PART IV

Randomness and Probability

CHAPTER 14

From Randomness to Probability

Early humans saw a world filled with random events. To help them make sense of the chaos around them, they sought out seers, consulted oracles, and read tea leaves. As science developed, we learned to recognize some events as predictable. We can now forecast the change of seasons, tell when eclipses will occur precisely, and even make a reasonably good guess at how warm it will be tomorrow. But many other events are still essentially random. Will the stock market go up or down today? When will the next car pass this corner? And we now know from quantum mechanics that the universe is in some sense random at the most fundamental levels of subatomic particles.

But we have also learned to understand randomness. The surprising fact is that in the long run, even truly random phenomena settle down in a way that's consistent and predictable. It's this property of random phenomena that makes the next steps we're about to take in Statistics possible.

Dealing with Random Phenomena

Every day you drive through the intersection at College and Main. Even though it may seem that the light is never green when you get there, you know this can't really be true. In fact, if you try really hard, you can recall just sailing through the green light once in a while.

What's random here? The light itself is governed by a timer. Its pattern isn't haphazard. In fact, the light may even be red at precisely the same times each day. It's the pattern of *your driving* that is random. No, we're certainly not insinuating that you can't keep the car on the road. At the precision level of the 30 seconds or so that the light spends being red or green, the time you arrive at the light *is random.* Even if you try to leave your house at exactly the same time every day, whether the light is red or green as *you* reach the intersection is a **random phenomenon.**[1]

[1] If you somehow managed to leave your house at *precisely* the same time every day and there was *no* variation in the time it took you to get to the light, then there wouldn't be any randomness, but that's not very realistic.

Is the color of the light completely unpredictable? When you stop to think about it, it's clear that you do expect some kind of *regularity* in your long-run experience. Some *fraction* of the time, the light will be green as you get to the intersection. How can you figure out what that fraction is?

You might record what happens at the intersection each day and graph the *accumulated percentage* of green lights like this:

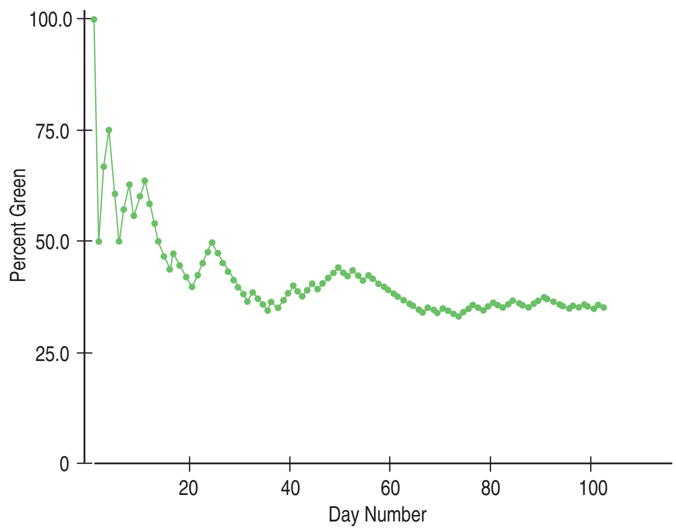


Figure 14.1
The overall percentage of times the light is green settles down as you see more outcomes.

Day	Light	% Green
1	Green	100
2	Red	50
3	Green	66.7
4	Green	75
5	Red	60
6	Red	50
⋮	⋮	⋮

The first day you recorded the light, it was green. Then on the next five days, it was red, then green again, then green, red, and red. If you plot the percentage of green lights against days, the graph would start at 100% (because the first time, the light was green, so 1 out of 1, for 100%). Then the next day it was red, so the accumulated percentage dropped to 50% (1 out of 2). The third day it was green again (2 out of 3, or 67% green), then green (3 out of 4, or 75%), then red twice in a row (3 out of 5, for 60% green, and then 3 out of 6, for 50%), and so on. As you collect a new data value for each day, each new outcome becomes a smaller and smaller fraction of the accumulated experience, so, in the long run, the graph settles down. As it settles down, you can see that, in fact, the light is green about 35% of the time.

A phenomenon consists of trials. Each trial has an outcome. Outcomes combine to make events.

When talking about random phenomena such as this, it helps to define our terms. You aren't interested in the traffic light *all* the time. You pull up to the intersection only once a day, so you care about the color of the light only at these particular times.[2] In general, each occasion upon which we observe a random phenomenon is called a **trial.** At each trial, we note the value of the random phenomenon, and call that the trial's **outcome.** (If this language reminds you of Chapter 11, that's *not* unintentional.)

For the traffic light, there are really three possible outcomes: red, yellow, or green. Often we're more interested in a combination of outcomes rather than in the individual ones. When you see the light turn yellow, what do *you* do? If you race through the intersection, then you treat the yellow more like a green light. If you step on the brakes, you treat it more like a red light. Either way, you might want to group the yellow with one or the other. When we combine outcomes like that, the resulting combination is an **event.**[3] We sometimes talk about the collection of *all possible outcomes* and call that event the **sample space.**[4] We'll denote the sample

[2] Even though the randomness here comes from the uncertainty in our arrival time, we can think of the light itself as showing a color at random.
[3] Each individual outcome is also an event.
[4] Mathematicians like to use the term "space" as a fancy name for a set. Sort of like referring to that closet colleges call a dorm room as "living space." But remember that it's really just the set of all outcomes.

space **S**. (Some books are even fancier and use the Greek letter Ω.) For the traffic light, **S** = {red, green, yellow}.

