Chapter Overview

While the graphic and numeric methods of Chapters 2 and 3 provide us with tools for summarizing data, probability theory, the subject of this chapter, provides a foundation for developing statistical theory. Most people have an intuitive feeling for probability, but care is needed as intuition can lead you astray if it does not have a sure foundation.

There are different ways of thinking about what probability means, and three such ways are discussed in Sections 4.1 through 4.3. After talking about the meaning of events and their probabilities in Section 4.4, we introduce a number of rules for calculating probabilities in Section 4.5. Then follow two important ideas: conditional probability in Section 4.6 and independent events in Section 4.7. We shall exploit two aids in developing these ideas. The first is the two-way table of counts or proportions, which enables us to understand in a simple way how many probabilities encountered in real life are computed. The second is the tree diagram. This is often useful in clarifying our thinking about sequences of events. A number of case studies will highlight some practical features of probability and its possible pitfalls. We make linkages between probabilities and proportions of a finite population and point out that both can be manipulated using the same rules. An important function of this chapter is to give the reader a facility for dealing with data encountered in the form of counts and proportions, particularly in two-way tables.

4.1 INTRODUCTION

"If I toss a coin, what is the probability that it will turn up heads?" It's a rather silly question isn't it? Everyone knows the answer: "a half," "one chance in two," or "fifty-fifty." But let us look a bit more deeply behind this response that everyone makes.
First, why is the probability one half? When you ask a class, a dialogue like the one following often develops. “The probability is one half because the coin is equally likely to come down heads or tails.” Well, it could conceivably land on its edge, but we can fix that by tossing again. So why are the two outcomes, heads and tails, equally likely? “Because it’s a fair coin.” Sounds as if somebody has taken statistics before, but that’s just jargon, isn’t it? What does it mean? “Well, it’s symmetrical.” Have you ever seen a symmetrical coin? They are always different on both sides, with different bumps and indentations. These might influence the chances that the coin comes down heads. How could we investigate this? “We could toss the coin lots of times and see what happens.” This leads us to an intuitively attractive approach to probabilities for repeatable “experiments” such as coin tossing or die rolling. Probabilities are described in terms of long-run relative frequencies from repeated trials. Assuming these relative frequencies become stable after a large enough number of trials, the probability could be defined as the limiting relative frequency. Well, it turns out that several people have tried this.

English mathematician John Kerrich was lecturing at the University of Copenhagen when World War II broke out. He was arrested by the Germans and spent the war interned in a camp in Jutland. To help pass the time he performed some experiments in probability. One of these involved tossing a coin 10,000 times and recording the results. Figure 4.1.1 graphs the proportion of heads Kerrich obtained, up to and including toss numbers 10, 20, . . . , 100, 200, . . . , 900, 1,000, 2,000, . . . , 9,000, 10,000. The data, from Kerrich [1964], are also given in Freedman et al. [1991, Table 1, p. 248].

![Proportion of heads versus number of tosses for John Kerrich's coin-tossing experiment.](image)

---

John Kerrich hasn’t been the only person with time on his hands. In the eighteenth century, French naturalist Comte de Buffon performed 4,040 tosses for 2,048 heads; around 1900, Karl Pearson, one of the pioneers of modern statistics, made 24,000 tosses for 12,012 heads.
For the coin Kerrich was using, the proportion of heads certainly seems to be settling close to $\frac{1}{2}$. This is an empirical or experimental approach to probability. We never know the exact probability this way, but we can get a pretty good estimate. Kerrich's data give us an example of the frequently observed fact that in a long sequence of independent repetitions of a random phenomenon, the proportion of times in which an outcome occurs gets closer and closer to a fixed number, which we can call the probability that the outcome occurs. This is the relative frequency definition of probability. The long-run predictability of relative frequencies follows from a mathematical result called the law of large numbers.

