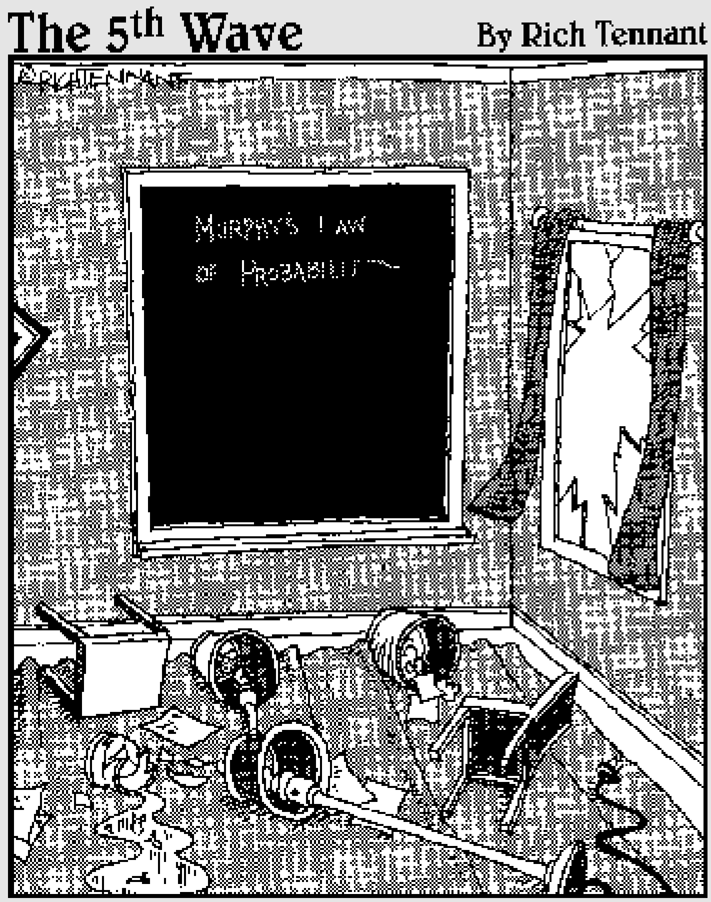
Part I

The Certainty of Uncertainty: Probability Basics



In this part . . .

In Part I, you get started with the basics of probability — the terminology, the basic ideas of finding a probability, and, perhaps most importantly, how to organize and set up all the information you have in order to successfully calculate a probability. You also discover ways in which people use probability in the real world.

But let's be honest. When it comes to a class that involves probability, is there truly a *real* world? Maybe, maybe not. Counting the number of ways to pick three green balls and four red balls from an urn that contains twenty green balls and thirty red balls doesn't sound all that relevant — and it isn't. That's why you won't see a single "urn problem" anywhere in this part. However, if you do run across an "urn problem" in your life, you'll know how to answer it, using the techniques from Part I.

Chapter 1

The Probability in Everyday Life

In This Chapter

- Recognizing the prevalence and impact of probability in your everyday life
- Taking different approaches to finding probabilities
- Steering clear of common probability misconceptions

You've heard it, thought it, and said it before: "What are the odds of that happening?" Someone wins the lottery not once, but twice. You accidentally run into a friend you haven't seen since high school during a vacation in Florida. A cop pulls you over the one time you forget to put your seatbelt on. And you wonder . . . what *are* the odds of this happening? That's what this book is about: figuring, interpreting, and understanding how to quantify the random phenomena of life. But it also helps you realize the limitations of probability and why probabilities can take you only so far.

In this chapter, you observe the impact of probability on your everyday life and some of the ways people come up with probabilities. You also find out that with probability, situations aren't always what they seem.

Figuring Out what Probability Means

Probabilities come in many different disguises. Some of the terms people use for probability are *chance, likelihood, odds, percentage,* and *proportion.* But the basic definition of *probability* is the long-term chance that a certain outcome will occur from some random process. A probability is a number between zero and one — a proportion, in other words. You can write it as a percentage, because people like to talk about probability as a percentage chance, or you can put it in the form of odds. The term "odds," however, isn't exactly the same as probability. *Odds* refers to the ratio of the denominator of a probability to the numerator of a probability. For example, if the probability of a horse winning a race is 50 percent (½), the odds of this horse winning are 2 to 1.

Understanding the concept of chance

The term *chance* can take on many meanings. It can apply to an individual ("What are my chances of winning the lottery?"), or it can apply to a group ("The overall percentage of adults who get cancer is . . ."). You can signify a chance with a percent (80 percent), a proportion (0.80), or a word (such as "likely"). The bottom line of all probability terms is that they revolve around the idea of a long-term chance. When you're looking at a random process (and most occurrences in the world are the results of random processes for which the outcomes are never certain), you know that certain outcomes can happen, and you often weigh those outcomes in your mind. It all comes down to long-term chance; what's the chance that this or that outcome is going to occur in the long term (or over many individuals)?

If the chance of rain tomorrow is 30 percent, does that mean it won't rain because the chance is less than 50 percent? No. If the chance of rain is 30 percent, a meteorologist has looked at many days with similar conditions as tomorrow, and it rained on 30 percent of those days (and didn't rain the other 70 percent). So, a 30-percent chance for rain means only that it's unlikely to rain.

Interpreting probabilities: Thinking large and long-term

You can interpret a probability as it applies to an individual or as it applies to a group. Because probabilities stand for long-term percentages (see the previous section), it may be easier to see how they apply to a group rather than to an individual. But sometimes one way makes more sense than the other, depending on the situation you face. The following sections outline ways to interpret probabilities as they apply to groups or individuals so you don't run into misinterpretation problems.

Playing the instant lottery

Probabilities are based on long-term percentages (over thousands of trials), so when you apply them to a group, the group has to be large enough (the larger the better, but at least 1,500 or so items or individuals) for the probabilities to really apply. Here's an example where long-term interpretation makes sense in place of short-term interpretation. Suppose the chance of winning a prize in an instant lottery game is $\frac{1}{10}$, or 10 percent. This probability means that in the long term (over thousands of tickets), 10 percent of all instant lottery tickets purchased for this game will win a prize, and 90 percent won't. It doesn't mean that if you buy 10 tickets, one of them will automatically win.

If you buy many sets of 10 tickets, on average, 10 percent of your tickets will win, but sometimes a group of 10 has multiple winners, and sometimes it has no winners. The winners are mixed up amongst the total population of tickets. If you buy exactly 10 tickets, each with a 10 percent chance of winning, you might expect a high chance of winning at least one prize. But the chance of you winning at least one prize with those 10 tickets is actually only 65 percent, and the chance of winning nothing is 35 percent. (I calculate this probability with the binomial model; see Chapter 8).

Pondering political affiliation

You can use the following example as an illustration of the limitation of probability — namely that actual probability often applies to the percentage of a large group. Suppose you know that 60 percent of the people in your community are Democrats, 30 percent are Republicans, and the remaining 10 percent are Independents or have another political affiliation. If you randomly select one person from your community, what's the chance the person is a Democrat? The chance is 60 percent. You can't say that the person is surely a Democrat because the chance is over 50 percent; the percentages just tell you that the person is more likely to be a Democrat. Of course, after you ask the person, he or she is either a Democrat or not; you can't be 60-percent Democrat.

Seeing probability in everyday life

Probabilities affect the biggest and smallest decisions of people's lives. Pregnant women look at the probabilities of their babies having certain genetic disorders. Before you sign the papers to have surgery, doctors and nurses tell you about the chances that you'll have complications. And before you buy a vehicle, you can find out probabilities for almost every topic regarding that vehicle, including the chance of repairs becoming necessary, of the vehicle lasting a certain number of miles, or of you surviving a front-end crash or rollover (the latter depends on whether you wear a seatbelt — another fact based on probability).

While scanning the Internet, I quickly found several examples of probabilities that affect people's everyday lives — two of which I list here:

- **Distributing prescription medications in specially designed blister packages rather than in bottles may increase the likelihood that consumers will take the medication properly, a new study suggests. (Source: Ohio State University Research News, June 20, 2005)**

  In other words, the probability of consumers taking their medications properly is higher if companies put the medications in the new packaging than it is when the companies put the medicines in bottles. You don't know what the probability of taking those medications correctly was originally or how much the probability increases with this new packaging, but you do know that according to this study, the packaging is having some effect.

- **According to State Farm Insurance, the top three cities for auto theft in Ohio are Toledo (580.23 thefts per 100,000 vehicles), Columbus (558.19 per 100,000), and Dayton-Springfield (525.06 per 100,000).**

  The information in this example is given in terms of rate; the study recorded the number of cars stolen each year in various metropolitan areas of Ohio. Note that the study reports the information as the number of thefts per 100,000 vehicles. The researchers needed a fixed number of vehicles in order to be fair about the comparison. If the study used only the number of thefts, cities with more cars would always rank higher than cities with fewer cars.

  How did the researchers get the specific numbers for this study? They took the actual number of thefts and divided it by the total number of vehicles to get a very small decimal value. They multiplied that value by 100,000 to get a number that's fair for comparison. To write the rates as probabilities, they simply divided them by 100,000 to put them back in decimal form. For Toledo, the probability of car theft is $580.23 \div 100{,}000 = 0.0058023$, or 0.58 percent; for Columbus, the probability of car theft is 0.0055819, or 0.56 percent; and for Dayton-Springfield, the probability is 0.0052506, or 0.53 percent.

Be sure to understand exactly what format people use to discuss or report a probability, and be sure that the format allows for a fair and equitable comparison.

Coming Up with Probabilities

You can figure or compute probabilities in a variety of ways, depending on the complexity of the situation and what exactly is possible to quantify. Some probabilities are very difficult to figure, such as the probability of a tropical storm developing into a hurricane that will ultimately make landfall at a certain place and time — a probability that depends on many elements that are themselves nearly impossible to determine. If people calculate actual probabilities for hurricane outcomes, they make estimates at best.

Some probabilities, on the other hand, are very easy to calculate for an exact number, such as the probability of a fair die landing on a 6 (1 out of 6, or 0.167). And many probabilities are somewhere in between the previous two examples in terms of how difficult it is to pinpoint them numerically, such as the probability of rain falling tomorrow in Seattle. For middle-of-the-road probabilities, past data can give you a fairly good idea of what's likely to happen.

After you analyze the complexity of the situation, you can use one of four major approaches to figure probabilities, each of which I discuss in this section.

Be subjective

The subjective approach to probability is the most vague and the least scientific. It's based mostly on opinions, feelings, or hopes, meaning that you typically don't use this type of probability approach in real scientific endeavors. You basically say, "Here's what I think the probability is." For example, although the actual, true probability that the Ohio State football team will win the national championship is out there somewhere, no one knows what it is, even though every fan and analyst will have ideas about what that chance is, based on everything from dreams they had last night, to how much they love or hate Ohio State, to all the statistics from Ohio State football over the last 100 years. Other people will take a slightly more scientific approach — evaluating players' stats, looking at the strength of the competition, and so on. But in the end, the probability of an event like this is mostly subjective, and although this approach isn't scientific, it sure makes for some great sports talk amongst the fans!

Take a classical approach

The classical approach to probability is a mathematical, formula-based approach. You can use math and counting rules to calculate exact probabilities in many cases (for more on the counting rules, see Chapter 5). Anytime you have a situation where you can enumerate the possible outcomes and figure their individual probabilities by using math, you can use the classical approach to getting the probability of an outcome or series of outcomes from a random process.

For example, when you roll two die, you have six possible outcomes for the first die, and for each of those outcomes, you have another six possible outcomes for the second die. All together, you have $6 * 6 = 36$ possible outcomes for the pair. In order to get a sum of two on a roll, you have to roll two 1s, meaning it can happen in only one way. So, the probability of getting a sum of two is $1/36$. The probability of getting a sum of three is $2/36$, because only two of the outcomes result in a sum of three: 1-2 or 2-1. A sum of seven has a probability of $6/36$, or $1/6$ — the highest probability of any sum of two die. Why is seven the sum with highest probability? Because it has the most possible ways of coming up: 1-6, 2-5, 3-4, 4-3, 5-2, and 6-1. That's why the number seven is so important in the gambling game craps. (For more on this example, see Chapter 2.)

You also use the classical approach when you make certain assumptions about a random process that's occurring. For example, if you can assume that the probability of achieving success when you're trying to make a sale is the same on each trial, you can use the binomial probability model for figuring out the probability of making 5 sales in 20 tries. Many types of probability models are available, and I discuss many of them in this book. (For more on the binomial probability model, see Chapter 8.)

The classical approach doesn't work when you can't describe the possible individual outcomes and come up with some mathematical way of determining the probabilities. For example, if you have to decide between different brands of refrigerators to buy, and your criterion is having the least chance of needing repairs in the next five years, the classical approach can't help you for a couple reasons. First, you can't assume that the probability of a refrigerator needing one repair is the same as the probability of needing two, three, or four repairs in five years. Second, you have no math formula to figure out the chances of repairs for different brands of refrigerators; it depends on past data that's been collected regarding repairs.

Find relative frequencies

In cases where you can't come up with a mathematical formula or model to figure a probability, the relative frequency approach is your best bet. The approach is based on collecting data and, based on that data, finding the percentage of time that an event occurred. The percentage you find is the *relative frequency* of that event — the number of times the event occurred divided by the total number of observations made. (You can find the probabilities for the refrigerator repairs example in the previous section with the relative frequency approach by collecting data on refrigerator repair records.)