The Law of Large Numbers

"For even the most stupid of men . . . is convinced that the more observations have been made, the less danger there is of wandering from one's goal."

—Jacob Bernoulli, 1713, discoverer of the LLN

What's the *probability* of a green light at College and Main? Based on the graph, it looks like the relative frequency of green lights settles down to about 35%, so saying that the probability is about 0.35 seems like a reasonable answer. But do random phenomena always behave well enough for this to make sense? Perhaps the relative frequency of an event can bounce back and forth between two values forever, never settling on just one number.

Fortunately, a principle called the **Law of Large Numbers** (LLN) gives us the guarantee we need. It simplifies things if we assume that the events are **independent.** Informally, this means that the outcome of one trial doesn't affect the outcomes of the others. (We'll see a formal definition of independent events in the next chapter.) The LLN says that as the number of independent trials increases, the long-run *relative frequency* of repeated events gets closer and closer to a single value.

Although the LLN wasn't proven until the 18th century, everyone expects the kind of long-run regularity that the Law describes from everyday experience. In fact, the first person to prove the LLN, Jacob Bernoulli, thought it was pretty obvious, too, as his remark quoted in the margin shows.[5]

Because the LLN guarantees that relative frequencies settle down in the long run, we can now officially give a name to the value that they approach. We call it the **probability** of the event. If the relative frequency of green lights at that intersection settles down to 35% in the long run, we say that the probability of encountering a green light is 0.35, and we write $P(\text{green}) = 0.35$. Because this definition is based on repeatedly observing the event's outcome, this definition of probability is often called **empirical probability.**

Empirical Probability

For any event **A**,

$$P(\mathbf{A}) = \frac{\#\text{ times }\mathbf{A}\text{ occurs}}{\text{total }\#\text{ of trials}}$$

in the long run.

The Nonexistent Law of Averages

Don't let yourself think that there's a Law of Averages that promises short-term compensation for recent deviations from expected behavior. A belief in such a "Law" can lead to money lost in gambling and to poor business decisions.

Even though the LLN seems natural, it is often misunderstood because the idea of the *long run* is hard to grasp. Many people believe, for example, that an outcome of a random event that hasn't occurred in many trials is "due" to occur. Many gamblers bet on numbers that haven't been seen for a while, mistakenly believing that they're likely to come up sooner. A common term for this is the "Law of Averages." After all, we know that in the long run, the relative frequency will settle down to the probability of that outcome, so now we have some "catching up" to do, right?

Wrong. The Law of Large Numbers says nothing about short-run behavior. Relative frequencies even out *only in the long run.* And, according to the LLN, the long run is *really* long (*infinitely* long, in fact).

The so-called Law of Averages doesn't exist at all. But you'll hear people talk about it as if it does. Is a good hitter in baseball who has struck out the last six times *due* for a hit his next time up? If you've been doing particularly well in weekly quizzes in Statistics class, are you *due* for a bad grade? No. This isn't the way random phenomena work. There is *no* Law of Averages for short runs.

The lesson of the LLN is that sequences of random events don't compensate in the *short* run and don't need to do so to get back to the right long-run probability.

"Slump? I ain't in no slump. I just ain't hittin'."

—Yogi Berra

[5] In case you were wondering, Jacob's reputation was that he was every bit as nasty as this quotation suggests. He and his brother, who was also a mathematician, fought publicly over who had accomplished the most.

The Law of Averages in Everyday Life

"Dear Abby: My husband and I just had our eighth child. Another girl, and I am really one disappointed woman. I suppose I should thank God she was healthy, but, Abby, this one was supposed to have been a boy. Even the doctor told me that the law of averages was in our favor 100 to one." (Abigail Van Buren, 1974. Quoted in Karl Smith, *The Nature of Mathematics.* 6th ed. Pacific Grove, CA: Brooks/Cole, 1991, p. 589)

If the probability of an outcome doesn't change and the events are independent, the probability of any outcome in another trial is *always* what it was, no matter what has happened in other trials.

Coins, Keno, and the Law of Averages You've just flipped a fair coin and seen six heads in a row. Does the coin "owe" you some tails? Suppose you spend that coin and your friend gets it in change. When she starts flipping the coin, should she expect a run of tails? Of course not. Each flip is a new event. The coin can't "remember" what it did in the past, so it can't "owe" any particular outcomes in the future.

Just to see how this works in practice, we ran a simulation of 100,000 flips of a fair coin. We collected 100,000 random numbers, letting the numbers 0 to 4 represent heads and the numbers 5 to 9 represent tails. In our 100,000 "flips," there were 2981 streaks of at least 5 heads. The "Law of Averages" suggests that the next flip after a run of 5 heads should be tails more often to even things out. Actually, the next flip was heads more often than tails: 1550 times to 1431 times. That's 51.9% heads. You can perform a similar simulation easily on a computer. Try it!