But let's not abandon the first approach via symmetry quite so soon. The answer it gave us agreed well with Kerrich's experiment. It was based on an artificial idealization of reality—what we call a model. The coin is imagined as being completely symmetrical, so there should be no preference for landing heads or tails. Each should occur half the time. No coin is exactly symmetrical, however, nor in all likelihood has a probability of landing heads of exactly $\frac{1}{2}$. Yet experience has told us that the answer the model gives is close enough for all practical purposes. Although real coins are not symmetrical, they are close enough to being symmetrical for the model to work well.

At a University of London seminar series, N.Z. statistician Brian Dawkins asked the very question we used to open this discussion. He then left the stage and came back with an enormous coin, almost as big as himself.

The best he could manage was a single half turn. There was no element of chance at all. Some people can even make the tossing of an ordinary coin predictable. Program 15 of the Against All Odds video series (COMAP [1989]) shows Harvard probabilist Persi Diaconis tossing a fair coin so that it lands heads every time.² It is largely a matter of being able to repeat the same action. By just trying to make the action very similar we could probably shift the chances from 50-50, but most of us don't have that degree of interest or muscular control. A model in which coins turn up heads half the time and tails the other half in a totally unpredictable order is an excellent description of what we see when we toss a coin. Our physical model of a symmetrical coin gives rise to a probability model for the experiment. A probability model has two essential components: the sample space, which is simply a list of all the outcomes the experiment can have, and a list of probabilities, where each probability is intended to be the probability that the corresponding outcome occurs (see Section 4.4.4).

²Diaconis was a performing magician before training as a mathematician.
4. If A and B are events, when does “A or B” occur? When does “A and B” occur?  
(Section 4.4.5)

5. What are the two essential properties of a probability distribution \( p_1, p_2, \ldots, p_n \)?  
(Section 4.4.4)

6. How do we get the probability of an event from the probabilities of outcomes that make up that event?  

7. If all outcomes are equally likely, how do we calculate \( \text{pr}(A) \)?  

8. How do the concepts of a proportion and a probability differ? Under what circumstances are they numerically identical? What does this imply about using probability formulas to manipulate proportions of a population?  
(Section 4.4.5)

### 4.5 Probability Rules

The following rules are clear for finite sample spaces in which we obtain the probability of an event by adding the \( p_i \)'s for all outcomes in that event. The outcomes do not need to be equally likely.

- **Rule 1:** The sample space is certain to occur. \( \text{pr}(S) = 1 \)

- **Rule 2:** \( \text{pr}(A \text{ does not occur}) = 1 - \text{pr}(A \text{ does occur}) \). [Alternatively, \( \text{pr}(\text{an event occurs}) = 1 - \text{pr}(\text{it doesn’t occur}) \).]

- **Rule 3:** Addition rule for mutually exclusive (non-overlapping) events:
  \( \text{pr}(A \text{ or } B \text{ occurs}) = \text{pr}(A \text{ occurs}) + \text{pr}(B \text{ occurs}) \)

- **Aside:** What happens if \( A \text{ and } B \) overlap? In adding \( \text{pr}(A) + \text{pr}(B) \), we use the \( p_i \)'s relating to outcomes in \( A \text{ and } B \) twice. Thus we adjust by subtracting \( \text{pr}(A \text{ and } B) \). This gives us:
  \[ \text{pr}(A \text{ or } B \text{ occurs}) = \text{pr}(A \text{ occurs}) + \text{pr}(B \text{ occurs}) - \text{pr}(A \text{ and } B \text{ occur}) \]

\( ^8 \)The rules apply generally, not just to finite sample spaces.

\( ^9 \)The sample space contains all possible outcomes.
**Example 4.5.1 Random Numbers**

A random number from 1 to 10 is selected from a table of random numbers. Let $A$ be the event that "the number selected is 9 or less." As all 10 possible outcomes are equally likely, $\text{pr}(A) = \frac{9}{10} = 0.9$. The complement of event $A$ has only one outcome, so $\text{pr}(\overline{A}) = 1 - \frac{9}{10} = 0.1$. Note that

$$\text{pr}(A) = 1 - \text{pr}(\overline{A}) = 1 - 0.1 = 0.9$$

This formula is very useful in any situation where $\text{pr}(\overline{A})$ is easier to obtain than $\text{pr}(A)$. It is used in this way in Example 4.5.4 and frequently thereafter.

