experiments, it is then possible to compute the probability that both of two events occur by applying the next result that follows directly from Eq. (2.1.1) and the analogous definition of  $\Pr(B|A)$ .

**Theorem**  
**2.1.1**

**Multiplication Rule for Conditional Probabilities.** Let  $A$  and  $B$  be events. If  $\Pr(B) > 0$ , then

$$\Pr(A \cap B) = \Pr(B) \Pr(A|B).$$

If  $\Pr(A) > 0$ , then

$$\Pr(A \cap B) = \Pr(A) \Pr(B|A). \quad \blacksquare$$

**Example**  
**2.1.6**

**Selecting Two Balls.** Suppose that two balls are to be selected at random, without replacement, from a box containing  $r$  red balls and  $b$  blue balls. We shall determine the probability  $p$  that the first ball will be red and the second ball will be blue.

Let  $A$  be the event that the first ball is red, and let  $B$  be the event that the second ball is blue. Obviously,  $\Pr(A) = r/(r + b)$ . Furthermore, if the event  $A$  has occurred, then one red ball has been removed from the box on the first draw. Therefore, the

probability of obtaining a blue ball on the second draw will be

$$\Pr(B|A) = \frac{b}{r+b-1}.$$

It follows that

$$\Pr(A \cap B) = \frac{r}{r+b} \cdot \frac{b}{r+b-1}. \quad \blacktriangleleft$$

The principle that has just been applied can be extended to any finite number of events, as stated in the following theorem.

**Theorem**  
**2.1.2**

**Multiplication Rule for Conditional Probabilities.** Suppose that  $A_1, A_2, \dots, A_n$  are events such that  $\Pr(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$ . Then

$$\begin{aligned} \Pr(A_1 \cap A_2 \cap \dots \cap A_n) \\ = \Pr(A_1) \Pr(A_2|A_1) \Pr(A_3|A_1 \cap A_2) \dots \Pr(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}). \end{aligned} \quad (2.1.2)$$

**Proof** The product of probabilities on the right side of Eq. (2.1.2) is equal to

$$\Pr(A_1) \cdot \frac{\Pr(A_1 \cap A_2)}{\Pr(A_1)} \cdot \frac{\Pr(A_1 \cap A_2 \cap A_3)}{\Pr(A_1 \cap A_2)} \dots \frac{\Pr(A_1 \cap A_2 \cap \dots \cap A_n)}{\Pr(A_1 \cap A_2 \cap \dots \cap A_{n-1})}.$$

Since  $\Pr(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$ , each of the denominators in this product must be positive. All of the terms in the product cancel each other except the final numerator  $\Pr(A_1 \cap A_2 \cap \dots \cap A_n)$ , which is the left side of Eq. (2.1.2). ■

**Example**  
**2.1.7**

**Selecting Four Balls.** Suppose that four balls are selected one at a time, without replacement, from a box containing  $r$  red balls and  $b$  blue balls ( $r \geq 2, b \geq 2$ ). We shall determine the probability of obtaining the sequence of outcomes red, blue, red, blue.

If we let  $R_j$  denote the event that a red ball is obtained on the  $j$ th draw and let  $B_j$  denote the event that a blue ball is obtained on the  $j$ th draw ( $j = 1, \dots, 4$ ), then

$$\begin{aligned} \Pr(R_1 \cap B_2 \cap R_3 \cap B_4) &= \Pr(R_1) \Pr(B_2|R_1) \Pr(R_3|R_1 \cap B_2) \Pr(B_4|R_1 \cap B_2 \cap R_3) \\ &= \frac{r}{r+b} \cdot \frac{b}{r+b-1} \cdot \frac{r-1}{r+b-2} \cdot \frac{b-1}{r+b-3}. \end{aligned} \quad \blacktriangleleft$$