Of course, sometimes an apparent drift from what we expect means that the probabilities are, in fact, *not* what we thought. If you get 10 heads in a row, maybe the coin has heads on both sides!

Keno is a simple casino game in which numbers from 1 to 80 are chosen. The numbers, as in most lottery games, are supposed to be equally likely. Payoffs are made depending on how many of those numbers you match on your card. A group of graduate students from a Statistics department decided to take a field trip to Reno. They (*very* discreetly) wrote down the outcomes of the games for a couple of days, then drove back to test whether the numbers were, in fact, equally likely. It turned out that some numbers were *more likely* to come up than others. Rather than bet on the Law of Averages and put their money on the numbers that were "due," the students put their faith in the LLN—and all their (and their friends') money on the numbers that had come up before. After they pocketed more than $50,000, they were escorted off the premises and invited never to show their faces in that casino again.

TI-*nspire*

The Law of Large Numbers. Watch the relative frequency of a random event approach the true probability *in the long run*.

JUST CHECKING

1. One common proposal for beating the lottery is to note which numbers have come up lately, eliminate those from consideration, and bet on numbers that have not come up for a long time. Proponents of this method argue that in the long run, every number should be selected equally often, so those that haven't come up are due. Explain why this is faulty reasoning.

Modeling Probability

A S ***Activity:*** **What Is Probability?** The best way to get a feel for probabilities is to experiment with them. We'll use this random-outcomes tool many more times.

Probability was first studied extensively by a group of French mathematicians who were interested in games of chance.[6] Rather than *experiment* with the games (and risk losing their money), they developed mathematical models of **theoretical probability.** To make things simple (as we usually do when we build models), they started by looking at games in which the different outcomes were equally likely. Fortunately, many games of chance are like that. Any of 52 cards is equally

[6] Ok, gambling.

likely to be the next one dealt from a well-shuffled deck. Each face of a die is equally likely to land up (or at least it *should be*).

It's easy to find probabilities for events that are made up of several *equally likely* outcomes. We just count all the outcomes that the event contains. The probability of the event is the number of outcomes in the event divided by the total number of possible outcomes. We can write

$$P(\mathbf{A}) = \frac{\text{\# outcomes in } \mathbf{A}}{\text{\# of possible outcomes}}.$$

For example, the probability of drawing a face card (JQK) from a deck is

$$P(\text{face card}) = \frac{\text{\# face cards}}{\text{\# cards}} = \frac{12}{52} = \frac{3}{13}.$$

NOTATION ALERT:

We often use capital letters—and usually from the beginning of the alphabet—to denote events. We *always* use P to denote probability. So,

$$P(\mathbf{A}) = 0.35$$

means "the probability of the event **A** is 0.35."

When being formal, use decimals (or fractions) for the probability values, but sometimes, especially when talking more informally, it's easier to use percentages.

Is that all there is to it? Finding the probability of any event when the outcomes are equally likely is straightforward, but not necessarily easy. It gets hard when the number of outcomes in the event (and in the sample space) gets big. Think about flipping two coins. The sample space is **S** = {HH,HT,TH,TT} and each outcome is equally likely. So, what's the probability of getting *exactly* one head and one tail? Let's call that event **A**. Well, there are two outcomes in the event **A** = {HT,TH} out of the 4 possible equally likely ones in **S,** so $P(\mathbf{A}) = \frac{2}{4}$, or $\frac{1}{2}$.

OK, now flip 100 coins. What's the probability of exactly 67 heads? Well, first, how many outcomes are in the sample space? **S** = {HHHHHHHHHHH . . . H, HH . . . T, . . .} Hmm. A lot. In fact, there are 1,267,650,600,228,229,401,496,703,205,376 different outcomes possible when flipping 100 coins. To answer the question, we'd still have to figure out how many ways there are to get 67 heads. That's coming in Chapter 17; stay tuned!

Activity: **Multiple Discrete Outcomes.** The world isn't all heads or tails. Experiment with an event with 4 random alternative outcomes.

Don't get trapped into thinking that random events are always equally likely. The chance of winning a lottery—especially lotteries with very large payoffs—is small. Regardless, people continue to buy tickets. In an attempt to understand why, an interviewer asked someone who had just purchased a lottery ticket, "What do you think your chances are of winning the lottery?" The reply was, "Oh, about 50–50." The shocked interviewer asked, "How do you get that?" to which the response was, "Well, the way I figure it, either I win or I don't!"

The moral of this story is that events are *not* always equally likely.

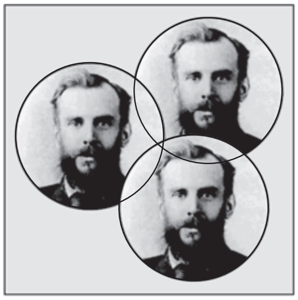
Personal Probability

What's the probability that your grade in this Statistics course will be an **A**? You may be able to come up with a number that seems reasonable. Of course, no matter how confident or depressed you feel about your chances for success, your probability should be between 0 and 1. How did you come up with this probability? Is it an empirical probability? Not unless you plan on taking the course over and over (and over . . .), calculating the proportion of times you get an **A.** And, unless you assume the outcomes are equally likely, it will be hard to find the theoretical probability. But people use probability in a third sense as well.