Suppose, for example, that you're watching your birdfeeder, and you notice a lot of cardinals coming for dinner. You want to find the probability that the next bird that comes to the feeder is a cardinal. You can estimate this probability by counting the number of birds that come to your feeder over a period of time and noting how many cardinals you see. If you count 100 bird visits, and 27 of the visitors are cardinals, you can say that for the period of time you observe, 27 out of 100 visits — or 27 percent, the relative frequency — were made by cardinals. Now, if you have to guess the probability that the next bird to visit is a cardinal, 27 percent would be your best guess. You come up with a probability based on relative frequency.

PROBABILITY IN ACTION

Consuming data with *Consumer Reports*

The magazine *Consumer Reports* — put out by the Consumers Union, a nonprofit group that helps provide consumer protection information — does thousands of studies to test different makes and models of products so it can report on how safe, reliable, effective, and efficient the models are, along with how much they cost. In the end, the group comes up with a list of recommendations regarding which models are the best values for your money. *Consumer Reports* bases its reports on a relative frequency approach. For example, when comparing refrigerators, it tests various models for energy efficiency, temperature performance, noise, ease of use, and energy cost per year. The researchers figure out what percentage of time the refrigerators need repairs, don't perform properly, and so on, and they base their reports on what they find.

REMEMBER

A limitation of the relative frequency approach is that the probabilities you come up with are only estimates because you base them on finite samples of data you collect. And those estimates are only as good as the data that you collect. For example, if you collected your birdfeeder data when you offered sunflower seeds, but now you offer thistle seed (loved by smaller birds), your probability of seeing a cardinal changes. Also, if you look at the feeder only at 5 p.m. each day, when cardinals are more likely to be out than any other bird, your predictions work only at that same time period, not at noon when all the finches are also out and about. The issue of collecting good data is a statistical one; see *Statistics For Dummies* (Wiley) for more information.

Use simulations

The simulation approach is a process that creates data by setting up a certain scenario, playing out that scenario over and over many times, and looking at the percentage of times a certain outcome occurs. It may sound like the relative frequency approach (see the previous section), but it's different in three ways:

- You create the data (usually with a computer); you don't collect it out in the real world.
- The amount of data is typically much larger than the amount you could observe in real life.
- You use a certain model that scientists come up with, and models have assumptions.

PROBABILITY IN ACTION

Tracking down hurricanes

One major area where professionals use computer models is in predicting the arrival, intensity, and path of tropical storms, including hurricanes. Computer hurricane models help scientists and leaders perform integrated cost-benefit studies; evaluate the effects of regulatory policies; and make decisions during crises. Insurance companies use the models to make predictions regarding the number of and estimated damage due to future hurricanes, which helps them adjust their premiums appropriately to be ready to pay out the huge claims that come with large hurricanes.

Computer models for tropical storms are best at predicting long-run (versus short-term) losses across large (versus small) geographic areas, due to the high margin of error. *Margin of error* is the amount by which your results are expected to change from sample to sample. You can't look at a single storm and say exactly what's going to happen. AIR Worldwide, whose computer models are used by half the residential property insurance markets in Florida and 85 percent of the companies that underwrite insurers, calculates projections over storms across a 50,000-year span. Another modeling expert recently lengthened its computer modeling from 100,000 to 300,000 years to get results within an acceptable margin of error.

The models contain so many variables that it takes many trials to approach a predictable average. Flipping a coin, for instance, has only one variable with two outcomes. If you want to estimate the probability of flipping heads by using a model, it takes about 2,500 trials to get a result within a 2-percent margin of error. The more variables, the more trials required to get a dependable outcome. And with hurricanes, the number of variables is huge. The computer models used by the National Hurricane Center include variables such as the initial latitude and longitude of the storm, the components of the "storm motion vector," and the initial storm intensity, just to name a few.

You can see an example of a simulation if you let a computer play out a game of chance for you. You can tell it to credit you with $1.00 if a head comes up on a coin flip and deduct $1.00 if a tail comes up. Repeat the bet thousands of times and see what you end up with. Change the probabilities of heads and tails to see what happens. Your experiments are examples of simple simulations.

One commonality between simulations and the relative frequency approach is that your results are only as good as the data you come up with. I remember very clearly a simulation that a student performed to predict which team would win the NCAA basketball tournament some years ago. The student gave each of the 64 teams in the tournament a probability of winning its game based on certain statistics that the sports gurus came up with. The student fed those probabilities into the computer and made the computer repeat the tournament over and over millions of times, recording who won each game and who won the entire tournament. On 96 percent of the simulations, Duke University won the whole thing. So, of course, it seemed as if Duke was a shoe-in that year. Guess how long Duke actually lasted? The team went down in the second of six rounds.

Probability Misconceptions to Avoid

No matter how researchers calculate a probability or what kind of information or data they base it on, the probability is often misinterpreted or applied in the wrong way by the media, the public, and even other researchers who don't quite understand the limitations of probability. The main idea is that probability often goes against your intuition, and you have to be very careful about not letting your intuition get the better of you when thinking in terms of probability. This section highlights some of the most common misconceptions about probability.

Thinking in 50-50 terms when you have two outcomes

Resist the urge to think that a situation with only two possible outcomes is a 50-50 situation. The only time a situation with two possible outcomes is a 50-50 proposition is when both outcomes are equally likely to occur, as in the flip of a fair coin.

I often ask students to tell me what they think the probability is that a basketball player will make a free throw. Most students tell me the probability depends on the player and his or her free-throw percentage (number of made shots divided by the number of attempts). For example, basketball professional Shaquille O'Neal's career best is 62 percent, shot in the 2002-2003 season. When Shaq stepped up to the line that season, he made his free throws 62 percent of the time, and he missed them 38 percent of the time. At any particular moment during that season when he was standing at the line to make a free throw, the chance of him making it was 62 percent. However, a few students look at me and say, "Wait a minute. He either makes it or he doesn't. So, shouldn't his chance be 50-50?"

If you look at it from a strictly basketball point of view, that reasoning doesn't make sense, because everyone would be a 50-percent free throw shooter — no more, no less — including people who don't even play basketball! The probability of making a free throw on your next try is based on a relative frequency approach (see the section "Find relative frequencies" earlier in this chapter) — it depends on what percentage you've made over the long haul, and that depends on many factors, not chance alone.

However, if you look at the situation from a probability point of view, it may be hard to escape this misconception. After all, you have two outcomes: make it or miss it. If you flip a coin, the probability of getting a head is 50 percent, and the probability of getting a tail is 50 percent, so why doesn't this hold true for free throws? Because free throws aren't set up like a fair coin. Fair coins are equally likely to turn up heads or tails, and unless your free-throw percentage is exactly 50 percent, you don't shoot free throws like you toss coins.

Thinking that patterns can't occur

WARNING!

What you perceive as random and what's actually random are two different things. Be careful not to misinterpret outcomes by identifying them as being less probable because they don't look random enough. In other words, don't rule out the fact that patterns can and do occur over the long term, just by chance.

The most important idea here is to not let your intuition get in the way of reality. Here are two examples to help you recognize what's real and what's not when it comes to probability.

Picking a number from one to ten

Suppose that you ask a group of 100 people to pick a number from one to ten. (Go ahead and pick a number before reading on, just for fun.) You should expect about ten people to pick one, ten people to pick two, and so on (not exactly, but fairly close). What happens, however, is that more people pick either three or seven than the other numbers. (Did you?) Why is this so? Because most people don't want to pick one or ten because these numbers are on the ends, and they don't want to pick five because it rests in the middle, so they go for numbers that *appear* more random — the middle of the numbers from one to five (which is three) and the middle of the numbers from five to ten (which is seven). So, you throw the assumption that all ten numbers are equally likely for selection out the window because people don't think as objectively as real random numbers do!

TIP

Research has shown that people can't be objective enough in choosing random numbers, so to be sure that your probabilities can be repeated, you need to make sure that you base them on random processes where each individual outcome has an equal chance of selection. If you put the numbers in a hat, shake, and pull one out, you create a random process.

Flipping a coin ten times

Suppose that you flip a coin ten times and get the following result: H, T, H, T, T, T, T, T, T, H. People who see your recorded outcome may think that you made up the results, because "you just don't get six tails in a row." Observers may think your outcome just doesn't look random enough. Their intuition fuels their doubts, but their intuition is wrong. In fact, you're very likely to have *runs* of heads or tails amongst a data set.

If you flip a coin ten times, with two possible outcomes on each flip, you have $2 * 2 * 2 * 2 * 2 * 2 * 2 * 2 * 2 * 2 = 1{,}024$ possible outcomes, each one being equally likely. Your outcome with the coin is just as likely as one that may look to be more random: H, T, H, T, H, T, H, T, H, T.

Chapter 2

Coming to Terms with Probability

In This Chapter

- Nailing down the basic definitions and terms associated with probability
- Examining how probability relates different events
- Solving probability problems with the rules and formulas of probability
- Identifying independent and mutually exclusive events
- Exploring the difference between independence and exclusivity

The first step toward probability success is having a clear knowledge of the terms, the notation, and the different types of probabilities you come across. If you use and understand the terms, notation, and types when working on easy problems, you have an edge from the start when the problems get more complex. This chapter sets you on the right track.

A Set Notation Overview

Probability has its own set of notations, symbols, and definitions that provide a shorthand way of expressing what you want to do. *Notation* refers to the symbols that you use as shorthand to talk about probability; for example, P(A) means the probability that A will occur. *Definition* refers to the statistical meanings of the terms used in probability. Every probability problem starts out by defining the information you have and the quantity you're trying to get, which all comes down to notation and terms.

Noting outcomes: Sample spaces

A probability is the chance that a certain *outcome,* or result, will occur out of all the possible outcomes for the process at hand. The process is called a *random process* because you conduct an experiment, or other form of data collection, and you don't know how the results will come out. Before you can figure out the probability of the result you're interested in, you list all the possible outcomes; this list is called the *sample space* and is typically denoted by S.

Any collection of items in probability is called a *set.* Notice that S is a set, so you use set notation to list its outcomes and probabilities for those outcomes (such as using brackets around the list with commas that separate each outcome).

For example, if your random process is rolling a single die, S = {1, 2, 3, 4, 5, 6} denotes the sample space. The set S can take on three different types: finite, countably infinite, and uncountably infinite.

Finite samples spaces

If you can write and count all the elements in a set, the set is *finite.* Rolling a single die is an example of a finite random process because you can achieve only six possible outcomes, and you can account for them all. Probability models that you can use for finite sample spaces include the binomial (see Chapter 8), the discrete uniform (see Chapter 7), and the hypergeometric (see Chapter 16).

Countably infinite sample spaces

Countably infinite means that you have a way to show the progression of the values, but they can go on into infinity. For example, if your random process involves the number of phone calls that come in to a switchboard during a week's time, the possible outcomes of S aren't finite, but rather countably infinite. In this case, S = {0, 1, 2, 3, 4 . . .}. S goes to infinity because you can't be sure of the maximum number of calls coming in. If you count all the calls, you get a fixed number with a countably infinite sample space S, but to be sure you allow for any maximum, you let S go on to infinity.

TIP

The way to get around the strange countably infinite situation is to give progressively smaller and smaller probabilities to the larger and larger values of S so eventually they become irrelevant. (More on this probability model in Chapter 13.)

Uncountably infinite sample spaces

Uncountably infinite means that you have situations where the possible outcomes are too numerous to write down in a listing, so you include an interval to describe them. An *interval* is a subset of the number line that falls entirely between two values — [1, 2] is the set of all real numbers between 1 and 2. It may seem weird to have an uncountably infinite set consisting of numbers between 1 and 2, but there are too many numbers in this interval to count!

For a real-world example, imagine that you're measuring the lengths of time it could take a computer to complete a task, and the maximum allowable time is 5 seconds. Your measurements of the actual time taken could be anywhere from 0 to 5 seconds (excluding 0) and to an infinite number of decimal places. You denote that as S = {all real numbers x such that $0 < x \le 5$}. For this example, S is uncountably infinite. (For more on this probability model, see Chapter 18.)

Noting subsets of sample spaces: Events

Probability problems typically involve figuring the probability of one or more subsets of the sample space, S. A subset of the sample space S is called an *event,* and the notations for events are capital letters: A, B, C, D, and so on. For example, if you roll a single die, S = {1, 2, 3, 4, 5, 6}. Event A may be the event that you roll an odd number: A = {1, 3, 5}. Event B may be the event that you roll a number greater than 2: B = {3, 4, 5, 6}.

If you have to monitor calls that come in to a switchboard for a week's time, you may be interested in the event that at least 10 calls come in (call it event E): E = {10, 11, 12 . . .}. If you have to monitor the time it takes a computer to complete tasks up to 5 seconds, you may want to find the probability that it takes no more than 4 seconds to complete the task (call this event D): D = {all real numbers x such that $0 < x \leq 4$}.

Translating inequalities

Many probabilities involve a series of outcomes described by phrases such as "at least," "at most," "not more than," "not less than," "more than," or "less than." You need to understand exactly what these phrases mean and be able to translate them into math symbols.