**Note: Conditional Probabilities Behave Just Like Probabilities.** In all of the situations that we shall encounter in this text, every result that we can prove has a conditional version given an event  $B$  with  $\Pr(B) > 0$ . Just replace *all* probabilities by conditional probabilities given  $B$  and replace all conditional probabilities given other events  $C$  by conditional probabilities given  $C \cap B$ . For example, Theorem 1.5.3 says that  $\Pr(A^c) = 1 - \Pr(A)$ . It is easy to prove that  $\Pr(A^c|B) = 1 - \Pr(A|B)$  if  $\Pr(B) > 0$ . (See Exercises 11 and 12 in this section.) Another example is Theorem 2.1.3, which is a conditional version of the multiplication rule Theorem 2.1.2. Although a proof is given for Theorem 2.1.3, we shall not provide proofs of all such conditional theorems, because their proofs are generally very similar to the proofs of the unconditional versions.

**Theorem**  
**2.1.3**

Suppose that  $A_1, A_2, \dots, A_n, B$  are events such that  $\Pr(B) > 0$  and  $\Pr(A_1 \cap A_2 \cap \dots \cap A_{n-1}|B) > 0$ . Then

$$\begin{aligned} \Pr(A_1 \cap A_2 \cap \dots \cap A_n|B) &= \Pr(A_1|B) \Pr(A_2|A_1 \cap B) \dots \\ &\quad \times \Pr(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1} \cap B). \end{aligned} \quad (2.1.3)$$

**Proof** The product of probabilities on the right side of Eq. (2.1.3) is equal to

$$\frac{\Pr(A_1 \cap B)}{\Pr(B)} \cdot \frac{\Pr(A_1 \cap A_2 \cap B)}{\Pr(A_1 \cap B)} \cdots \frac{\Pr(A_1 \cap A_2 \cap \cdots \cap A_n \cap B)}{\Pr(A_1 \cap A_2 \cap \cdots \cap A_{n-1} \cap B)}.$$

Since  $\Pr(A_1 \cap A_2 \cap \cdots \cap A_{n-1} | B) > 0$ , each of the denominators in this product must be positive. All of the terms in the product cancel each other except the first denominator and the final numerator to yield  $\Pr(A_1 \cap A_2 \cap \cdots \cap A_n \cap B) / \Pr(B)$ , which is the left side of Eq. (2.1.3). ■

## Conditional Probability and Partitions

Theorem 1.4.11 shows how to calculate the probability of an event by partitioning the sample space into two events  $B$  and  $B^c$ . This result easily generalizes to larger partitions, and when combined with Theorem 2.1.1 it leads to a very powerful tool for calculating probabilities.

### Definition 2.1.2

**Partition.** Let  $S$  denote the sample space of some experiment, and consider  $k$  events  $B_1, \dots, B_k$  in  $S$  such that  $B_1, \dots, B_k$  are disjoint and  $\bigcup_{i=1}^k B_i = S$ . It is said that these events form a *partition* of  $S$ .

Typically, the events that make up a partition are chosen so that an important source of uncertainty in the problem is reduced if we learn which event has occurred.

### Example 2.1.8

**Selecting Bolts.** Two boxes contain long bolts and short bolts. Suppose that one box contains 60 long bolts and 40 short bolts, and that the other box contains 10 long bolts and 20 short bolts. Suppose also that one box is selected at random and a bolt is then selected at random from that box. We would like to determine the probability that this bolt is long. ◀

Partitions can facilitate the calculations of probabilities of certain events.