We use the language of probability in everyday speech to express a degree of uncertainty *without* basing it on long-run relative frequencies or mathematical models. Your personal assessment of your chances of getting an **A** expresses your uncertainty about the outcome. That uncertainty may be based on how comfortable you're feeling in the course or on your midterm grade, but it can't be based on long-run behavior. We call this third kind of probability a subjective or **personal probability.**

Although personal probabilities may be based on experience, they're not based either on long-run relative frequencies or on equally likely events. So they don't display the kind of consistency that we'll need probabilities to have. For that reason, we'll stick to formally defined probabilities. You should be alert to the difference.

The First Three Rules for Working with Probability



John Venn (1834–1923) created the Venn diagram. His book on probability, The Logic of Chance, *was "strikingly original and considerably influenced the development of the theory of Statistics," according to John Maynard Keynes, one of the luminaries of Economics.*

1. Make a picture.
2. Make a picture.
3. Make a picture.

We're dealing with probabilities now, not data, but the three rules don't change. The most common kind of picture to make is called a Venn diagram. We'll use Venn diagrams throughout the rest of this chapter. Even experienced statisticians make Venn diagrams to help them think about probabilities of compound and overlapping events. You should, too.

Formal Probability

Surprising Probabilities

We've been careful to discuss probabilities only for situations in which the outcomes were finite, or even countably infinite. But if the outcomes can take on *any* numerical value at all (we say they are *continuous*), things can get surprising. For example, what is the probability that a randomly selected child will be *exactly* 3 feet tall? Well, if we mean 3.00000 . . . feet, the answer is zero. No randomly selected child—even one whose height would be recorded as 3 feet, will be *exactly* 3 feet tall (to an infinite number of decimal places). But, if you've grown taller than 3 feet, there must have been a time in your life when you actually *were* exactly 3 feet tall, even if only for a second. So this is an outcome with probability 0 that not only has happened—it has happened to *you.*

We've seen another example of this already in Chapter 6 when we worked with the Normal model. We said that the probability of any *specific* value—say, $z = 0.5$—is zero. The model gives a probability for any *interval* of values, such as $0.49 < z < 0.51$. The probability is smaller if we ask for $0.499 < z < 0.501$, and smaller still for $0.49999999 < z < 0.50000001$. Well, you get the idea. Continuous probabilities are useful for the mathematics behind much of what we'll do, but it's easier to deal with probabilities for countable outcomes.

For some people, the phrase "50/50" means something vague like "I don't know" or "whatever." But when we discuss probabilities of outcomes, it takes on the precise meaning of *equally likely*. Speaking vaguely about probabilities will get us into trouble, so whenever we talk about probabilities, we'll need to be precise.[7] And to do that, we'll need to develop some formal rules[8] about how probability works.

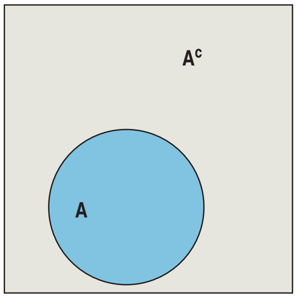
1. If the probability is 0, the event can't occur, and likewise if it has probability 1, it *always* occurs. Even if you think an event is very unlikely, its probability can't be negative, and even if you're sure it will happen, its probability can't be greater than 1. So we require that

A probability is a number between 0 and 1.

For any event A, $0 \le P(A) \le 1$.

[7] And to be precise, we will be talking only about sample spaces where we can enumerate all the outcomes. Mathematicians call this a countable number of outcomes.

[8] Actually, in mathematical terms, these are axioms—statements that we assume to be true of probability. We'll derive other rules from these in the next chapter.



The set **A** *and its complement* $\mathbf{A^C}$*. Together, they make up the entire sample space* **S**.

NOTATION ALERT:

We write $P(\mathbf{A}$ or $\mathbf{B})$ as $P(\mathbf{A} \cup \mathbf{B})$. The symbol $\cup$ means "union," representing the outcomes in event **A** *or* event **B** (or both). The symbol $\cap$ means "intersection," representing outcomes that are in both event **A** *and* event **B**. We write $P(\mathbf{A}$ *and* $\mathbf{B})$ as $P(\mathbf{A} \cap \mathbf{B})$.

2. If a random phenomenon has only one possible outcome, it's not very interesting (or very random). So we need to distribute the probabilities among all the outcomes a trial can have. How can we do that so that it makes sense? For example, consider what you're doing as you read this book. The possible outcomes might be

A: You read to the end of this chapter before stopping.
B: You finish this section but stop reading before the end of the chapter.
C: You bail out before the end of this section.

When we assign probabilities to these outcomes, the first thing to be sure of is that we distribute all of the available probability. Something always occurs, so the probability of the entire sample space is 1.