- *At least* means greater than or equal to and is denoted by $\geq$. For example, rolling at least a 3 on a fair die means $x \geq 3$, where x represents the outcomes from S that you're interested in: 3, 4, 5, or 6. Or you may be looking at grade point average (on a 4.00 scale), where at least 3.00 means all possible numbers from 3.00 to 4.00, including 3.00. Whether the statement refers to integers or all real numbers depends on the type of sample space you're dealing with.
- *At most* means you can go up to the number in question but not beyond it, so the notation is $\leq$. To roll a die and get a number that's at most 3 means $x \leq 3$: 1, 2 or 3. If you're looking at grade point average, at most 3.00 means anything from 0.00 to 3.00, including 3.00.
- *Not more than* means you can have the number in question or anything less, but you can't have more. It means the same as less than or equal to or at most.
- *Not less than* has the same meaning as greater than or equal to or at least.
- Strictly *less than* means you don't want to include the number in question itself, so you use <; for example, less than 3 on a single die means {1, 2}. For grade point average, less than 3.00 means everything from 0.00 to 3.00 but not including 3.00.
- Strictly *greater than* means you don't want to include the number in question, but you include every number beyond it, so you use >. To roll a die and get a number greater than 3 is to roll {4, 5, 6}. A grade point average of greater than 3.00 means every average from 3.00 to 4.00, not including 3.00.

The easiest way to remember these phrases and their notations is to think of easy-to-remember examples that make sense to you. For example, to remember how "at least" works, you can recall that you need to be at least 21 to go to a bar, which means {$x \geq 21$}, where x is your age.

TIP

You can simplify set notation by using *interval notation,* which indicates you want an interval of numbers on the number line. In interval notation, you write the left endpoint first, a comma, and then the right endpoint. You indicate whether you want to include the two endpoints by the type of brackets you use:

- If you want to include the endpoint in a set, you use a square bracket.
- If you don't want to include the endpoint in a set, you use a round bracket.

In the previous computer example, you have S = (0, 4]. The square bracket indicates that you do include the endpoint 4, and the round bracket indicates that you don't include 0. You include every number in between.

REMEMBER

The type of notation that's used to indicate an interval is totally up to the instructor. Both types — the one involving brackets, such as [0, 2], and the one involving inequalities that involve x, such as $0 \le x \le 2$ — are used very commonly, so you should get used to both notations. I typically use the inequality notation throughout this book.

Noting a void in the set: Empty sets

The last basic probability definition is the empty set, or null set. If an event, or subset of the sample space S, doesn't have any outcomes in it, you have an *empty set,* or *null set.* The most common notation used for the empty set is ø, although you sometimes see the notation { }. You may see an empty set if you're looking for elements that are common to two sets, and you don't find any. For example, let A = {1, 2, 3} and B = {4, 5, 6}. Which outcomes are common to both sets? None. How do you indicate this, using set notation? The set of all outcomes common to sets A and B is the empty set: { }.

Putting sets together: Unions, intersections, and complements

After you identify the sample space, S, and the various events, or subsets, of S that compose the space (see the previous sections), you can put those sets together with unions, intersections, or complements. This action is similar to adding and subtracting, only you're dealing with sets, so you use a different notation.

When you find the union of two sets, you produce a set that's at least as big as the largest of your two sets; it may be as large as the sample space itself. When you find the intersection of two sets, you produce a set that's at most as big as the smaller of the two sets; it can even be as small as the empty set. In general, unions make the sets stay the same or get larger, and intersections make the sets stay the same or get smaller.

Unions

To put two sets together into one possibly larger set is called *forming a union* of those two sets. The notation for a union is $\cup$. $A \cup B$ represents the union of two sets A and B. ***Note:*** The union of two sets is itself a set.

Say, for example, that you're rolling a single die. Event A is the event that you roll an odd number, and event B is the event that you roll a number greater than 2 (in other words, at least a 3). What would the set $A \cup B$ look like? You know that $A = \{1, 3, 5\}$, and you know that $B = \{3, 4, 5, 6\}$, which means $A \cup B$ is $\{1, 3, 4, 5, 6\}$. ***Note:*** The union is itself a set, so you need to use set notation, or brackets, around the new set. Notice that you represent each distinct outcome in either set only one time in the union set (the number 3 shows up in both sets, but in the union, it appears only one time). Now, suppose you have a third event, C, which represents an even number on the die: $C = \{2, 4, 6\}$. If you want to look at the set $A \cup C$, you get $\{1, 2, 3, 4, 5, 6\}$, which is equal to the sample space S.

Intersections

To find only the common outcomes between two sets is to find the *intersection* of those sets. The notation for intersection is $\cap$. $A \cap B$ represents the intersection of two sets A and B. ***Note:*** The intersection of two sets is itself a set.

In the previous die rolling example, you have $A = \{1, 3, 5\}$ and $B = \{3, 4, 5, 6\}$. What are the common outcomes of these two sets? The only common outcomes are the numbers 3 and 5, so $A \cap B$ is $\{3, 5\}$. If you look at the intersection of A and C ($\{2, 4, 6\}$), what would you find? The sets A and C have no common elements, so you use the notation for the empty set, $\emptyset$.

Complements

The *complement* of an event A is the set of all outcomes from the sample space, S, that don't reside in A. (Maybe the outcomes used to reside in A but didn't receive enough compliments, so they left . . .) The notation for the complement of A is A^c. For example, if $S = \{1, 2, 3, 4, 5, 6\}$ and $A = \{1, 3, 5\}$, the complement of A is the set $A^c = \{2, 4, 6\}$. In other words, if you're rolling a die, set A represents the outcomes where the die comes up odd, and set A^c represents the outcomes where the die comes up even.

Probabilities of Events Involving A and/or B

The types of probabilities you want to find will vary according to the question under review, but all probabilities boil down to the big five: marginal probability, union probability, intersection probability, conditional probability, and the probability of a complement. Suppose, for example, that you roll a single fair die. You may want to know the probability that the number you roll is even (marginal probability); the probability that the number is even or less than 4 (union probability); the probability that the number is even *and* less than 4 (intersection probability); or the probability that the number is 5 if you know it came up odd (conditional probability). I address each type of probability in the pages that follow, along with how probability notation fits into the picture.

Even though it may seem easier to figure probabilities intuitively without the use of formulas, you need to resist the urge and stick to using the definitions and formulas to figure them out. When the problems get more complicated, you'll be glad that you've established a process to make the calculations.

Probability notation

To describe the probability of an outcome (or set of outcomes), you need shorthand notation. But in order to understand the notation, you need to know what a probability really means. You start with a set, such as A = {1, 2, 3} outcomes of the roll of a die (see the section "A Set Notation Overview" earlier in this chapter). To find the probability of A occurring (of the die coming up 1, 2, or 3), you give set A a number between zero and one. (In this case, the number is $\frac{3}{6}$ or $\frac{1}{2}$, because a die boasts six possible numbers and three of them are in set A.) The $\frac{1}{2}$ is a probability. (This is the classical approach to finding a probability — see Chapter 1 — and I use this approach throughout the rest of the book except where otherwise noted.)

A probability is really a mapping (or a correspondence) that goes from the sample space S to the numbers on the number line between zero and one. Table 2-1 shows this mapping for a single roll of a die. Note that each probability is $\frac{1}{6}$ because you have six possible outcomes and each one has an equal chance of occurring, assuming the die is fair.

Table 2-1 Probability Mappings for a Single Die (S = {1, 2, 3, 4, 5, 6})

Outcome from S	*Probability*
{1}	$\frac{1}{6}$
{2}	$\frac{1}{6}$
{3}	$\frac{1}{6}$
{4}	$\frac{1}{6}$
{5}	$\frac{1}{6}$
{6}	$\frac{1}{6}$

Because the probability of the set {1} is $\frac{1}{6}$, you write P(1) = $\frac{1}{6}$, or 0.167. You say, "The probability of 1 is equal to one-sixth, or 0.167." Similarly, you write P(2) = $\frac{1}{6}$, and so on. If set A is {1, 3, 5}, you write its probability as P(A) = $\frac{3}{6}$, or 0.50, because it contains three elements, each with the probability $\frac{1}{6}$. (You say, "The probability of A is equal to $\frac{1}{2}$, or 0.50.") The important idea here is that you find the probability of a set of outcomes. In the end, the probability you get is a number between zero and one. If you have an empty set, you always give it a probability of zero.

WARNING!

Be clear about the different parts of a probability equation/statement. Consider the example P(A) = 0.50 from the previous paragraph. The letter A *represents* the set {1, 3, 5}; it doesn't *equal* 0.50. The probability of A is what equals 0.50. Also, be careful when using probability notation. P(1) = $\frac{1}{6}$ is correct, but P(1) = P($\frac{1}{6}$) is incorrect.

Marginal probabilities

If you're finding the probability of a set A all by itself, the probability you're finding is called the *marginal probability* of A. For example, suppose you roll a fair die and want the probability that the number you roll is even. The event in question here is A = {2, 4, 6}. Because three equally likely outcomes make up this set, the probability that the die will come up even is P(A) = $\frac{3}{6}$, or $\frac{1}{2}$.

REMEMBER

You aren't concerned with more than one single characteristic of an outcome when you're looking at a marginal probability. For the die example, the die coming up even is the only characteristic of concern.

Union probabilities

The probability of the union of two events, say A and B, is called a *union probability* and is written $P(A \cup B)$. The language often associated with unions is the word "or"; in other words, $P(A \cup B)$ means the probability of "A or B." For example, suppose you roll a fair die. You let A be the event that the outcome is even and B represent the event that the outcome is less than 4: $A = \{2, 4, 6\}$ and $B = \{1, 2, 3\}$. The union of sets A and B is the set of all numbers present in A or B or both: $\{1, 2, 3, 4, 6\}$. This union contains five equally likely elements, so $P(A \cup B) = 5/6$. (In the section "Understanding and Applying the Rules of Probability" later in this chapter, you find out more about how to calculate union probabilities.)

In the case of union probabilities, you're concerned about two characteristics of an outcome, and the chance that one or the other (or both) is present. Keep in mind, however, that "or" doesn't mean "either or." It means A or B or both.

Intersection (joint) probabilities

The probability of the intersection of two events, say A and B, is called an *intersection probability,* or a *joint probability,* and is written $P(A \cap B)$. The "joint" part of joint probability means "happening at the same time."

The language often associated with intersection is the word "and" — $P(A \cap B)$ means the probability of "A and B." The joint probability for A and B is the probability of all outcomes jointly located in A and B at the same time. For example, suppose that you roll a fair die. You let A be the event that you roll an even number and B represent the event that the number is less than 4: $A = \{2, 4, 6\}$ and $B = \{1, 2, 3\}$. The intersection of A and B is the set $\{2\}$. The intersection contains only one element, which has a probability of $1/6$. That means $P(A \cap B) = 1/6$. (In the section "Understanding and Applying the Rules of Probability" later in this chapter, you discover a shortcut for calculating joint probabilities under certain circumstances.)

Complement probabilities

The *complement* of an event A, A^c, is every item in the sample space, S, that isn't in A. For example, suppose you're rolling a single die, and the event $A = \{2, 4\}$. The complement of A is the event $A^c = \{1, 3, 5, 6\}$. Because this set contains the four die outcomes that don't appear in A, the probability of

A^c is $\frac{4}{6}$, or $\frac{2}{3}$, because each outcome of the die is equally likely. (In the section "Understanding and Applying the Rules of Probability," you find out how to get the probability of complements in a general way and how very helpful complements can be in terms of calculating a probability.)

Conditional probabilities

Sometimes knowing prior information about an outcome can change the probability of the outcome. And when you break down outcomes into subgroups (for example, odd or even die outcomes), the probabilities change. Conditional probabilities deal with the change that comes from factoring in prior information. The probability of one event given that another event has already occurred is a *conditional probability*.

Solving conditional probabilities without a formula

Conditional probabilities provide a way to compare groups or to use information that you already know about a situation to your advantage. The notation for the probability of event A given that B has already occurred is $P(A|B)$ — translated as "probability of A given B."

Be careful not to confuse conditional notation. $P(A|B)$ isn't $P(A)$ divided by $P(B)$. The notation separates the event you already know has occurred (B) from the event you want to find the probability for (A). And remember that the event that follows the "|" is the known or given event; the conditional probability of B given A is entirely different from the conditional probability of A given B.

For example, say that you're rolling a single die, and one roll of the die comes up odd. What's the probability that the roll is a 5? In probability notation, you want P(die is a 5 | die is odd), or $P(C|A)$, where A is the event that the die is odd, and C is the event the die is a 5. After you know that the die is odd, you have only three possibilities — 1, 3, or 5 — and each are equally likely. Therefore, you can say that P(die is 5 | die is odd) is $\frac{1}{3}$, or 0.33.

Solving conditional probabilities with a formula

The definition of $P(A|B)$ in equation form is $P(A|B) = \frac{P(A \cap B)}{P(B)}$. To find an answer, you take the joint probability of A and B and divide it by the probability of B. The numerator is the joint probability because you want the outcomes from B that also appear in A. You divide by the probability of B because B is your new sample space; you know that the item in question is already in set B.

Using conditional probabilities to evaluate disease testing

Doctors often want to know how effective a certain test is for detecting a disease. To find out, they have to test it on people who have the disease and on people who don't have the disease. They assign the set A to the event that someone has the disease and set B to the event that someone tests positive for the disease. If the test works properly, the probability of testing positive for people who have the disease should be very high, and the probability for testing negative for people who don't have the disease should also be very high. After all, you don't want to make a mistake and scare people by telling them they have the disease when they actually don't.