### Theorem 2.1.4

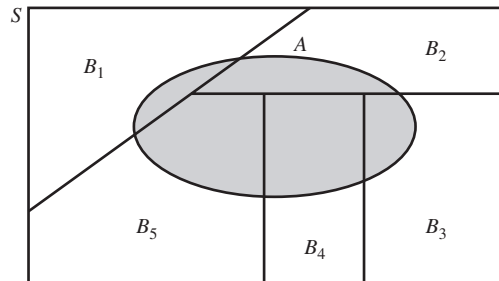
**Law of total probability.** Suppose that the events  $B_1, \dots, B_k$  form a partition of the space  $S$  and  $\Pr(B_j) > 0$  for  $j = 1, \dots, k$ . Then, for every event  $A$  in  $S$ ,

$$\Pr(A) = \sum_{j=1}^k \Pr(B_j) \Pr(A|B_j). \quad (2.1.4)$$

**Proof** The events  $B_1 \cap A, B_2 \cap A, \dots, B_k \cap A$  will form a partition of  $A$ , as illustrated in Fig. 2.2. Hence, we can write

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_k \cap A).$$

**Figure 2.2** The intersections of  $A$  with events  $B_1, \dots, B_5$  of a partition in the proof of Theorem 2.1.4.



Furthermore, since the  $k$  events on the right side of this equation are disjoint,

$$\Pr(A) = \sum_{j=1}^k \Pr(B_j \cap A).$$

Finally, if  $\Pr(B_j) > 0$  for  $j = 1, \dots, k$ , then  $\Pr(B_j \cap A) = \Pr(B_j) \Pr(A|B_j)$  and it follows that Eq. (2.1.4) holds. ■

**Example  
2.1.9**

**Selecting Bolts.** In Example 2.1.8, let  $B_1$  be the event that the first box (the one with 60 long and 40 short bolts) is selected, let  $B_2$  be the event that the second box (the one with 10 long and 20 short bolts) is selected, and let  $A$  be the event that a long bolt is selected. Then

$$\Pr(A) = \Pr(B_1) \Pr(A|B_1) + \Pr(B_2) \Pr(A|B_2).$$

Since a box is selected at random, we know that  $\Pr(B_1) = \Pr(B_2) = 1/2$ . Furthermore, the probability of selecting a long bolt from the first box is  $\Pr(A|B_1) = 60/100 = 3/5$ , and the probability of selecting a long bolt from the second box is  $\Pr(A|B_2) = 10/30 = 1/3$ . Hence,

$$\Pr(A) = \frac{1}{2} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{1}{3} = \frac{7}{15}. \quad \blacktriangleleft$$

**Example  
2.1.10**

**Achieving a High Score.** Suppose that a person plays a game in which his score must be one of the 50 numbers  $1, 2, \dots, 50$  and that each of these 50 numbers is equally likely to be his score. The first time he plays the game, his score is  $X$ . He then continues to play the game until he obtains another score  $Y$  such that  $Y \geq X$ . We will assume that, conditional on previous plays, the 50 scores remain equally likely on all subsequent plays. We shall determine the probability of the event  $A$  that  $Y = 50$ .

For each  $i = 1, \dots, 50$ , let  $B_i$  be the event that  $X = i$ . Conditional on  $B_i$ , the value of  $Y$  is equally likely to be any one of the numbers  $i, i + 1, \dots, 50$ . Since each of these  $(51 - i)$  possible values for  $Y$  is equally likely, it follows that

$$\Pr(A|B_i) = \Pr(Y = 50|B_i) = \frac{1}{51 - i}.$$

Furthermore, since the probability of each of the 50 values of  $X$  is  $1/50$ , it follows that  $\Pr(B_i) = 1/50$  for all  $i$  and

$$\Pr(A) = \sum_{i=1}^{50} \frac{1}{50} \cdot \frac{1}{51 - i} = \frac{1}{50} \left( 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{50} \right) = 0.0900. \quad \blacktriangleleft$$

**Note: Conditional Version of Law of Total Probability.** The law of total probability has an analog conditional on another event  $C$ , namely,

$$\Pr(A|C) = \sum_{j=1}^k \Pr(B_j|C) \Pr(A|B_j \cap C). \quad (2.1.5)$$

The reader can prove this in Exercise 17.