Making this more formal gives the **Probability Assignment Rule.**

The set of all possible outcomes of a trial must have probability 1.

$$P(\mathbf{S}) = 1$$

3. Suppose the probability that you get to class on time is 0.8. What's the probability that you don't get to class on time? Yes, it's 0.2. The set of outcomes that are *not* in the event **A** is called the **complement** of **A,** and is denoted $\mathbf{A^C}$**.** This leads to the **Complement Rule:**

The probability of an event occurring is 1 minus the probability that it doesn't occur.

$$P(\mathbf{A}) = 1 - P(A^C)$$

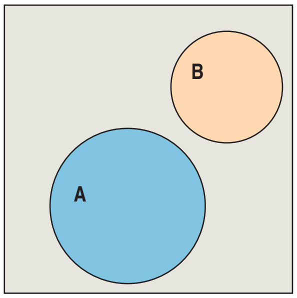
FOR EXAMPLE Applying the Complement Rule

Recap: We opened the chapter by looking at the traffic light at the corner of College and Main, observing that when we arrive at that intersection, the light is green about 35% of the time.

Question: If $P(\text{green}) = 0.35$, what's the probability the light isn't green when you get to College and Main?

"Not green" is the complement of "green," so $P(\text{not green}) = 1 - P(\text{green})$
$= 1 - 0.35 = 0.65$

There's a 65% chance I won't have a green light.



Two disjoint sets, **A** *and* **B**.

4. Suppose the probability that (**A**) a randomly selected student is a sophomore is 0.20, and the probability that (**B**) he or she is a junior is 0.30. What is the probability that the student is *either* a sophomore *or* a junior, written $P(\mathbf{A} \cup \mathbf{B})$? If you guessed 0.50, you've deduced the **Addition Rule,** which says that you can add the probabilities of events that are disjoint. To see whether two events are disjoint, we take them apart into their component outcomes and check whether they have any outcomes in common. **Disjoint** (or **mutually exclusive**) events have no outcomes in common. The **Addition Rule** states,

For two disjoint events A and B, the probability that one *or* the other occurs is the sum of the probabilities of the two events.

$$P(\mathbf{A} \cup \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B}), \text{ provided that A and B are disjoint.}$$

FOR EXAMPLE **Applying the Addition Rule**

Recap: When you get to the light at College and Main, it's either red, green, or yellow. We know that $P(\text{green}) = 0.35$.

Question: Suppose we find out that $P(\text{yellow})$ is about 0.04. What's the probability the light is red?

To find the probability that the light is green or yellow, I can use the Addition Rule because these are disjoint events: The light can't be both green and yellow at the same time.

$$P(\text{green} \cup \text{yellow}) = 0.35 + 0.04 = 0.39$$

Red is the only remaining alternative, and the probabilities must add up to 1, so

$$\begin{aligned} P(\text{red}) &= P(\text{not } (\text{green} \cup \text{yellow}) \\ &= 1 - P(\text{green} \cup \text{yellow}) \\ &= 1 - 0.39 = 0.61 \end{aligned}$$

"Baseball is 90% mental. The other half is physical."

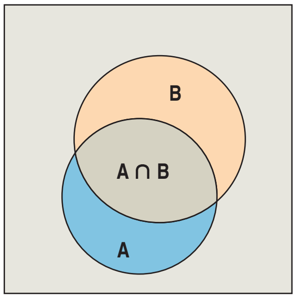
—Yogi Berra

A S ***Activity:*** **Addition Rule for Disjoint Events.** Experiment with disjoint events to explore the Addition Rule.

Because sample space outcomes are disjoint, we have an easy way to check whether the probabilities we've assigned to the possible outcomes are **legitimate.** The Probability Assignment Rule tells us that the sum of the probabilities of all possible outcomes must be exactly 1. No more, no less. For example, if we were told that the probabilities of selecting at random a freshman, sophomore, junior, or senior from all the undergraduates at a school were 0.25, 0.23, 0.22, and 0.20, respectively, we would know that something was wrong. These "probabilities" sum to only 0.90, so this is not a legitimate probability assignment. Either a value is wrong, or we just missed some possible outcomes, like "pre-freshman" or "postgraduate" categories that soak up the remaining 0.10. Similarly, a claim that the probabilities were 0.26, 0.27, 0.29, and 0.30 would be wrong because these "probabilities" sum to more than 1.

But be careful: The Addition Rule doesn't work for events that aren't disjoint. If the probability of owning an MP3 player is 0.50 and the probability of owning a computer is 0.90, the probability of owning either an MP3 player or a computer may be pretty high, but it is *not* 1.40! Why can't you add probabilities like this? Because these events are not disjoint. You *can* own both. In the next chapter, we'll see how to add probabilities for events like these, but we'll need another rule.

5. Suppose your job requires you to fly from Atlanta to Houston every Monday morning. The airline's Web site reports that this flight is on time 85% of the time. What's the chance that it will be on time two weeks in a row? That's the same as asking for the probability that your flight is on time this week *and* it's on time again next week. For independent events, the answer is very simple. Remember that independence means that the outcome of one event doesn't influence the outcome of the other. What happens with your flight this week doesn't influence whether it will be on time next week, so it's reasonable to assume that those events are independent. The **Multiplication Rule** says that for independent events, to find the probability that both events occur, we just multiply the probabilities together. Formally,



Two sets **A** *and* **B** *that are not disjoint. The event (***A** ∩ **B***) is their intersection.*

For two independent events A and B, the probability that both A *and* B occur is the product of the probabilities of the two events.

$$P(A \cap B) = P(A) \times P(B), \text{ provided that A and B are independent.}$$

A S *Activity:* **Multiplication Rule for Independent Events.** Experiment with independent random events to explore the Multiplication Rule.