**Example 4.5.2 Probability Rules and a Two-Way Table**

Suppose that between the hours of 9.00 a.m. and 5.30 p.m., Dr. Wild is available for student help 70% of the time, Dr. Seber is available 60% of the time, and both are available (simultaneously) 50% of the time. A student comes for help at some random time in these hours. Let $A = \text{"Wild in"}$ and $B = \text{"Seber in."}$ Then $A$ and $B = \text{"both in."}$ If a time is chosen at random, then $\text{pr}(A) = 0.7$, $\text{pr}(B) = 0.6$ and $\text{pr}(A$ and $B) = 0.5$. What is the probability, $\text{pr}(A$ or $B)$, that at least one of Wild and Seber are in?

We can obtain this probability, and many other probabilities, without using any algebra if we put our information into a two-way table, as in Fig. 4.5.1. The left-hand (descriptive) table categorizes the possible outcomes of Wild being in or out and Seber being in or out. Each interior cell in the table is to contain the probability that a particular combination occurs. At present we have information on only one combination, namely, $\text{pr(Wild in and Seber in)} = 0.5$, and we have placed that in the table. The left column total tells us about Seber being in regardless of where Wild is, so that is where $\text{pr(Seber in)} = 0.6$ belongs. Similarly, $\text{pr(Wild in)} = 0.7$ belongs in the top row total position. Everything else in the table is unknown at present. The right-hand (algebraic) table relates all this information to the algebraic events and notation we set up earlier.

<table>
<thead>
<tr>
<th></th>
<th>Seber In</th>
<th>Seber Out</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild In</td>
<td>0.5</td>
<td>0.7</td>
<td>1.00</td>
</tr>
<tr>
<td>Wild Out</td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.6</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

**Descriptive Table**

**Algebraic Table**

$$\begin{align*}
\text{pr}(A) & = 0.7 \\
\text{pr}(B) & = 0.6 \\
\text{pr}(A$ and $B) & = 0.5 \\
\text{pr}(A$ or $B) & = ?
\end{align*}$$

**Example 4.5.3 Putting Wild–Seber information into a two-way table.**

The missing entries in the left-hand table can now be filled in using the simple idea that rows add to row totals and columns add to column totals.

<table>
<thead>
<tr>
<th></th>
<th>Seber In</th>
<th>Seber Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild In</td>
<td>0.5</td>
<td>0.7</td>
</tr>
<tr>
<td>Wild Out</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Total</td>
<td>0.6</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Partial Table**

$$\begin{align*}
\text{pr}(\text{Wild in}) & = 0.7 \\
\text{pr}(\text{Seber in}) & = 0.6 \\
\text{pr}(\text{Wild and Seber in}) & = 0.5
\end{align*}$$

**Complete Table**

$$\begin{align*}
\text{pr}(\text{Wild in and Seber in}) & = 0.5 \\
\text{pr}(\text{Wild in}) & = 0.7 \\
\text{pr}(\text{Seber in}) & = 0.6 \\
\text{pr}(\text{Wild in or Seber in}) & = ?
\end{align*}$$
Our final table is reproduced in Table 4.5.1. To find the probability that at least one person is in, namely pr( or ), we now sum the probabilities in the (row 1, column 1), (row 1, column 2), and (row 2, column 1) positions:

\[ pr(A \text{ or } B) = 0.5 + 0.2 + 0.1 = 0.8 \]

We can also find a number of other probabilities from the table, such as the following. (Algebraic expressions are appended to each line that follows, not because we want you to work with complicated algebraic expressions, but just so that you will appreciate the power and elegance of the table.)

\[
\begin{align*}
pr(\text{Wild in and Seber out}) &= 0.2 &\text{[pr}(A \text{ and } \overline{B})]\npr(\text{At least one of Wild and Seber are out}) &= 0.2 + 0.1 + 0.2 &\text{[pr}(A \text{ or } \overline{B})]\npr(\text{One of Wild and Seber is in and the other is out}) &= 0.2 + 0.1 &\text{[pr}(A \text{ and } \overline{B}) \text{ or } (\overline{A} \text{ and } B)]\npr(\text{Both Wild and Seber are in or both are out}) &= 0.5 + 0.2 &\text{[pr}(A \text{ and } B) \text{ or } (\overline{A} \text{ and } \overline{B})]}
\end{align*}
\]

The preceding example is typical of a number of situations where we have partial information about various combinations of events and we wish to find some new probability. To do this we simply fill in the appropriate two-way table. We have used some ideas here that seem natural in the context of the table, but we have not formalized them yet. One idea used was adding up the probabilities of mutually exclusive events.