The notation for the event that someone who has the disease tests positive is $B|A$. This means you use the set "B given A," where A means you have the disease, and B means you test positive. You use the word "given" because you know the person belongs to event or group A already (the given part, or the part indicated after the "|" sign). So, of those people who have the disease, what's the probability that they test positive? This probability is represented by $P(B|A)$, and it should be high.

For the test to work properly, you also have to check how often it gives the negative diagnosis to people who don't have the disease. In other words, given that you don't have the disease, what's the probability that you'll test negative? The known part is that you don't have the disease; the unknown part is whether you'll test negative. Participants who don't have the disease are noted by A^c, the complement of set A. Giving the correct result for these people means that the test should come out negative, indicated by B^c, the complement of set B. Therefore, you want to look at $P(B^c|A^c)$ — the conditional probability of B complement, given A complement. You want this probability to be high.

You can't find the conditional probability of A given B if the probability of B is zero — in other words, if B is the empty set. But that's no problem; if B is the empty set, you shouldn't be interested in finding the probability of A given B anyway, because B can't happen.

You can use the formula for conditional probability to find the answer to the example problem in the previous section, where $A = \{1, 3, 5\}$ and $C = \{5\}$. According to the definition of conditional probability, $P(C|A) = \frac{P(C \cap A)}{P(A)}$. You know that $P(C \cap A)$ equals $\frac{1}{6}$ because the intersection of those two sets is the outcome $\{5\}$, and its probability is $\frac{1}{6}$. Now you can say $P(A) = P\{1, 3, 5\} = \frac{3}{6}$. Dividing these two, you get $\frac{1}{6} \div \frac{3}{6}$, which is $\frac{1}{6} * \frac{6}{3} = \frac{1}{3}$, or 0.33. You get the same answer as when you don't use the formula.

Understanding and Applying the Rules of Probability

You can find the probability of outcomes, events, or combinations of outcomes and/or events by adding or subtracting, multiplying, or dividing the probabilities of the original outcomes and events. You use some combinations so often that they have their own rules and formulas. The better you understand the ideas behind the formulas, the more likely it is that you'll remember them and be able to use them successfully.

Any probability has to follow three basic properties:

- Every probability has to be a number between zero and one. If you ever report that the probability of an event is greater than one or negative, you've made a mistake!
- To find the probability of a set of individual outcomes from S, the sample space, you sum their probabilities. (This isn't necessarily true for combining events, but it is for individual outcomes.)
- If you take the probabilities of all the outcomes in S, they have to sum to one.

The following pages contain the basic rules and formulas of probability that build on these three basic properties.

The complement rule (for opposites, not for flattering a date)

The *complement* of an event A, A^c, is the set of every item or individual in the sample space, S, that isn't in A. The probability of the complement of event A is the chance that A didn't occur. By definition, if you take the union of A and A complement, you get S, and the probability is $1/1$; therefore, $P(A^c) + P(A) = 1$. Solving for $P(A^c)$, you get what's called the *complement rule:* $P(A^c) = 1 - P(A)$.

Suppose that you're rolling a single die. The sample space S = {1, 2, 3, 4, 5, 6}. If you let A = {1, 3, 5}, the complement of A is the set A^c = {2, 4, 6}. Say, for example, that you want the probability of rolling a number greater than 1 (in other words, at least 2), which means you create an event D: {2, 3, 4, 5, 6}. You can see that the probability of D, P(D), is $5/6$, but you can also find this probability by using the complement rule. You know that D^c = {1} and $P(D^c) = 1/6$; therefore, according to the complement rule, $P(D) = 1 - P(D^c) = 1 - 1/6 = 5/6$.

Oftentimes, the event you have to deal with is complicated and difficult to get a handle on. In tough cases, before you start pulling your hair out, think of the complement; it may be easier to grasp the outcomes you don't want than the outcomes you do want.

To illustrate this idea, look at the situation where you roll two dice. You have $6 * 6 = 36$ possible outcomes, from (1, 1) all the way to (6, 6). The following table shows the entire set of outcomes.

(1, 1)	(2, 1)	(3, 1)	(4, 1)	(5, 1)	(6, 1)
(1, 2)	(2, 2)	(3, 2)	(4, 2)	(5, 2)	(6, 2)
(1, 3)	(2, 3)	(3, 3)	(4, 3)	(5, 3)	(6, 3)
(1, 4)	(2, 4)	(3, 4)	(4, 4)	(5, 4)	(6, 4)
(1, 5)	(2, 5)	(3, 5)	(4, 5)	(5, 5)	(6, 5)
(1, 6)	(2, 6)	(3, 6)	(4, 6)	(5, 6)	(6, 6)

Suppose A is the event that at least one of the dice comes up greater than 1 on a roll. To find this probability, you need to add the probabilities of all the outcomes that make up this event. That's a lot of possibilities! All the outcomes in the first column of the previous table are included, except the first one — (1, 1). All the outcomes in the remaining five columns are also included. In fact, all the outcomes except (1, 1) fit this description. So, you can calculate P(A) by finding all the outcomes that meet the description and summing up their probabilities to get $35/36$, but you may find it easier to take a look at the complement. The complement of A is the set of all outcomes in S in which you don't have at least one of the dice greater than 1. The only outcome that meets this criterion is (1, 1); therefore, you have $A^c = (1, 1)$. You know that $P(A^c) = 1/36$, and by the complement rule, $P(A) = 1 - P(A^c) = 1 - 1/36 = 35/36$. The complement rule makes it much easier to find P(A) in this case.

The multiplication rule (for intersections, not for rabbits)

You use the *multiplication rule* to find the probability of an intersection of two events A and B. It makes sense that the definition of conditional probability involves an intersection (see the section "Conditional probabilities" earlier in this chapter) because conditional probability is where the multiplication rule comes from. The conditional probability of A given B is defined as $P(A|B) = \frac{P(A \cap B)}{P(B)}$. If you cross-multiply this formula, you get $P(A \cap B) = P(B) * P(A|B)$. Allow me to translate: The probability that A and B occurred is equal to the probability that B occurred times the probability that A occurred given that B occurred. The multiplication rule splits up the joint probability into two stages; first B occurs, and then A occurs given that B has occurred.

If you're given a marginal probability and a conditional probability, you can use the multiplication rule to find the joint probability.

Suppose, for example, that a class is made up of 60 percent women, and of these women, 40 percent are married. What's the chance that a person you select at random from the class is a woman and married? To answer this, let the event $W = \{woman\}$ and $M = \{married\}$. What you want is P(W and M), which is the joint probability $P(W \cap M)$. You know that 60 percent of the class is made up of women, which means $P(W) = 0.60$. You also know that of the women in the class, 40 percent, 0.40, are married. You have to use a conditional probability to solve the problem because you split up the women and look at the probability that they're married — $P(M|W) = 0.40$. By the multiplication rule, to find $P(W \cap M)$, you take $P(W) * P(M|W) = 0.60 * 0.40 = 0.24$. Of all the people in the class, 24 percent are women and are married, which also means that the chance of you picking a married woman from the class is 24 percent.

A probability, technically, is a number between zero and one, but you often see it expressed as a percentage because it's easier to interpret that way. You write a probability as a percentage by multiplying the probability by 100.

Be aware of the difference between a joint probability and a conditional probability. You need a *joint probability* when you select someone from the entire group who has two characteristics. You need a *conditional probability* when you pull out a subgroup that has one of the characteristics already, and you want the probability that someone from that subgroup has a second characteristic.

The addition rule (for unions of the nonmarital nature)

The *union* of two events A and B is the set of all outcomes in the sample space, S, that are in either A or B or both. To find the probability of the union of two events A and B, you do what appears to be the most intuitive calculation — you add the two probabilities together. Only you can't stop there. When you add P(A) and P(B), you double count the outcomes that are in both A and B; in other words, you double count the outcomes in $A \cap B$. If you count those outcomes twice, it makes the probability of the union too large. So how do you fix it? Because you count the outcomes in $A \cap B$ twice, you should subtract them out once. The probability of A union B is given by the following formula: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. This formula is called the *addition rule.*

For example, suppose that you have a class made up of 60 percent women, 40 percent of which are married (for the introduction of this example, see the previous section). Suppose you also know that 50 percent of all the people in the class are married. You want to find the percentage of people in the class who are women *or* married (or both): $P(M \cup W)$. Using the addition rule, you have $P(M \cup W) = P(M) + P(W) - P(M \cap W)$. You know that 60 percent of the class is made up of women, so P(W) = 0.60. You know that 50 percent of the class is married, so P(M) = 0.50. By using the multiplication rule in the previous section, you find that $P(M \cap W) = 0.24$. Therefore, $P(M \cup W) = 0.50 + 0.60 - 0.24 = 0.86$. You calculate that 86 percent of the class is married or female (or both).

Notice that if you don't subtract the intersection in the previous example, you get 0.50 + 0.60 = 1.1, which is greater than one. You can never have a probability greater than one or less than zero.

Recognizing Independence in Multiple Events

One of the most important assumptions of the basic probability models is independence. Multiple events are *independent* if knowledge that one event has happened doesn't affect the probability of the other event happening. In other words, knowing that A has occurred doesn't change the probability of B occurring given A. If events A and B are independent, you can say bye-bye to conditional probabilities, which makes your life much easier! (See the section "Conditional probabilities" earlier in this chapter for more information on the topic.)

You have two ways to check for independence (assuming A and B aren't empty):

- **Use the definition of independence.** Check to see if $P(A|B) = P(A)$ or if $P(B|A) = P(B)$.
- **Check the multiplication rule for independence.** You can also check to see if $P(A \cap B) = P(A) * P(B)$. If so, A and B are independent.

Checking independence for two events with the definition

Suppose that you're rolling a single die. The sample space, S, = {1, 2, 3, 4, 5, 6}. If you let event A = {the die comes up odd} and event B = {the die comes up 1}, are these two events independent? To answer that, first ask the question, "If I know the die is odd, what's the probability that it's a 1?" The answer is $\frac{1}{3}$. Now ask, "What's the probability that the die is a 1 without knowing whether it's odd?" The answer is $\frac{1}{6}$. The probabilities are different, so events A and B are not independent. Knowledge of one event affects the probability of the other event.

Now suppose you add the event C = {the die is a 1 or 2}. Are events A and C independent? To find out, you check to see if P(C) equals P(C | A) (you could also check to see if P[A] = P[A | C]). The probability of C is $\frac{2}{6}$, or 0.33. The probability of C given A is the probability of the die being a 1 or 2 given that it's odd. Using the definition of conditional probability (see the section "Conditional probabilities"), you have $P(C \mid A) = \frac{P(C \cap A)}{P(A)}$. The set $C \cap A$ is the set {1}, whose probability is $\frac{1}{6}$. The probability of A is $\frac{3}{6}$. Dividing these probabilities, you get $\frac{1}{6} \div \frac{3}{6}$, which is $\frac{1}{3}$, or 0.33. Events A and C are independent because knowing that the die is odd doesn't change the probability that the die is a 1 or a 2. Some information isn't really worth knowing, because it doesn't affect the chances.

One big word of caution when it comes to independent events: If two events are independent, it doesn't mean that they can't happen at the same time. Many people make the mistake of thinking of independent events as being totally separate from each other. In probability, two independent events can happen at the same time, and in essence coexist; they just don't affect each other in terms of their probabilities.

Utilizing the multiplication rule for independent events

I can't overemphasize how wonderful life is when events are independent. Suppose you want the probability of five events happening at the same time. If these events aren't independent, you have to find the conditional probabilities at every stage of the process: The second event would depend on the first event; the third event would depend on the first and second events; the fourth event would depend on the first three events; and the fifth event would depend on the first four events. What a complicated mess! If all the events are independent, their five-way joint probability is just the product of the five probabilities of the individual events. Much easier! You can extend the multiplication rule for independent events to any number of events. To find the joint probability of two events, for example, you simply multiply their individual probabilities.

If you know that two events are independent, or if you can show that they're independent, your calculations are much easier for any probabilities affiliated with these events. The reason is because when you know A and B are independent, you can say that P(A | B) = P(A) and P(B | A) = P(B). "Well, what's so great about that," you ask? When you go to find the joint probability of A and B, $P(A \cap B)$, it equals P(A) * P(B | A). But because A and B are independent, it equals P(A) * P(B), because P(B | A) = P(B). Therefore, to get the joint probability of events A and B, you multiply the marginal (or individual) probabilities of A and B together, if A and B are independent.

It may be tempting to take $P(A) * P(B)$ whenever you need to find $P(A \cap B)$, but you can do this only if A and B are independent. If not, you have to use the formula $P(A) * P(B|A)$ and deal with conditional probabilities.

Suppose, for example, that you're rolling dice. You can safely assume that the outcome on one die doesn't affect the outcome on the other die. If you roll two dice, what's the probability of getting a 1 and a 1? According to the multiplication rule, it's the probability of getting a 1, $\frac{1}{6}$, times the probability of getting a 1, $\frac{1}{6}$. The result is $\frac{1}{36}$. If you roll five dice, the chance of getting all 1s is $\frac{1}{6} * \frac{1}{6} * \frac{1}{6} * \frac{1}{6} * \frac{1}{6}$, which equals $(\frac{1}{6})^5$. In general, if you roll n dice, the probability of getting all 1s is $(\frac{1}{6})^n$. Now, if you want to find the probability of tossing a single 1 when you roll five dice five times, that's more difficult, because you can use different ways to get a single 1, and you have to take them all into account. (See Chapter 7 for info on handling situations that involve n dice.)