**Augmented Experiment** In some experiments, it may not be clear from the initial description of the experiment that a partition exists that will facilitate the calculation of probabilities. However, there are many such experiments in which such a partition exists if we imagine that the experiment has some additional structure. Consider the following modification of Examples 2.1.8 and 2.1.9.

**Example  
2.1.11**

**Selecting Bolts.** There is one box of bolts that contains some long and some short bolts. A manager is unable to open the box at present, so she asks her employees what is the composition of the box. One employee says that it contains 60 long bolts and 40 short bolts. Another says that it contains 10 long bolts and 20 short bolts. Unable to reconcile these opinions, the manager decides that each of the employees is correct with probability  $1/2$ . Let  $B_1$  be the event that the box contains 60 long and 40 short bolts, and let  $B_2$  be the event that the box contains 10 long and 20 short bolts. The probability that the first bolt selected is long is now calculated precisely as in Example 2.1.9. ◀

In Example 2.1.11, there is only one box of bolts, but we believe that it has one of two possible compositions. We let the events  $B_1$  and  $B_2$  determine the possible compositions. This type of situation is very common in experiments.

**Example  
2.1.12**

**A Clinical Trial.** Consider a clinical trial such as the study of treatments for depression in Example 2.1.4. As in many such trials, each patient has two possible outcomes, in this case relapse and no relapse. We shall refer to relapse as “failure” and no relapse as “success.” For now, we shall consider only patients in the imipramine treatment group. If we knew the effectiveness of imipramine, that is, the proportion  $p$  of successes among all patients who might receive the treatment, then we might model the patients in our study as having probability  $p$  of success. Unfortunately, we do not know  $p$  at the start of the trial. In analogy to the box of bolts with unknown composition in Example 2.1.11, we can imagine that the collection of all available patients (from which the 40 imipramine patients in this trial were selected) has two or more possible compositions. We can imagine that the composition of the collection of patients determines the proportion that will be success. For simplicity, in this example, we imagine that there are 11 different possible compositions of the collection of patients. In particular, we assume that the proportions of success for the 11 possible compositions are  $0, 1/10, \dots, 9/10, 1$ . (We shall be able to handle more realistic models for  $p$  in Chapter 3.) For example, if we knew that our patients were drawn from a collection with the proportion  $3/10$  of successes, we would be comfortable saying that the patients in our sample each have success probability  $p = 3/10$ . The value of  $p$  is an important source of uncertainty in this problem, and we shall partition the sample space by the possible values of  $p$ . For  $j = 1, \dots, 11$ , let  $B_j$  be the event that our sample was drawn from a collection with proportion  $(j - 1)/10$  of successes. We can also identify  $B_j$  as the event  $\{p = (j - 1)/10\}$ .

Now, let  $E_1$  be the event that the first patient in the imipramine group has a success. We defined each event  $B_j$  so that  $\Pr(E_1|B_j) = (j - 1)/10$ . Suppose that, prior to starting the trial, we believe that  $\Pr(B_j) = 1/11$  for each  $j$ . It follows that

$$\Pr(E_1) = \sum_{j=1}^{11} \frac{1}{11} \frac{j-1}{10} = \frac{55}{110} = \frac{1}{2}, \quad (2.1.6)$$

where the second equality uses the fact that  $\sum_{j=1}^n j = n(n+1)/2$ . ◀

The events  $B_1, B_2, \dots, B_{11}$  in Example 2.1.12 can be thought of in much the same way as the two events  $B_1$  and  $B_2$  that determine the mixture of long and short bolts in Example 2.1.11. There is only one box of bolts, but there is uncertainty about its composition. Similarly in Example 2.1.12, there is only one group of patients, but we believe that it has one of 11 possible compositions determined by the events  $B_1, B_2, \dots, B_{11}$ . To call these events, they must be subsets of the sample space for the experiment in question. That will be the case in Example 2.1.12 if we imagine that



the experiment consists not only of observing the numbers of successes and failures among the patients but also of potentially observing enough additional patients to be able to compute  $p$ , possibly at some time very far in the future. Similarly, in Example 2.1.11, the two events  $B_1$  and  $B_2$  are subsets of the sample space if we imagine that the experiment consists not only of observing one sample bolt but also of potentially observing the entire composition of the box.