A S *Activity:* **Probabilities of Compound Events.** The Random tool also lets you experiment with Compound random events to see if they are independent.

This rule can be extended to more than two independent events. What's the chance of your flight being on time for a month—four Mondays in a row? We can multiply the probabilities of it happening each week:

$$0.85 \times 0.85 \times 0.85 \times 0.85 = 0.522$$

or just over 50–50. Of course, to calculate this probability, we have used the assumption that the four events are independent.

Many Statistics methods require an **Independence Assumption,** but *assuming* independence doesn't make it true. Always *Think* about whether that assumption is reasonable before using the Multiplication Rule.

FOR EXAMPLE Applying the Multiplication Rule (and others)

Recap: We've determined that the probability that we encounter a green light at the corner of College and Main is 0.35, a yellow light 0.04, and a red light 0.61. Let's think about your morning commute in the week ahead.

Question: What's the probability you find the light red both Monday and Tuesday?

Because the color of the light I see on Monday doesn't influence the color I'll see on Tuesday, these are independent events; I can use the Multiplication Rule:

$$\begin{aligned} P(\text{red Monday} \cap \text{red Tuesday}) &= P(\text{Red}) \times P(\text{red}) \\ &= (0.61)(0.61) \\ &= 0.3721 \end{aligned}$$

There's about a 37% chance I'll hit red lights both Monday and Tuesday mornings.

Question: What's the probability you don't encounter a red light until Wednesday?

For that to happen, I'd have to see green or yellow on Monday, green or yellow on Tuesday, and then red on Wednesday. I can simplify this by thinking of it as not red on Monday and Tuesday and then red on Wednesday.

$$P(\text{not red}) = 1 - P(\text{red}) = 1 - 0.61 = 0.39, \text{ so}$$

$$\begin{aligned} P(\text{not red Monday} \cap \text{not red Tuesday} \cap \text{red Wednesday}) &= P(\text{not red}) \times P(\text{not red}) \times P(\text{red}) \\ &= (0.39)(0.39)(0.61) \\ &= 0.092781 \end{aligned}$$

There's about a 9% chance that this week I'll hit my first red light there on Wednesday morning.

Question: What's the probability that you'll have to stop *at least once* during the week?

Having to stop at least once means that I have to stop for the light either 1, 2, 3, 4, or 5 times next week. It's easier to think about the complement: never having to stop at a red light. Having to stop at least once means that I didn't make it through the week with no red lights.

$$\begin{aligned} &P(\text{having to stop at the light at least once in 5 days}) \\ &= 1 - P(\text{no red lights for 5 days in a row}) \\ &= 1 - P(\text{not red} \cap \text{not red} \cap \text{not red} \cap \text{not red} \cap \text{not red}) \\ &= 1 - (0.39)(0.39)(0.39)(0.39)(0.39) \\ &= 1 - 0.0090 \\ &= 0.991 \end{aligned}$$

There's over a 99% chance I'll hit at least one red light sometime this week.

Note that the phrase "at least" is often a tip-off to think about the complement. Something that happens *at least once* does happen. Happening at least once is the complement of not happening at all, and that's easier to find.

In informal English, you may see "some" used to mean "at least one." "What's the probability that some of the eggs in that carton are broken?" means at least one.

JUST CHECKING

2. Opinion polling organizations contact their respondents by telephone. Random telephone numbers are generated, and interviewers try to contact those households. In the 1990s this method could reach about 69% of U.S. households. According to the Pew Research Center for the People and the Press, by 2003 the contact rate had risen to 76%. We can reasonably assume each household's response to be independent of the others. What's the probability that . . .

a) the interviewer successfully contacts the next household on her list?

b) the interviewer successfully contacts both of the next two households on her list?

c) the interviewer's first successful contact is the third household on the list?

d) the interviewer makes at least one successful contact among the next five households on the list?

STEP-BY-STEP EXAMPLE Probability

The five rules we've seen can be used in a number of different combinations to answer a surprising number of questions. Let's try one to see how we might go about it.

In 2001, Masterfoods, the manufacturers of M&M's® milk chocolate candies, decided to add another color to the standard color lineup of brown, yellow, red, orange, blue, and green. To decide which color to add, they surveyed people in nearly every country of the world and asked them to vote among purple, pink, and teal. The global winner was purple!

In the United States, 42% of those who voted said purple, 37% said teal, and only 19% said pink. But in Japan the percentages were 38% pink, 36% teal, and only 16% purple. Let's use Japan's percentages to ask some questions:

1. What's the probability that a Japanese M&M's survey respondent selected at random preferred either pink or teal?
2. If we pick two respondents at random, what's the probability that they both selected purple?
3. If we pick three respondents at random, what's the probability that *at least one* preferred purple?

THINK

The probability of an event is its long-term relative frequency. It can be determined in several ways: by looking at many replications of an event, by deducing it from equally likely events, or by using some other information. Here, we are told the relative frequencies of the three responses.