Including Mutually Exclusive Events

In probability, you often see independent events that can occur at the same time without affecting each other when they do, but you also see the opposite situation, where two events can't occur at the same time, and hence affect each other greatly. Two events A and B are *mutually exclusive* if they can't occur at the same time. In other words, they exclude each other from occurring: $A \cap B = \emptyset$, or $P(A \cap B) = 0$. If you know A has occurred, you know that B can't occur; and if you know B has occurred, you know that A can't occur.

Like with independent events (see the previous section of this chapter), mutually exclusive events can make your calculations much easier, so you should keep an eye out and try to control for them in probability models where possible.

Recognizing mutually exclusive events

Tagging two events as mutually exclusive doesn't always mean that one event or the other must occur; it only means that if one of the events occurs, the other event can't occur.

Look at the outcomes of a traffic light, where the sample space, S, = {red, green, yellow}. Let A equal the event that the light is green, and let B equal the event that the light is red. If you know the light is red, you know it can't be green, and vice versa. Those two events are mutually exclusive because mutually exclusive events directly affect each other's probabilities. No matter what P(A) is, you know that $P(A|B)$ has to be zero, and vice versa.

If A and B are mutually exclusive, note that $P(A \cap B) = 0$, so
$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)} = 0$.

You can also use the definition of mutually exclusive events in reverse order to see if A and B are mutually exclusive, because definitions always go in both directions. You know that if A and B are mutually exclusive, $P(A \cap B) = 0$. Therefore, to check to see if two events are mutually exclusive, you can check to see if $P(A \cap B) = 0$. If so, the events are mutually exclusive; if not, the events aren't mutually exclusive.

One special case of mutually exclusive events is events that are *complements.* Such events are the opposite of each other in terms of the outcomes they contain.

For example, say you flip a coin twice. The sample space is {HH, HT, TH, and TT} — H denotes getting a head, and T denotes getting a tail. If you let A be the event that the outcome is two heads, the complement, A^c, is the event that you don't get two heads, so A = {HH} and A^c = {HT, TH, TT}. By definition, because A^c contains the outcomes in S that don't appear in A, the events A and A^c can't intersect — $A \cap A^c = \emptyset$, which means $P(A \cap A^c) = 0$. That makes the events mutually exclusive.

Simplifying the addition rule with mutually exclusive events

If two events are mutually exclusive, they have no intersection, which makes using the addition rule for probability much easier. The addition rule finds the probability of the union of two events A and B (refer to the section "The addition rule [for unions of the nonmarital nature]" earlier in this chapter); it's the probability of the set of all outcomes in A or B or both. With mutually exclusive events, no outcomes lie in both sets, so the addition rule is simplified to the sum of the two events. Using probability notation, the addition rule looks like this: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. If A and B are mutually exclusive, $P(A \cap B) = 0$, so the addition rule becomes $P(A \cup B) = P(A) + P(B)$. You don't have to deal with the intersection probabilities, which makes the calculations much easier.

Suppose, for example, that you select a card from a standard 52-card deck, and you want the probability that the card is a 2 or a 3. Let A = {card is a 2} and B = {card is a 3}. Note that $P(A) = 4/52$, because you have 52 cards in a standard deck, and 4 of the cards are 2s. Similarly, $P(B) = 4/52$. Because you want the probability that the card is a 2 *or* a 3, you want a union probability, or $P(A \cup B)$. Events A and B are mutually exclusive because a card can't be a 2 and a 3 at the same time, so $P(A \cap B) = 0$. Therefore, $P(A \cup B) = P(A) + P(B) = 4/52 + 4/52 = 8/52 = 0.154$, or a 15.4 percent probability.

Reviewing the standard 52-card deck

A standard card deck contains 52 cards. Each card has one of 13 denominations on it: ace, 2, 3, 4, 5, 6, 7, 8, 9,10, jack, queen, or king. Sometimes the ace is considered the highest denomination, and sometimes it's considered the lowest denomination; it all depends on what type of game you play with the cards. Each card is also labeled with one of four possible *suits.* The four suits are diamonds, denoted by ♦; hearts, denoted by ♥; clubs, denoted by ♣; and spades, denoted by ♠. Diamonds and hearts are red cards (their denomination and suit labels are marked in red), and clubs and spades are black (their denomination and suit labels are marked in black). Thirteen cards make up each suit, which makes 26 of the cards red and 26 black. The deck has four cards in each denomination. The cards denoted by J, Q, or K are called *face cards* or *court cards,* because on those cards you see the face (and body) of a jack (or prince), a queen, or a king. One deck contains 12 face cards.

REMEMBER

You may be tempted to take $P(A) + P(B)$ whenever you need to find $P(A \cup B)$, but you can do this only if A and B are mutually exclusive. If not, you have to use the formula $P(A) + P(B) - P(A \cap B)$ and deal with intersection probabilities.

Distinguishing Independent from Mutually Exclusive Events

One challenge students of probability face is understanding the difference between independent events and mutually exclusive events. Both events are often easy to understand separately, but when compared to each other, the concepts seem to fall apart. However, if you take the time to study the definitions of each event, you'll see a stunning difference between the two; it all boils down to the intersection probabilities and how they compare.

Comparing and contrasting independence and exclusivity

If events A and B are independent, they can occur at the same time, which means they can intersect. Their intersection (or joint) probability is given by $P(A \cap B) = P(A) * P(B)$. So, to find the intersection probability of two events, you multiply their marginal probabilities together. If, however, events A and B are mutually exclusive, they can't occur at the same time, which means they can't intersect. Their intersection (or joint) probability is given by $P(A \cap B) = 0$.

So now comes the big question. Suppose A and B are nonempty events (which means their probabilities aren't zero) and independent. Can they be mutually exclusive? No, because for them to be mutually exclusive, they must have no intersection, so $P(A \cap B)$ must be zero. Because $P(A \cap B) = P(A) * P(B)$ and A and B are independent, the only way to get zero is if P(A) is zero or P(B) is zero, and that isn't the case. So, if two events are nonempty and independent, they can't be mutually exclusive.

Turning this example around, if A and B are mutually exclusive events and nonempty, can they be independent? No. To see this clearly, look at the definition of independence. Suppose events A and B are mutually exclusive. These two events are independent if $P(A|B) = P(A)$, but you know that $P(A|B) = \frac{P(A \cap B)}{P(B)}$, and the numerator is $P(A \cap B) = 0$ because A and B are mutually exclusive. This fact forces the entire conditional probability, $P(A|B)$, to be zero. But for A and B to be independent, $P(A|B)$ has to be equal to P(A), and it can't be zero unless A is the empty set. So, if events A and B are nonempty and mutually exclusive, they can't be independent.

REMEMBER

Mutually exclusive events can't be independent and independent events can't be mutually exclusive unless one (or both) of the events is the empty set.

Checking for independence or exclusivity in a 52-card deck

Suppose you pull a card from a standard 52-card deck. Let A = {the card is a 2}; let B = {the card is black}; let C = {the card is a face card); and let D = {the card isn't a face card}. You find that $P(A) = \frac{4}{52}$, or $\frac{1}{13}$; $P(B) = \frac{26}{52}$, or $\frac{1}{2}$; $P(C) = \frac{12}{52}$, or $\frac{3}{13}$; and $P(D) = 1 - \frac{3}{13} = \frac{10}{13}$ (by the complement rule, because C and D are complement events; see the section "The complement rule [for opposites, not for flattering a date]"). Are events A and B mutually exclusive? No, because they have an intersection: Two of the cards in the deck are 2s and black (♠2 and ♣2).

Are events A and B independent? You find that $P(A \cap B) = \frac{2}{52}$, or $\frac{1}{26}$. You can also find that $P(A) * P(B) = \frac{1}{13} * \frac{1}{2} = \frac{1}{26}$. Because these probabilities are equal, the events are independent. This makes sense, because if you know that the card is black, the probability of it being a 2 is $\frac{2}{26}$, which is the same as the probability of the card being a 2 without knowing it's black (which is $\frac{4}{52} = \frac{2}{26}$). So, knowing that the card is black doesn't affect the probability of it being a 2.

What about events A and C? Are they independent? Because a card can't be a 2 and a face card at the same time, the intersection is the empty set, so A and C are mutually exclusive — their probabilities directly affect each other. $P(A) = \frac{4}{52}$, and $P(A|C)$ equals 0; because these numbers aren't equal, the events aren't independent.

Pinpointing probability models: It's all in the way you say it

The word "or" is a clue that you need to find the probability of a union — for example, when you have to find the probability that someone owns more than one cell phone *or* more than one land-line phone. The word "and" is a clue that you need to find the probability of an intersection — for example, when you need to find the probability that someone owns at least one cell phone *and* at least one land-line phone. The word "of" is a good indicator that you're looking at a conditional probability — for example, *of* those people who own land-line phones, what's the probability that they also own cell phones?

What about events A and D? Their intersection is the set of all cards that are 2s and not face cards, of which there are four (all the 2s), so $P(A \cap B) = \frac{4}{52}$. Therefore, A and D aren't mutually exclusive. Are the events independent? Because $P(A) = \frac{4}{52}$, or $\frac{1}{13}$, and $P(D) = \frac{40}{52}$, or $\frac{10}{13}$, $P(A) * P(D) = \frac{1}{13} * \frac{10}{13} = 10/169 = 0.059$. But remember, the intersection of A and D contains four cards (all the 2s), so its probability is $\frac{4}{52} = 0.077$. These two probabilities aren't equal, so events A and D aren't independent.

REMEMBER

For two events to be independent, P(A) has to be equal to $P(A|B)$ — not close, but equal. In the previous example, the probabilities 0.059 and 0.077 may seem close, but close doesn't count when it comes to independence. The numbers have to be exact.

Chapter 3

Picturing Probability: Venn Diagrams, Tree Diagrams, and Bayes' Theorem

In This Chapter

- Organizing probability information with Venn diagrams and tree diagrams
- Using diagrams to solve complex probability problems
- Figuring marginal probabilities with the Law of Total Probability
- Conquering multi-stage probabilities with Bayes' Theorem

Probability problems can get complicated quickly — especially if you have more than one event to deal with at the same time or if the information appears to come in stages. The most complicated probability problems you come across are complicated only because of the information you're given, compared to what you need. Two very common situations come to mind. First, you're given the conditional probability of A given B and its related complements (see Chapter 2), as well as marginal probabilities for B and B^c, and you need to find the marginal probability of A. In another situation, you may be given the conditional probability for B given A and its related complements, and you need to find the conditional probability of A given B (in other words, the conditional probability in the opposite order). Yes, you can solve both of these problems, but they each require two major tools: a good picture and a good formula.

In this chapter, you discover different methods of putting probabilities into picture form and using those pictures to help you develop techniques for solving more complex probability problems. You also get formulas that you need to solve those complex problems. Although some methods are better to use than others in certain situations (I explain which ones are better when), the most important idea is to make the pictures and methods work for you so you can gain control over complicated problems.

Diagramming Probabilities with Venn Diagrams

One way to organize information given in a probability problem is to draw a picture that represents the sample space (the set of all possible outcomes; see Chapter 2), all the events involved (represented by *sets,* or subsets of the sample space; see Chapter 2), and all the subsets that are formed when events are allowed to intersect. In other words, you want a picture that takes a complex situation and breaks it into bite-sized pieces that you can easily identify and work with to get the solution. One of the most common pictures used to represent probabilities is a Venn diagram.

A *Venn diagram* is a picture that uses a box to represent the sample space, S, and circles to represent the various events involved in the problem. If the events can intersect, the circles overlap. If the events are mutually exclusive (which means no intersection takes place; see Chapter 2), the circles appear separated. Figure 3-1 shows two examples of Venn diagrams with two events, A and B. Figure 3-1a shows the possibility of intersecting events, and Figure 3-1b shows mutually exclusive events.

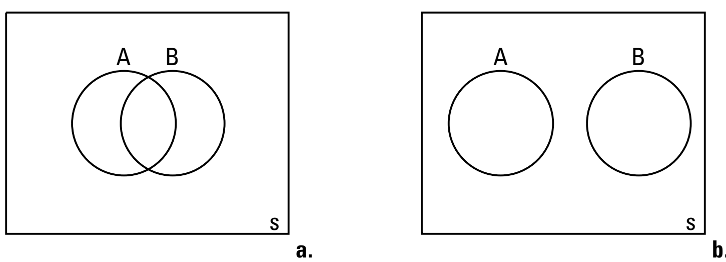


Figure 3-1: Two examples of Venn diagrams.

TIP: The S in the lower-right corner of each box indicates that the entire set is the sample space S. You can omit this notation unless you reduce the sample space; for example, if you look only at the group of females in a class, you can write F in that space to indicate your new sample space.

Utilizing Venn diagrams to find probabilities beyond those given

The most important use of a Venn diagram is to help you find probabilities beyond those given in the problem. You know the drill; you're given certain pieces of information in a problem, and you have to answer tons of questions with the pieces. Sometimes it seems like you have to take straw and turn it into gold, right? With a Venn diagram, you can organize the information you have, and with the rules of sets and probability, the diagram helps you organize, identify, and figure out other probabilities that you need to find.

TIP

Before you start to work out your solution to the problem, set up and fill out your Venn diagram completely first. That's the key to success.