### 4.5.1 Addition Rule for More Than Two Mutually Exclusive Events

Events \(A_1, A_2, \ldots, A_k\) are all mutually exclusive if they have no overlap, that is, if no two of them can occur at the same time. If they are mutually exclusive, we can get the probability that at least one of them occurs (sometimes referred to as the union)
simply by adding:

$$\Pr(A_1 \text{ or } A_2 \text{ or } \ldots \text{ or } A_k) = \Pr(A_1) + \Pr(A_2) + \cdots + \Pr(A_k)$$

The addition law therefore applies to any number of mutually exclusive events.

**Example 4.5.3  Equally Likely Random Numbers**

A number is drawn at random from 1 to 10 so that the sample space $S = \{1, 2, \ldots, 10\}$. Let $A_1 = \{\text{an even number chosen}\}$, $A_2 = \{1, 5\}$, and $A_3 = \{7\}$. If $B = A_1$ or $A_2$ or $A_3$, find $\Pr(B)$.

Since the 10 outcomes are equally likely, $\Pr(A_1) = \frac{5}{10}$, $\Pr(A_2) = \frac{2}{10}$, and $\Pr(A_3) = \frac{1}{10}$. The three events $A_i$ are mutually exclusive so that, by the addition law,

$$\Pr(B) = \Pr(A_1) + \Pr(A_2) + \Pr(A_3) = \frac{5}{10} + \frac{2}{10} + \frac{1}{10} = \frac{9}{10}$$

Alternatively, $B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ so that $n(B) = 9$ and, by rule 2,

$$\Pr(B) = 1 - \Pr(\overline{B}) = 1 - \frac{1}{10} = \frac{9}{10}$$

**Example 4.5.4  Events Relating to the Marital Status of U.S. Couples**

Table 4.5.2 is a two-way table of proportions that cross-classifies male-female couples in the United States who are not married to each other by the marital status of the partners. Each entry within the table is the proportion of couples with a given combination of marital statuses. Let us consider choosing a couple at random. We shall take all of the couples represented in the table as our sample space. The table proportions give us the probabilities of the events defined by the row and column titles (see Example 4.4.9 and Section 4.4.5). Thus the probability of getting a couple in which

<table>
<thead>
<tr>
<th>Male</th>
<th>Female</th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never married</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.554</td>
</tr>
<tr>
<td>Divorced</td>
<td>.401</td>
<td>.111</td>
<td>.017</td>
<td>.025</td>
<td>.535</td>
</tr>
<tr>
<td>Widowed</td>
<td>.117</td>
<td>.195</td>
<td>.024</td>
<td>.017</td>
<td>.353</td>
</tr>
<tr>
<td>Married to other</td>
<td>.006</td>
<td>.008</td>
<td>.016</td>
<td>.001</td>
<td>.031</td>
</tr>
<tr>
<td>Total</td>
<td>.545</td>
<td>.336</td>
<td>.060</td>
<td>.059</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Source: Constructed, with permission, from data in The World Almanac (1993, p. 942). © PRIMEDIA Reference Inc. All rights reserved.
the male has never been married and the female is divorced is 0.111. Here ‘married to other’ means living as a member of a couple with one person while married to someone else.