The M&M's Web site reports the proportions of Japanese votes by color. These give the probability of selecting a voter who preferred each of the colors:

$P(\text{pink}) = 0.38$
$P(\text{teal}) = 0.36$
$P(\text{purple}) = 0.16$

Make sure the probabilities are legitimate. Here, they're not. Either there was a mistake, or the other voters must have chosen a color other than the three given. A check of the reports from other countries shows a similar deficit, so probably we're seeing those who had no preference or who wrote in another color.

Each is between 0 and 1, but they don't all add up to 1. The remaining 10% of the voters must have not expressed a preference or written in another color. I'll put them together into "no preference" and add $P(\text{no preference}) = 0.10$.

With this addition, I have a legitimate assignment of probabilities.

Question 1. What's the probability that a Japanese M&M's survey respondent selected at random preferred either pink or teal?

THINK	**Plan** Decide which rules to use and check the conditions they require.	The events "Pink" and "Teal" are individual outcomes (a respondent can't choose both colors), so they are disjoint. I can apply the Addition Rule.
SHOW	**Mechanics** Show your work.	$P(\text{pink} \cup \text{teal}) = P(\text{pink}) + P(\text{teal})$ $= 0.38 + 0.36 = 0.74$
TELL	**Conclusion** Interpret your results in the proper context.	The probability that the respondent said pink or teal is 0.74.

Question 2. If we pick two respondents at random, what's the probability that they both said purple?

THINK	**Plan** The word "both" suggests we want $P(\mathbf{A}\ and\ \mathbf{B})$, which calls for the Multiplication Rule. Think about the assumption.	✓ **Independence Assumption:** It's unlikely that the choice made by one random respondent affected the choice of the other, so the events seem to be independent. I can use the Multiplication Rule.
SHOW	**Mechanics** Show your work. For both respondents to pick purple, each one has to pick purple.	$P(\text{both purple})$ $= P(\text{first respondent picks purple} \cap \text{second respondent picks purple})$ $= P(\text{first respondent picks purple}) \times P(\text{second respondent picks purple})$ $= 0.16 \times 0.16 = 0.0256$
TELL	**Conclusion** Interpret your results in the proper context.	The probability that both respondents pick purple is 0.0256.

Question 3. If we pick three respondents at random, what's the probability that at least one preferred purple?

THINK

Plan The phrase "at least . . ." often flags a question best answered by looking at the complement, and that's the best approach here. The complement of "At least one preferred purple" is "None of them preferred purple."

$P(\text{at least one picked purple})$
$= P(\{\text{none picked purple}\}^C)$
$= 1 - P(\text{none picked purple}).$
$= 1 - P(\text{not purple} \cap \text{not purple} \cap \text{not purple}).$

Think about the assumption.

✔ **Independence Assumption:** These are independent events because they are choices by three random respondents. I can use the Multiplication Rule.

SHOW

Mechanics First we find *P*(not purple) with the Complement Rule.

$P(\text{not purple}) = 1 - P(\text{purple})$
$= 1 - 0.16 = 0.84$

Next we calculate *P*(none picked purple) by using the Multiplication Rule.

$P(\text{at least one picked purple})$
$= 1 - P(\text{none picked purple})$
$= 1 - P(\text{not purple} \cap \text{not purple} \cap \text{not purple})$
$= 1 - (0.84)(0.84)(0.84)$

Then we can use the Complement Rule to get the probability we want.

$= 1 - 0.5927$
$= 0.4073$

TELL

Conclusion Interpret your results in the proper context.

There's about a 40.7% chance that at least one of the respondents picked purple.

WHAT CAN GO WRONG?

- **Beware of probabilities that don't add up to 1.** To be a legitimate probability assignment, the sum of the probabilities for all possible outcomes must total 1. If the sum is less than 1, you may need to add another category ("other") and assign the remaining probability to that outcome. If the sum is more than 1, check that the outcomes are disjoint. If they're not, then you can't assign probabilities by just counting relative frequencies.
- **Don't add probabilities of events if they're not disjoint.** Events must be disjoint to use the Addition Rule. The probability of being under 80 *or* a female is not the probability of being under 80 *plus* the probability of being female. That sum may be more than 1.
- **Don't multiply probabilities of events if they're not independent.** The probability of selecting a student at random who is over 6′10″ tall *and* on the basketball team is *not* the probability the student is over 6′10″ tall *times* the probability he's on the basketball team. Knowing that the student is over 6′10″ changes the probability of his being on the basketball team. You can't multiply these probabilities. The multiplication of probabilities of events that are not independent is one of the most common errors people make in dealing with probabilities.
- **Don't confuse disjoint and independent.** Disjoint events *can't* be independent. If **A** = {you get an A in this class} and **B** = {you get a B in this class}, **A** and **B** are disjoint. Are they independent? If you find out that **A** is true, does that change the probability of **B**? You bet it does! So they can't be independent. we'll return to this issue in the next chapter.

CONNECTIONS

We saw in the previous three chapters that randomness plays a critical role in gathering data. That fact alone makes it important that we understand how random events behave. The rules and concepts of probability give us a language to talk and think about random phenomena. From here on, randomness will be fundamental to how we think about data, and probabilities will show up in every chapter.