Using Venn diagrams to organize and visualize relationships

Venn diagrams help you organize and account for all the possible sets and subsets that occur in a probability scenario. Each piece of the diagram has meaning and a probability. After you account for all the pieces in terms of their probabilities, you can solve many different types of problems. In Figure 3-2, you see a Venn diagram representing set A and its complement, A^c (the shaded area represents A^c; the complement of a set A is everything in the sample space that isn't included in set A [see Chapter 2]).

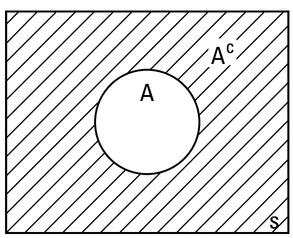


Figure 3-2: Sets A and A^c, represented by a Venn diagram.

Venn diagrams also help you to visualize important relationships that can exist between two events. In Figure 3-3, I show two sets, A and B, which represent two events, A and B, that can intersect, and I identify and label each part of each piece of the diagram in terms of set notation. The set $A \cap B$ represents the set of all outcomes in S that appear in both A and B. The set $A \cap B^c$ represents all outcomes in S that appear in A but not in B. The set $B \cap A^c$ represents all outcomes in S that appear in B but not in A. $(A \cup B)^c$ is the part of the box that appears outside of both circles.

All three of these sets together represent the set A union B, or $A \cup B$. Everything outside of these three sets must be the complement of A union B, which is represented by the set $(A \cup B)^c$. So, in the end, if you take the union of all four of these subsets, you get S, the entire sample space. A Venn diagram does a great job of helping you to visualize these intricate sets and the relationships between them.

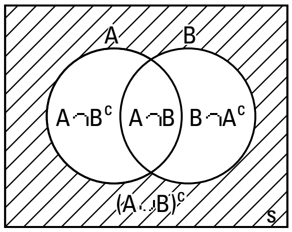


Figure 3-3: Dissecting sets by using a Venn diagram.

REMEMBER

If you take the union of all the sets you identify in the Venn diagram, the sum of all their probabilities must be one. For example, suppose you roll two dice — one red die and one green die. Let event A represent getting an odd number when you roll a red die, and let B represent getting an even number when you roll a green die. You identify four possible situations: odd on red and even on green; even on red and odd on green; odd on both red and green; or even on both red and green. Because these occurrences represent all possible cases, the sum of their probabilities must equal one. Using Figure 3-3, you see how a Venn diagram can help you visualize and represent all these situations:

- The set $A \cap B$ represents getting an odd outcome on the red die and an even outcome on the green die (for example, 1 on red and 2 on green).
- The set $A^c \cap B^c = (A \cup B)^c$ represents getting an even outcome on the red die and an odd outcome on the green die (for example, 2 on red and 1 on green).
- The set $A \cap B^c$ represents getting an odd outcome on both dice (for example, 1 on red and 3 on green).
- The set $A^c \cup B$ represents getting an even outcome on both die (for example, 2 on red and 4 on green).

TIP

Remember the commutative property in algebra? It says that $a + b = b + a$. In other words, if you add two numbers together, you can do it in either order and still get the same answer. The same is true of sets. If you union two sets A and B, you can do it in either order and get the same result. So, $A \cup B = B \cup A$. The same is true for intersections: $A \cap B = B \cap A$.

Proving intermediate rules about sets, using Venn diagrams

When working with complex sets (and their probabilities) that require the use of Venn diagrams, some intermediate rules of sets are very helpful for proving certain equalities. Two of these rules are given by the following:

- $(A \cup B)^c = A^c \cap B^c$
- $(A \cap B)^c = A^c \cup B^c$

The union inside the parentheses switches to an intersection in the first rule, and the intersection inside the parentheses switches to a union in the second rule. These two rules are sometimes known as *DeMorgan's Laws.*

TIP

You use these two intermediate rules of probability in a way similar to how you use the distributive property in algebra — $a(b + c) = ab + ac$. This is a way to break down parentheses and simplify expressions. The two rules of probability in the previous list apply that same idea to sets. Whenever you have parentheses in an expression involving sets, you can use the intermediate rules to break down the parentheses and simplify the expressions.

You can prove these rules of probability by using Venn diagrams, and in many cases, drawing pictures with Venn diagrams can prove certain results about sets and probabilities. Figure 3-4 shows the proof of the first rule from the previous list.

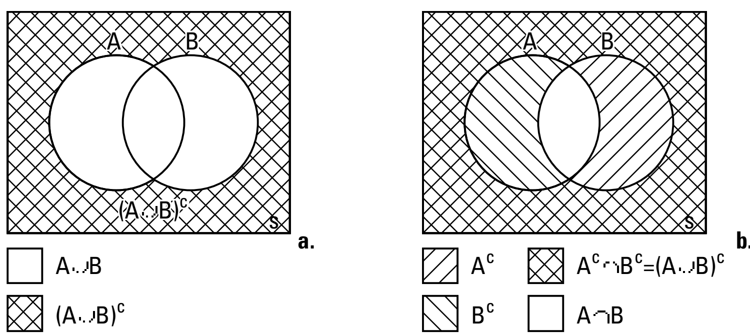


Figure 3-4: Proof that $(A \cup B)^c = A^c \cap B^c$, using Venn diagrams.

Figure 3-4a shows a picture representing the left side of the equation, and Figure 3-4b shows a picture representing the right side of the equation. Notice that the areas represented by the crosshatch pattern — intersecting left and right diagonal lines — are the same in both pictures and represent the sets on either side of the equation, which proves that the two sets are equal.

Figure 3-5 shows the proof of the second probability rule $(A \cap B)^c = A^c \cup B^c$.

Figure 3-5a shows a picture representing the left side of the equation, and Figure 3-5b shows a picture representing the right side of the equation. Because Figure 3-5b represents a union, any area that's covered with right diagonal lines, left diagonal lines, or both constitutes the set you're looking for. The area represented by diagonal lines in Figure 3-5 matches the area in Figure 3-5a, which proves that $(A \cap B)^c = A^c \cup B^c$.

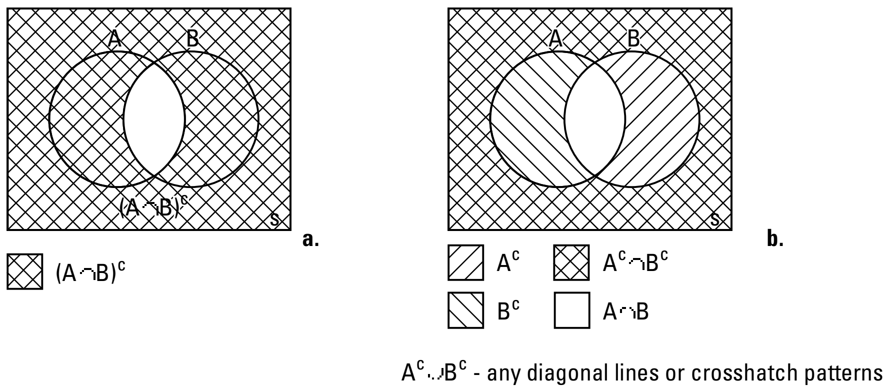


Figure 3-5: Proof that $(A \cap B)^c = A^c \cup B^c$, using Venn diagrams.

Exploring the limitations of Venn diagrams

Venn diagrams help you most if a problem gives you probabilities of events by themselves (in other words, marginal probabilities — see Chapter 2) and probabilities of intersections (also called joint probabilities — see Chapter 2). In such a situation, you can find probabilities for all other parts of the Venn diagram, such as the probability of neither A nor B happening or the probability of exactly A and/or exactly B happening. But Venn diagrams can't help you solve every type of probability problem.

Venn diagrams don't work as well if a problem gives you partial information, such as the conditional probability of A given B (written as $P[A|B]$ — see Chapter 2), or when the problem creates the sample space through a series of stages or a sequence of events that happen in a certain order. These situations require a different method for picturing the sample space (called tree diagrams; see the section "Mapping Out Probabilities with Tree Diagrams").

REMEMBER

Before starting a probability problem, think about the information you're given and determine the best way to organize the information in a picture format. You generally use Venn diagrams when you're given marginal and joint probabilities for A and B and you're asked for probabilities of combinations and/or complements of those events.

Finding probabilities for complex problems with Venn diagrams

Being able to draw out and visualize the probabilities you need to find can help you break down complex problems into pieces that you can easily put together for a solution. The following is an example of a problem that you can solve by using a Venn diagram.

Suppose your street has two traffic lights. The chance that the first light is red is 0.40, and the chance that the second light is red is 0.30. City officials set them so that the chance of them being red at the same time is only 0.10. Here are the two questions you must answer:

> **Question 1:** What's the probability that neither light is red?
>
> **Question 2:** What's the probability that exactly one of the lights is red?

In the following pages, I take you through the process of answering these questions.

Setting up your diagram

To answer the questions posed in the introduction to this section, you first define the sample space, S, and the events involved in the problem. In this case, S is composed of all the possible situations that can happen with the two traffic lights: Both are red, both are not red, or one is red and the other isn't. Notice that you don't worry about the other colors of the lights (green and yellow) because the issue of interest is whether the lights are red. So, S = {both lights are red, first light is red and second light isn't, first light isn't red and second light is, both lights are not red}. The Venn diagram for this situation is shown in Figure 3-6, where A = {first light is red} and B = {second light is red}.

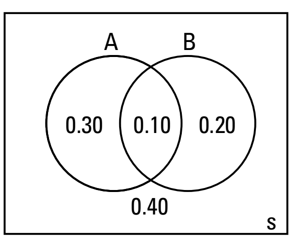


Figure 3-6: A Venn diagram playing red light, green light for a traffic problem.

I give a probability for each separate part of the Venn diagram. You can come up with these probabilities by identifying what each part of the Venn diagram represents and by using the rules of probability from Chapter 2.

First, let event A = {first light is red} and B = {second light is red}. You know that $P(A) = 0.40$ and $P(B) = 0.30$, but don't write those probabilities into the Venn diagram just yet; you have to split A into its two subsets: $A \cap B$ and $A \cap B^c$ (see the section "Using Venn diagrams to organize and visualize relationships"). Because the union of $A \cap B$ and $A \cap B^c$ equals A (by the addition rule; see Chapter 2), $P(A \cap B) + P(A \cap B^c) = P(A) = 0.40$. The probability of A and B is 0.10, so $P(A \cap B) = 0.10$. Solving for $P(A \cap B^c)$, you get $0.40 - 0.10 = 0.30$. You can now put this probability into the Venn diagram; it represents the part of set A that doesn't include B.

You use the same reasoning to find $P(B \cap A^c)$ — the probability of the part of set B that doesn't appear in set A. You know that set B is equal to the union of two subsets: $B \cap A$ and $B \cap A^c$. Because these sets are mutually exclusive (they don't intersect; see Chapter 2), their probabilities sum to $P(B) = 0.30$. And because $P(B \cap A) = 0.10$, $P(B \cap A^c) = 0.30 - 0.10 = 0.20$.

WARNING!

It may appear at this point that you're done, but you haven't filled out the entire Venn diagram. You know you're not done when the probabilities you have so far don't sum to one. Finishing too soon is the source of many mistakes in terms of calculating probabilities.

One probability remains: the probability of being outside of the circles (see the section "Using Venn diagrams to organize and visualize relationships"), because an outcome can be in S and not be in either A or B.

All the other parts of the Venn diagram have probabilities assigned to them, and because the probability of the entire sample space, S, is one, the remaining probability must be equal to one minus the sum of all the other probabilities: $1 - (0.30 + 0.10 + 0.20) = 0.40$. You represent this set in set notation as $(A \cup B)^c$, or everything in S that isn't included in A or B or both.

Answering the "neither" question

When you have the Venn diagram for this example problem completed (see the previous section), you can look at the questions you need to answer. The first question asks for the probability that neither light is red; that is, you want the probability that the first light isn't red and the second light isn't red. In probability notation, you want $P(A^c \cap B^c)$. Looking at Figure 3-6, you can see that this is the area outside the two circles representing A and B, so you don't need to use the notation to answer this question; you can see the answer on the Venn diagram. However, if you want to solve the problem by using probability notation, note that by the first intermediate rule of probability used in the opposite order, $P(A^c \cap B^c)$, equals $P(A \cup B)^c$, which equals $1 - P(A \cup B)$ by the complement rule (see Chapter 2). Using what you find in the previous section, that probability equals $1 - (0.30 + 0.10 + 0.20) = 0.40$, or 40 percent.

Answering the "exactly one" question

To answer the second question, you need to find the probability that exactly one of the lights is red. Before looking at the Venn diagram or plunging into the probability rules and formulas, it helps to think about what the question is asking. What does it mean for exactly one of the lights to be red? It means one of the lights is red and the other isn't. There are two ways this can happen, and you need to find the probability of both in order to get the right answer. Either the first light is red and the second isn't, or the second light is red and the first isn't.

On the Venn diagram, the crescent shaped area that includes set A but not set B represents the event "first light is red and the second isn't." (This area is marked with a "1" in Figure 3-7.) In probability notation, this area is noted by $P(A \cap B^c)$. You can see on the Venn diagram that this probability is 0.30. The crescent shaped area that includes set B but not set A represents the event "second light is red and the first isn't." (This area is marked with a "2" in Figure 3-7.) The probability notation for this set is $P(A^c \cap B)$. You can see on the Venn diagram that this probability is 0.20. To find the probability that exactly one light is red, you first notice that these two sets are mutually exclusive (they have no intersection), so you add their probabilities together by the addition rule (see Chapter 2): $0.30 + 0.20 = 0.50$, or 50 percent.