Throughout the remainder of this text, we shall implicitly assume that experiments are augmented to include outcomes that determine the values of quantities such as  $p$ . We shall not require that we ever get to observe the complete outcome of the experiment so as to tell us precisely what  $p$  is, but merely that there is an experiment that includes all of the events of interest to us, including those that determine quantities like  $p$ .

**Definition 2.1.3** **Augmented Experiment.** If desired, any experiment can be augmented to include the potential or hypothetical observation of as much additional information as we would find useful to help us calculate any probabilities that we desire.

Definition 2.1.3 is worded somewhat vaguely because it is intended to cover a wide variety of cases. Here is an explicit application to Example 2.1.12.

**Example 2.1.13**

**A Clinical Trial.** In Example 2.1.12, we could explicitly assume that there exists an infinite sequence of patients who could be treated with imipramine even though we will observe only finitely many of them. We could let the sample space consist of infinite sequences of the two symbols  $S$  and  $F$  such as  $(S, S, F, S, F, F, F, \dots)$ . Here  $S$  in coordinate  $i$  means that the  $i$ th patient is a success, and  $F$  stands for failure. So, the event  $E_1$  in Example 2.1.12 is the event that the first coordinate is  $S$ . The example sequence above is then in the event  $E_1$ . To accommodate our interpretation of  $p$  as the proportion of successes, we can assume that, for every such sequence, the proportion of  $S$ 's among the first  $n$  coordinates gets close to one of the numbers  $0, 1/10, \dots, 9/10, 1$  as  $n$  increases. In this way,  $p$  is explicitly the limit of the proportion of successes we would observe if we could find a way to observe indefinitely. In Example 2.1.12,  $B_2$  is the event consisting of all the outcomes in which the limit of the proportion of  $S$ 's equals  $1/10$ ,  $B_3$  is the set of outcomes in which the limit is  $2/10$ , etc. Also, we observe only the first 40 coordinates of the infinite sequence, but we still behave as if  $p$  exists and could be determined if only we could observe forever. ◀

In the remainder of the text, there will be many experiments that we assume are augmented. In such cases, we will mention which quantities (such as  $p$  in Example 2.1.13) would be determined by the augmented part of the experiment even if we do not explicitly mention that the experiment is augmented.



## The Game of Craps

We shall conclude this section by discussing a popular gambling game called craps. One version of this game is played as follows: A player rolls two dice, and the sum of the two numbers that appear is observed. If the sum on the first roll is 7 or 11, the player wins the game immediately. If the sum on the first roll is 2, 3, or 12, the player loses the game immediately. If the sum on the first roll is 4, 5, 6, 8, 9, or 10, then the two dice are rolled again and again until the sum is either 7 or the original value. If the original value is obtained a second time before 7 is obtained, then the

player wins. If the sum 7 is obtained before the original value is obtained a second time, then the player loses.

We shall now compute the probability  $\Pr(W)$ , where  $W$  is the event that the player will win. Let the sample space  $S$  consist of all possible sequences of sums from the rolls of dice that might occur in a game. For example, some of the elements of  $S$  are  $(4, 7)$ ,  $(11)$ ,  $(4, 3, 4)$ ,  $(12)$ ,  $(10, 8, 2, 12, 6, 7)$ , etc. We see that  $(11) \in W$  but  $(4, 7) \in W^c$ , etc.. We begin by noticing that whether or not an outcome is in  $W$  depends in a crucial way on the first roll. For this reason, it makes sense to partition  $W$  according to the sum on the first roll. Let  $B_i$  be the event that the first roll is  $i$  for  $i = 2, \dots, 12$ .