We began thinking about independence back in Chapter 3 when we looked at contingency tables and asked whether the distribution of one variable was the same for each category of another. Then, in Chapter 12, we saw that independence was fundamental to drawing a Simple Random Sample. For computing compound probabilities, we again ask about independence. And we'll continue to think about independence throughout the rest of the book.

Our interest in probability extends back to the start of the book. We've talked about "relative frequencies" often. But—let's be honest—that's just a casual term for probability. For example, you can now rephrase the 68–95–99.7 Rule to talk about the *probability* that a random value selected from a Normal model will fall within 1, 2, or 3 standard deviations of the mean.

Why not just say "probability" from the start? Well, we didn't need any of the formal rules of this chapter (or the next one), so there was no point to weighing down the discussion with those rules. And "relative frequency" is the right intuitive way to think about probability in this course, so you've been thinking right all along.

Keep it up.

WHAT HAVE WE LEARNED?

We've learned that probability is based on long-run relative frequencies. We've thought about the Law of Large Numbers and noted that it speaks only of long-run behavior. Because the long run is a very long time, we need to be careful not to misinterpret the Law of Large Numbers. Even when we've observed a string of heads, we shouldn't expect extra tails in subsequent coin flips.

Also, we've learned some basic rules for combining probabilities of outcomes to find probabilities of more complex events. These include

- the Probability Assignment Rule,
- the Complement Rule,
- the Addition Rule for disjoint events, and
- the Multiplication Rule for independent events.

Terms

Random phenomenon 324. A phenomenon is random if we know what outcomes could happen, but not which particular values will happen.

Trial 325. A single attempt or realization of a random phenomenon.

Outcome 325. The outcome of a trial is the value measured, observed, or reported for an individual instance of that trial.

Event 325. A collection of outcomes. Usually, we identify events so that we can attach probabilities to them. We denote events with bold capital letters such as **A, B,** or **C**.

Sample Space 325. The collection of all possible outcome values. The sample space has a probability of 1.

Law of Large Numbers 326. The Law of Large Numbers states that the long-run *relative frequency* of repeated independent events gets closer and closer to the *true* relative frequency as the number of trials increases.

Independence (informally) 326. Two events are *independent* if learning that one event occurs does not change the probability that the other event occurs.

Probability 326. The probability of an event is a number between 0 and 1 that reports the likelihood of that event's occurrence. We write $P(\mathbf{A})$ for the probability of the event **A.**

Empirical probability 326. When the probability comes from the long-run relative frequency of the event's occurrence, it is an **empirical probability.**

Theoretical probability 327. When the probability comes from a model (such as equally likely outcomes), it is called a **theoretical probability.**

Personal probability 328. When the probability is subjective and represents your personal degree of belief, it is called a **personal probability.**

The Probability Assignment Rule 330. The probability of the entire sample space must be 1. $P(\mathbf{S}) = 1$.

Complement Rule 330. The probability of an event occurring is 1 minus the probability that it doesn't occur.

$$P(\mathbf{A}) = 1 - P(\mathbf{A^C})$$

Disjoint (Mutually exclusive) 330. Two events are disjoint if they share no outcomes in common. If **A** and **B** are disjoint, then knowing that **A** occurs tells us that **B** cannot occur. Disjoint events are also called "mutually exclusive."

Addition Rule 330. If **A** and **B** are disjoint events, then the probability of **A** *or* **B** is

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

Legitimate probability assignment 331. An assignment of probabilities to outcomes is legitimate if

- each probability is between 0 and 1 (inclusive).
- the sum of the probabilities is 1.

Multiplication Rule 331. If **A** and **B** are independent events, then the probability of **A** *and* **B** is

$$P(\mathbf{A} \cap \mathbf{B}) = P(\mathbf{A}) \times P(\mathbf{B}).$$

Independence Assumption 332. We often require events to be independent. (So you should think about whether this assumption is reasonable.)

Skills

THINK

- Understand that random phenomena are unpredictable in the short term but show long-run regularity.
- Be able to recognize random outcomes in a real-world situation.
- Know that the relative frequency of a random event settles down to a value called the (empirical) probability. Know that this is guaranteed for independent events by the Law of Large Numbers.
- Know the basic definitions and rules of probability.
- Recognize when events are disjoint and when events are independent. Understand the difference and that disjoint events cannot be independent.

SHOW

- Be able to use the facts about probability to determine whether an assignment of probabilities is legitimate. Each probability must be a number between 0 and 1, and the sum of the probabilities assigned to all possible outcomes must be 1.
- Know how and when to apply the Addition Rule. Know that events must be disjoint for the Addition Rule to apply.
- Know how and when to apply the Multiplication Rule. Know that events must be independent for the Multiplication Rule to apply. Be able to use the Multiplication Rule to find probabilities for combinations of independent events.
- Know how to use the Complement Rule to make calculating probabilities simpler. Recognize that probabilities of "at least. . ." are likely to be simplified in this way.

TELL

- Be able to use statements about probability in describing a random phenomenon. You will need this skill soon for making statements about statistical inference.
- Know and be able to use the terms "sample space", "disjoint events", and "independent events" correctly.