The 16 cells in the table, which relate to different combinations of the status of the male and the female partner, correspond to mutually exclusive events (no couple belongs to more than one cell). For this reason we can obtain probabilities of collections of cells by simply adding cell probabilities. Moreover, these 16 mutually exclusive events account for all the couples in the sample space so that their probabilities add to 1 (by rule 1).^{10}

Each row total in the table tells us about the proportion of couples with males in the given category, regardless of the status of the females. Thus in 35.3% of couples the male is divorced, or equivalently, 0.353 is the probability that the male of a randomly chosen couple is divorced.

We shall continue to use Table 4.5.2 to further illustrate the use of the probability rules in the previous subsection. The complement of the event ‘at least one member of the couple has been married’ is the event that both are in the ‘never married’ category. Thus by rule 2,

\[
pr(\text{At least one has been married}) = 1 - pr(\text{Both never married})
\]

\[
= 1 - 0.401 = 0.599
\]

Note how much simpler it is in this case to use the complement than to calculate the desired probability directly by adding the probabilities of the 15 cells that fall into the event ‘at least one married.’

QUIZ ON SECTION 4.5

1. If A and B are mutually exclusive, what is the probability that both occur? What is the probability that at least one occurs?

2. If we have two or more mutually exclusive events, how do we find the probability that at least one of them occurs?

3. Why is it sometimes easier to compute pr(A) from 1 – pr(\overline{A})?

EXERCISES FOR SECTION 4.5

1. A house needs to be reroofed in the spring. To do this a dry, windless day is needed. The probability of getting a dry day is 0.7, a windy day is 0.4, and a wet and windy day is 0.2. What is the probability of getting
   (a) a wet day?
   (b) a day that is either wet, windy, or both?
   (c) a day when the house can be reroofed?

2. Using the data in Table 4.5.2, what is the probability that for a randomly chosen unmarried couple:
   (a) the male is divorced or married to someone else?
   (b) both the male and the female are either divorced or married to someone else?
   (c) neither is married to anyone else?

^{10} Such a collection of events is technically known as a partition of S.
(d) at least one is married to someone else?
(e) the male is married to someone else or the female is divorced or both?
(f) the female is divorced and the male is not divorced?

3. A young man with a barren love life feels tempted to become a contestant on the television game show “Blind Date.” He decides to watch a few programs first to assess his chances of being paired with a suitable date: someone whom he finds attractive and who is no taller than he is (as he is self-conscious about his height). After watching 50 female contestants, he decides that he is not attracted to 8, that 12 are too tall, and that 9 are attractive but too tall. If these figures are typical, what is the probability of getting someone
(a) who is both unattractive and too tall?
(b) whom he likes, that is, is not unattractive or too tall?
(c) who is not too tall but is unattractive?

4. For the data and situation in Example 4.4.9, what is the probability that a random job loss
(a) was not by a male who lost it because the workplace moved?
(b) was by a male or someone who lost it because the workplace moved?
(c) was by a male but for some reason other than the workplace moving?

4.6 CONDITIONAL PROBABILITY

4.6.1 Definition

Our assessment of the chances that an event will occur can be very different depending on the information that we have. An estimate of the probability that your house will collapse tomorrow should clearly be much larger if a violent earthquake is expected than it would be if there were no reason to expect unusual seismic activity. The two examples that follow give a more concrete demonstration of how an assessment of the chances of an event occurring may change radically if we are given information about whether event B has occurred or not.

**EXAMPLE 4.6.1 Conditional Probability of Zero**

Suppose we toss two fair coins and \( S = \{HH, HT, TH, TT\} \). Let \( A \) = “two tails” = \{TT\} and \( B \) = “at least one head” = \{HH, HT, TH\}. Since all four outcomes in \( S \) are equally likely, \( P(A) = \frac{1}{4} \). If we know that \( B \) has occurred, however, then \( A \) cannot occur. Hence the conditional probability of \( A \) given that \( B \) has occurred is 0.

**EXAMPLE 4.6.2 Conditional Probabilities from a Two-Way Table of Frequencies**

Table 4.6.1 was obtained by cross-classifying 400 patients with a form of skin cancer called malignant melanoma with respect to the histological type of their cancer and its location (site) on their bodies. We see, for example, that 35 patients have nodular

---

Histological type means the type of abnormality observed in the cells that make up the cancer.