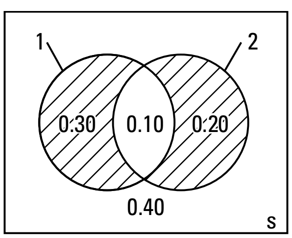


Figure 3-7: The Venn diagram when exactly one light is red.

Mapping Out Probabilities with Tree Diagrams

Some probability problems involve a multi-stage process or a sequence of events. In these cases, you need a method of visualizing the sample space that shows each of the stages in the process; that shows all the outcomes that can occur at each stage; and that allows you to see all the combinations that form as a result. The method of drawing a picture that represents a sample space that involves multiple stages or a sequence of events is the tree diagram method.

The *tree diagram* uses sets of branches to show each stage, and each possible outcome within a given stage is represented by another branch on the tree. In the end, you can follow the paths that the branches make to find all the elements of the sample space, S; each one has its own path of braches on the tree diagram.

For example, suppose you flip a coin twice. This is a two-stage process; first you flip the coin and get either heads or tails, and then you flip the coin again and get either heads or tails. The sample space shows the combinations of all possible outcomes of both stages, so S = {HH, HT, TH, TT}, where H signifies a head and T signifies a tail. The tree diagram for this sample space is shown in Figure 3-8.

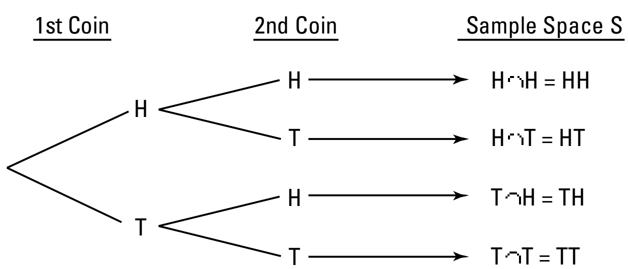


Figure 3-8: Tree diagram created by flipping a coin twice.

Two branches, H and T, represent the first coin flip. From each of those branches comes the possible outcomes of the second flip, H and T, which gives you a total of $2 * 2 = 4$ total pathways on the tree, representing each of the four possible outcomes of the sample space. For example, if you follow the top pathway of the tree diagram, you get the outcome HH, which in probability notation is $H \cap H$. If you follow the bottom pathway on the tree, you get the outcome TT, written as $T \cap T$. The second pathway from the top, HT (or $H \cap T$), denotes getting a head first and then a tail; the third pathway from the top, TH (or $T \cap H$), denotes getting a tail first and then a head.

Tree diagrams have a wide variety of uses and can help you solve many different types of problems in probability — in particular, problems that give you marginal and conditional probabilities (see Chapter 2). The following sections offer more detail on when tree diagrams are ideal.

Showing multi-stage outcomes with a tree diagram

One of the most common uses for a tree diagram is mapping out the sample space. For probability problems involving a large number of stages or a long sequence of events, you need to be able to picture each stage of the process and to identify and count all the possible outcomes in the sample space.

Suppose, for example, that you're taking orders at a pizza restaurant. The process of taking an order goes as follows:

1. **The customer can order one of three sizes: small (S), medium (M), or large (L).**
2. **The customer can order either thin crust (Tn) or thick crust (Tk).**
3. **The customer can order up to two toppings on the pizza: pepperoni (R) and/or mushrooms (MS).**

The sample space is the result of four stages:

- Pizza size (S, M, L)
- Type of crust (Tn, Tk)
- Pepperoni (R or R^c)
- Mushrooms (MS or MS^c)

The tree diagram for the pizza orders is shown in Figure 3-9. It has a total of $3 * 2 * 2 * 2 = 24$ possible pathways, which match the 24 possible outcomes in the sample space, S. In other words, with these choices, you can make up to 24 different pizzas. For example, the first pathway across the top denotes a small thin-crust pizza with pepperoni and mushrooms; the last pathway across the bottom denotes a large thick-crust pizza with no pepperoni and no mushrooms.

REMEMBER

You may have thought that this tree diagram should have three stages: pizza size (S, M, L), crust type (Tn, Tk), and toppings (pepperoni on one branch, mushrooms on one branch). However, a three-stage tree is wrong because if, for example, you want a pizza with pepperoni and mushrooms, you have a pathway available on the tree to get that outcome. You can use this line of thought to find out if you correctly set up your tree diagram. Each possible outcome in the sample space has to have a separate pathway on the tree that you can follow from beginning to end.

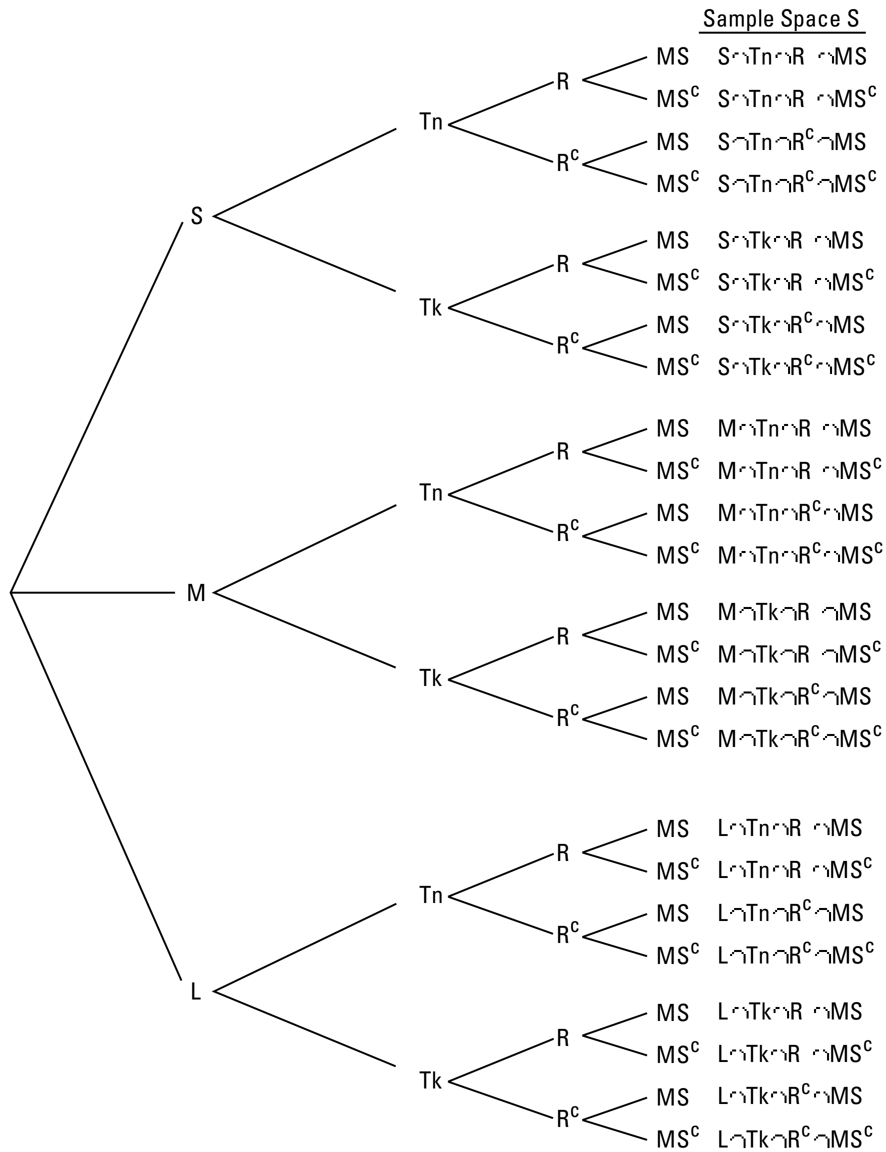


Figure 3-9: Tree diagram of pizza orders.

Organizing conditional probabilities with a tree diagram

After you organize all the possible outcomes in your tree diagram (see the previous section), the next step is to include the probabilities for each individual branch in the tree diagram, using what the problem gives you and the complement rule (see Chapter 2). The marginal probabilities (probability of the event that happened first and its complement) go on the first set of branches, and the conditional probabilities (probability of the second event that occurred, conditioned or dependent on the outcome of the first event) go on the second branches. This second set of branches of the tree diagram is especially useful when the two events are dependent, because the probabilities for the second branch depend on what happened on branch one, and the tree diagram helps you see and work with those conditional probabilities. In this section, you gain practice and insight organizing probabilities for situations where the events are independent or dependent.

Organizing probabilities for independent events

When two events don't influence each other, their outcomes are said to be *independent.* The probabilities for two flips of a coin are shown in Figure 3-10, over the lines that represent each branch. At each stage, you get either heads or tails, and the probabilities are ½ and ½. Because the coins don't influence each other, their outcomes are independent. To get the joint probability (the probability of getting the two outcomes of the coins to occur together), you multiply the marginal probabilities (the probability of an individual outcome on each coin) together. (See Chapter 2 for more information on joint and marginal probabilities.)

For example, to get the probability of the top pathway in the tree, $H \cap H$, you take $P(H) * P(H) = ½ * ½ = ¼$. Each of the probabilities for the branches as well as the joint probabilities for each outcome in the sample space are shown in Figure 3-10. Notice that the joint probabilities all sum to one because the entire sample space is represented on the tree diagram.

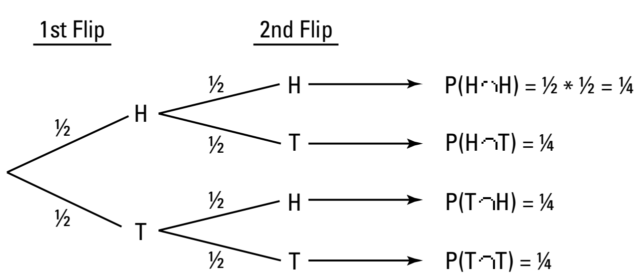


Figure 3-10: Tree diagram of probabilities for flipping a coin twice.

Organizing probabilities for dependent events

Tree diagrams are also useful when you have two dependent events. You know that events are dependent because the conditional probability of A given B isn't equal to the marginal probability of A (see Chapter 2). For example, suppose you have four females and two males in a group, and you need to select two people to be on a committee. You want the process to be democratic, so you select the participants at random; however, you know you can't choose the same person twice. The tree diagram for choosing this committee is shown in Figure 3-11. You're interested in the gender makeup of your committee, so you record the outcomes as M (for male) or F (for female) on each branch; you have two sets of branches because you're choosing two people.

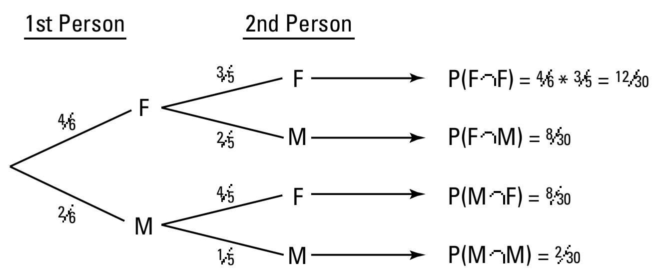


Figure 3-11: A tree diagram for choosing a committee of two people out of a group of six.

Notice that the tree diagram in Figure 3-11 has two branches at the first stage, because you can choose a female or male. For each of those branches, you make your next selection, which is either a female or a male. This gives you a total of $2 * 2 = 4$ possible pathways in the tree diagram, leading to the four possible outcomes in your sample space: S = {FF, FM, MF, MM}.

The difference between this example and flipping a coin twice (see the previous section) isn't in the appearance of the tree diagrams; it's in the probabilities. At the first stage of the committee example, you choose a female (of which there are four) or a male (of which there are two), which means the probability of choosing a female is 4 out of 6, or $\frac{4}{6}$, and the probability of choosing a male is 2 out of 6, or $\frac{2}{6}$. Figure 3-11 shows these stage-one probabilities under "1st Person."

REMEMBER

Note that the probabilities of the stage-one branches add up to one because they're complements of each other. At any given stage in a tree diagram, the sum of all the branches at that particular stage sum to one because they represent all the possible outcomes at that particular stage.

Now you must choose the second person from the group. This is where the probabilities change. You can't choose the same person twice; assuming that you choose a female with your first selection, you have a total of five people left to choose from — three females and two males. Therefore, coming off the first branch at stage one, the probability of choosing a female at stage two is

3 out of 5, or $\frac{3}{5}$, and the probability of choosing a male at stage two is 2 out of 5, or $\frac{2}{5}$. Notice that these two probabilities sum to one, because after you choose a female at stage one, you either choose a female or a male at stage two. So the conditional probabilities for all the second branches that come off of one specific first branch must sum to one. This reflects the fact that once the first event has occurred, the second event either occurs or it doesn't. This takes care of the top two pathways on the tree, resulting in the outcomes $F \cap F$ and $F \cap M$.

Now look at the bottom two pathways on the tree diagram in Figure 3-11. If you choose a male at stage one, you have only five people left to choose from at stage two — one male and four females. So, coming off that first branch of choosing a male at stage one, you can choose a female with probability $\frac{4}{5}$ or a male with probability $\frac{1}{5}$ (again, note that these branches sum to one). The two possible outcomes from these two pathways are $M \cap F$ and $M \cap M$.