Theorem 2.1.4 tells us that  $\Pr(W) = \sum_{i=2}^{12} \Pr(B_i) \Pr(W|B_i)$ . Since  $\Pr(B_i)$  for each  $i$  was computed in Example 1.6.5, we need to determine  $\Pr(W|B_i)$  for each  $i$ . We begin with  $i = 2$ . Because the player loses if the first roll is 2, we have  $\Pr(W|B_2) = 0$ . Similarly,  $\Pr(W|B_3) = 0 = \Pr(W|B_{12})$ . Also,  $\Pr(W|B_7) = 1$  because the player wins if the first roll is 7. Similarly,  $\Pr(W|B_{11}) = 1$ .

For each first roll  $i \in \{4, 5, 6, 8, 9, 10\}$ ,  $\Pr(W|B_i)$  is the probability that, in a sequence of dice rolls, the sum  $i$  will be obtained before the sum 7 is obtained. As described in Example 2.1.5, this probability is the same as the probability of obtaining the sum  $i$  when the sum must be either  $i$  or 7. Hence,

$$\Pr(W|B_i) = \frac{\Pr(B_i)}{\Pr(B_i \cup B_7)}.$$

We compute the necessary values here:

$$\begin{aligned} \Pr(W|B_4) &= \frac{\frac{3}{36}}{\frac{3}{36} + \frac{6}{36}} = \frac{1}{3}, & \Pr(W|B_5) &= \frac{\frac{4}{36}}{\frac{4}{36} + \frac{6}{36}} = \frac{2}{5}, \\ \Pr(W|B_6) &= \frac{\frac{5}{36}}{\frac{5}{36} + \frac{6}{36}} = \frac{5}{11}, & \Pr(W|B_8) &= \frac{\frac{5}{36}}{\frac{5}{36} + \frac{6}{36}} = \frac{5}{11}, \\ \Pr(W|B_9) &= \frac{\frac{4}{36}}{\frac{4}{36} + \frac{6}{36}} = \frac{2}{5}, & \Pr(W|B_{10}) &= \frac{\frac{3}{36}}{\frac{3}{36} + \frac{6}{36}} = \frac{1}{3}. \end{aligned}$$

Finally, we compute the sum  $\sum_{i=2}^{12} \Pr(B_i) \Pr(W|B_i)$ :

$$\begin{aligned} \Pr(W) &= \sum_{i=2}^{12} \Pr(B_i) \Pr(W|B_i) = 0 + 0 + \frac{3}{36} \frac{1}{3} + \frac{4}{36} \frac{2}{5} + \frac{5}{36} \frac{5}{11} + \frac{6}{36} \\ &\quad + \frac{5}{36} \frac{5}{11} + \frac{4}{36} \frac{2}{5} + \frac{3}{36} \frac{1}{3} + \frac{2}{36} + 0 = \frac{2928}{5940} = 0.493. \end{aligned}$$

Thus, the probability of winning in the game of craps is slightly less than 1/2.



## Summary

The revised probability of an event  $A$  after learning that event  $B$  (with  $\Pr(B) > 0$ ) has occurred is the conditional probability of  $A$  given  $B$ , denoted by  $\Pr(A|B)$  and computed as  $\Pr(A \cap B) / \Pr(B)$ . Often it is easy to assess a conditional probability, such as  $\Pr(A|B)$ , directly. In such a case, we can use the multiplication rule for conditional probabilities to compute  $\Pr(A \cap B) = \Pr(B) \Pr(A|B)$ . All probability results have versions conditional on an event  $B$  with  $\Pr(B) > 0$ : Just change *all* probabilities so that they are conditional on  $B$  in addition to anything else they were already

conditional on. For example, the multiplication rule for conditional probabilities becomes  $\Pr(A_1 \cap A_2|B) = \Pr(A_1|B) \Pr(A_2|A_1 \cap B)$ . A partition is a collection of disjoint events whose union is the whole sample space. To be most useful, a partition is chosen so that an important source of uncertainty is reduced if we learn which one of the partition events occurs. If the conditional probability of an event  $A$  is available given each event in a partition, the law of total probability tells how to combine these conditional probabilities to get  $\Pr(A)$ .

## Exercises

1. If  $A \subset B$  with  $\Pr(B) > 0$ , what is the value of  $\Pr(A|B)$ ?
2. If  $A$  and  $B$  are disjoint events and  $\Pr(B) > 0$ , what is the value of  $\Pr(A|B)$ ?
3. If  $S$  is the sample space of an experiment and  $A$  is any event in that space, what is the value of  $\Pr(A|S)$ ?
4. Each time a shopper purchases a tube of toothpaste, he chooses either brand A or brand B. Suppose that for each purchase after the first, the probability is  $1/3$  that he will choose the same brand that he chose on his preceding purchase and the probability is  $2/3$  that he will switch brands. If he is equally likely to choose either brand A or brand B on his first purchase, what is the probability that both his first and second purchases will be brand A and both his third and fourth purchases will be brand B?
5. A box contains  $r$  red balls and  $b$  blue balls. One ball is selected at random and its color is observed. The ball is then returned to the box and  $k$  additional balls of the same color are also put into the box. A second ball is then selected at random, its color is observed, and it is returned to the box together with  $k$  additional balls of the same color. Each time another ball is selected, the process is repeated. If four balls are selected, what is the probability that the first three balls will be red and the fourth ball will be blue?
6. A box contains three cards. One card is red on both sides, one card is green on both sides, and one card is red on one side and green on the other. One card is selected from the box at random, and the color on one side is observed. If this side is green, what is the probability that the other side of the card is also green?
7. Consider again the conditions of Exercise 2 of Sec. 1.10. If a family selected at random from the city subscribes to newspaper A, what is the probability that the family also subscribes to newspaper B?
8. Consider again the conditions of Exercise 2 of Sec. 1.10. If a family selected at random from the city subscribes to at least one of the three newspapers A, B, and C, what is the probability that the family subscribes to newspaper A?
9. Suppose that a box contains one blue card and four red cards, which are labeled A, B, C, and D. Suppose also that

two of these five cards are selected at random, without replacement.

- a. If it is known that card A has been selected, what is the probability that both cards are red?
- b. If it is known that at least one red card has been selected, what is the probability that both cards are red?
10. Consider the following version of the game of craps: The player rolls two dice. If the sum on the first roll is 7 or 11, the player wins the game immediately. If the sum on the first roll is 2, 3, or 12, the player loses the game immediately. However, if the sum on the first roll is 4, 5, 6, 8, 9, or 10, then the two dice are rolled again and again until the sum is either 7 or 11 or the original value. If the original value is obtained a second time before either 7 or 11 is obtained, then the player wins. If either 7 or 11 is obtained before the original value is obtained a second time, then the player loses. Determine the probability that the player will win this game.
11. For any two events  $A$  and  $B$  with  $\Pr(B) > 0$ , prove that  $\Pr(A^c|B) = 1 - \Pr(A|B)$ .
12. For any three events  $A$ ,  $B$ , and  $D$ , such that  $\Pr(D) > 0$ , prove that  $\Pr(A \cup B|D) = \Pr(A|D) + \Pr(B|D) - \Pr(A \cap B|D)$ .
13. A box contains three coins with a head on each side, four coins with a tail on each side, and two fair coins. If one of these nine coins is selected at random and tossed once, what is the probability that a head will be obtained?
14. A machine produces defective parts with three different probabilities depending on its state of repair. If the machine is in good working order, it produces defective parts with probability 0.02. If it is wearing down, it produces defective parts with probability 0.1. If it needs maintenance, it produces defective parts with probability 0.3. The probability that the machine is in good working order is 0.8, the probability that it is wearing down is 0.1, and the probability that it needs maintenance is 0.1. Compute the probability that a randomly selected part will be defective.