The total number of outcomes in this sample space is, therefore, four, but because the two stages aren't independent of each other, the probabilities change from stage one to stage two. And even though this example has four outcomes like in the coin-toss example (see the previous section), the probabilities of the final outcomes aren't equal!

TIP

If you multiply the first branch times the second branch for every possible combination of branches on a tree diagram, the grand total is one. That's because the combinations of branches collectively represent all the possible combinations of the two events than can occur in the entire sample space. (In other words, the sum of the joint probabilities for all possible events has to equal one; see Chapter 2.)

Connecting the tree's branches to the rules of probability

All the elements of a tree diagram have connections to the definitions and rules of probability (see Chapter 2). The stage-one probabilities from the previous section are called *marginal probabilities,* because you look only at the probability of choosing a male or female at stage one. In probability notation, at stage one, you have $P(F) = \frac{4}{6}$ and $P(M) = \frac{2}{6}$.

The stage-two probabilities are called *conditional probabilities,* because they depend on what happens during stage one. In this case, you have $P(F|F) = \frac{3}{5}$ representing the top branch of stage two, where you've already chosen a female and now you're going to choose another female. Similarly, the other stage-two probabilities are noted by $P(M|F) = \frac{2}{5}$, $P(F|M) = \frac{4}{5}$, and $P(M|M) = \frac{1}{5}$. To get the probabilities for the outcomes in the sample space, you have to multiply the stage-one probabilities by the stage-two probabilities, using the multiplication rule (see Chapter 2). For example, to find the probability of choosing two females, you want $P(F \cap F)$, which is equal to $P(F) * P(F|F) = \frac{4}{6} * \frac{3}{5}$ by the multiplication rule. The tree diagram makes setting up your calculations very easy because you just follow the branches that make up the pathway to get you the outcome $F \cap F$ and multiply those probabilities together.

In the coin-flipping example from the section "Organizing probabilities for independent events," you use the same rules and ideas, but because the stages are independent, you don't have any conditional probabilities to worry about. The stage-two probabilities are the same as the stage-one probabilities, so each pathway has the probability $\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$. This only happens if the initial outcomes have equal probability and the stages are independent.

Reviewing the limitations of tree diagrams

Tree diagrams help you most if your sample space is composed of a series of stages or a sequence of events. In these cases, problems typically give you probabilities of the first-stage events by themselves (in other words, marginal probabilities — see Chapter 2) and the conditional probabilities of the second-stage outcomes given the first-stage outcomes (see Chapter 2). You can then use a tree diagram to help you organize and enumerate all the outcomes of a multi-stage sample space that help you find probabilities of any pathway or combination of pathways on the tree. However, tree diagrams can't help you solve every probability problem.

Tree diagrams don't work well if problems give you *intersection probabilities* (the probability of two events occurring at the same time — see Chapter 2) rather than conditional probabilities. They also don't help if you can't break down your sample space into a series of stages or a sequence of events. In such cases, a Venn diagram (see the section "Diagramming Probabilities with Venn Diagrams" earlier in this chapter) is much more helpful.

REMEMBER

Before you start a probability problem, think about the information you're given and determine the best way to organize the information in a picture format. You generally use Venn diagrams when problems give you marginal and intersection probabilities, such as $P(A \cap B)$, and you have to find conditional probabilities, probabilities of combinations, and/or complements of those events. You generally use tree diagrams when problems give you marginal probabilities, such as $P(A)$, and conditional probabilities, such as $P(A|B)$, and you have to find probabilities of intersections, unions, or other conditional probabilities, such as $P(B|A)$.

Drawing a tree diagram to find probabilities for complex events

You can use a tree diagram to map out and find probabilities for complicated events that come from a sequence of events or stages in a process. Suppose, as you do in the section "Organizing probabilities for dependent events," that you have to choose two people to form a committee from a group of six — four females and two males. You want the probability of choosing

- Exactly one female
- At least one female
- Two people of the same gender

You can use the tree diagram and its probabilities shown in Figure 3-11 to address each of these goals.

Example 1: The probability of choosing exactly one female

To find the probability of choosing exactly one female, you have to think about what this probability means in terms of outcomes. You have two ways to get exactly one female: choosing a female first and then a male ($F \cap M$), as indicated by the second pathway on the tree (from the top), or choosing a male first and then a female ($M \cap F$), as indicated by the third pathway on the tree (from the top). If you put the second and third pathways together, you get the event that you choose exactly one female, because you have two ways to choose exactly one female: $F \cap M$ or $M \cap F$. The probability of choosing exactly one female is the probability of $(F \cap M) \cup (M \cap F)$.

To find this probability, you use the multiplication rule and the addition rule (see Chapter 2). Because these two events are mutually exclusive (they have no elements in common), the addition rule tells you to add the probabilities together to get $(\frac{4}{6} * \frac{2}{5}) + (\frac{2}{6} * \frac{4}{5}) = \frac{16}{30}$, or 0.53.

In terms of a tree diagram, you define union probabilities as sums of the probabilities of the pathways making up the unions. After you calculate the probabilities for all the outcomes in the sample space, you simply add them together for the elements that pertain to the event you want.

Example 2: The probability of choosing at least one female

To obtain the probability of choosing at least one female according to the tree diagram in Figure 3-11, you have to figure out what "at least one" means. You must choose two people to form your committee, so you can obtain zero, one, or two females. You want at least one female, which means one or more females. In other words, you want the probability of choosing exactly one female or two females.

The probability of exactly one female is the sum of pathways two and three — $\frac{16}{30}$ by your calculations from Example 1. The probability of two females is the probability of the first (top) pathway on the tree: $F \cap F$, which is $\frac{4}{6} * \frac{3}{5} = \frac{12}{30}$. Using the addition rule (see Chapter 2), you add these probabilities together to get $\frac{16}{30} + \frac{12}{30} = \frac{28}{30}$, or 0.93 for the probability of you choosing at least one female from the group. This probability is high because you have twice as many females as males to choose from in the original group of six.

Another way to solve Example 2 is to use the rule of complements (see Chapter 2). At least one female is the complement of zero females. The event of zero females is the fourth pathway on the tree in Figure 3-11, $M \cap M$. So, P(at least 1 female) = 1 – P(0 females) = $1 - (\frac{2}{6} * \frac{1}{5}) = 1 - \frac{2}{30} = \frac{28}{30}$, or 0.93.

Example 3: The probability that both people you choose have the same gender

To find the probability that both people you choose have the same gender, you break the probability down into the individual outcomes that this goal represents. You can choose both females — ($F \cap F$), represented by the top pathway on the tree diagram in Figure 3-11 — or you can choose both males — ($M \cap M$), represented by the fourth (bottom) pathway on the tree diagram. The probability of the top pathway is $\frac{4}{6} * \frac{3}{5} = \frac{12}{30}$, and the probability of the bottom pathway is $\frac{2}{6} * \frac{1}{5} = \frac{2}{30}$. Using the addition rule (see Chapter 2), you add these probabilities together to get $\frac{12}{30} + \frac{2}{30} = \frac{14}{30}$, or 0.47.

The Law of Total Probability and Bayes' Theorem

Two situations come up quite often in probability courses, and both of them come up in the context of multi-stage sample spaces. The first situation involves finding a marginal probability (see Chapter 2) for an event A, P(A), when you're given several conditional probabilities and marginal probabilities for events involved with A but not the direct probability of A. For example, suppose the satisfaction rates at three local grocery stores in a certain area appear in the newspaper, along with the percentage of business each store gets. How can you find the overall satisfaction rate, regardless of which store a customer goes to? You can solve this type of problem by using the Law of Total Probability.

The other situation involves finding the conditional probability (see Chapter 2) of event A given event B, P(B), when you know $P(B|A)$ and its complements, as well as the marginal probability for B and its complement. Sticking with the grocery example, suppose you know a person went grocery shopping in this area and came away satisfied. You want to know which store the person most likely went to (after the fact). You can solve this type of problem by using Bayes' Theorem.

In the following pages, you discover how to use the formulas for the Law of Total Probability and Bayes' Theorem. The formulas may appear daunting if you just look at them cold-turkey, but after you understand what's going on behind the scenes and how you use the information you're given to find the probability you want, the formulas don't seem nearly as complicated.

Finding a marginal probability using the Law of Total Probability

Sometimes a problem gives you several different conditional probabilities and/or intersection probabilities (see Chapter 2) that all involve an event A, but it never gives you the probability of A, P(A). And, of course, you need to find P(A), because most problems ask you to. For example, you may know the probability that a person will be late given that she flew on airline A, given that she flew on airline B, or given that she flew on airline C, but you have to find the overall probability of her being late, regardless of what airline she flew on. This is where the Law of Total Probability comes in.

The *Law of Total Probability* says that the marginal probability of an event happening at stage two is equal to the sum of the products of the marginal (stage one) and conditional (stage two given stage one) probabilities over all the possible ways to achieve the event. For this example, you total up all the probabilities from the various conditional scenarios, weighted by the proportion of time each one occurs. For example, if the girl is late 60 percent of the time flying airline A, but airline A only gets 5 percent of the overall flights, this is a small issue in the grand scheme of being late. A big issue arises if the girl is late 60 percent of the time on airline A, and airline A gets 90 percent of the business.

To picture this point, suppose that customers have three restaurants to choose from in a certain town: Restaurant 1, Restaurant 2, and Restaurant 3. Previous data collection has shown that these restaurants get 50 percent, 30 percent, and 20 percent of the business, respectively. Suppose you also know that 70 percent of the customers who dine at Restaurant 1 are satisfied (and 30 percent are not); 60 percent of the customers who dine at Restaurant 2 are satisfied (and 40 percent are not); and 50 percent of the customers who dine at Restaurant 3 are satisfied (and 50 percent are not). Now, suppose you want to know the probability that someone who eats at a restaurant in this town will be satisfied — in other words, you want to know the probability of an outcome from stage two. To find this probability, you formalize the Law of Total Probability, set up a tree diagram, and put the probabilities together to get your answer.

Formalizing the Law of Total Probability

Using set notation, the formula for the Law of Total Probability is the following: $P(B) = \sum_i P(A_i) * P(B \mid A_i)$.

To use the Law of Total Probability, you add up the probabilities of all the pathways that lead you to event B at stage two. For the restaurant example problem posed in the previous section, A_i represents each of the stage-one events that you go through to achieve event B at stage two. For each event A_i, you multiply $P(A_i)$ times $P(B \mid A_i)$ to get the probability of that particular pathway, and then you sum up the probabilities of all the pathways to get the total probability of event B.

For some people, the formula for the Law of Total Probability is too confusing and intimidating to be useful, but don't give up! As long as you set up the branches and probabilities of your tree diagram properly, the law makes solving complicated problems much easier. The trick is remembering when to use the Law of Total Probability: When the sample space comes in stages, and you want the total marginal probability of an event happening at stage two, fire away with the law.

For the restaurant example, you're interested in the overall probability of a customer being satisfied, regardless of which restaurant the customer eats at. You know the percentage of business each restaurant gets, which means you know how much of an impact each restaurant contributes to the overall satisfaction for the town, so you use the Law of Total Probability to figure out the overall satisfaction rate. But before you jump into the calculations, it helps to map the situation out.

Setting up the tree diagram

The first step in using the Law of Total Probability is to map out the events and the probabilities with a tree diagram. Using the three-restaurant example I present earlier in this section, the customer can go to one of the three restaurants (stage one); after the customer attends one of the restaurants, he will either be satisfied or unsatisfied (stage two). The tree diagram of this situation is shown in Figure 3-12, where R1, R2, and R3 represent the three restaurants, S denotes that the customer is satisfied, and S^c denotes that the customer is unsatisfied.

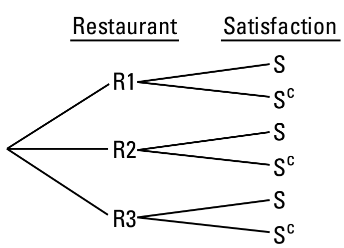


Figure 3-12: A tree diagram illustrating a customer's satisfaction.

Plugging in the probabilities

After you set up your tree diagram, you fill in the probabilities. At the first stage of the restaurant example, the customer goes to Restaurant 1 with a probability of 50 percent, because 0.50 of the customers go to that restaurant. In a similar fashion, the marginal probabilities for R2 and R3 are 0.30 and 0.20, respectively. Branching off of R1, the probability of the customer being satisfied is 70 percent, because 70 percent of the customers who go to Restaurant 1 are satisfied.

Note: This is a conditional probability, because you know the customers went to Restaurant 1 (the given part), and you want the probability that the customers are satisfied. Similarly, off of R1, the branch indicating that the customer is unsatisfied is 1 – 0.70 = 0.30, because being satisfied and unsatisfied at Restaurant 1 are compliments of each other.

Next, $P(S \mid R2) = 0.60$, because 60 percent of the customers who go to Restaurant 2 are satisfied, which means that $P(S^c \mid R2) = 1 - 0.60 = 0.40$.

For Restaurant 3, the probability of the customer being satisfied is 0.50 — $P(S \mid R3) = 0.50$ — which means that $P(S^c \mid R3)$ is the probability of being unsatisfied with Restaurant 3, 1– 0.50 = 0.50. The probabilities for this tree diagram are shown in Figure 3-13.

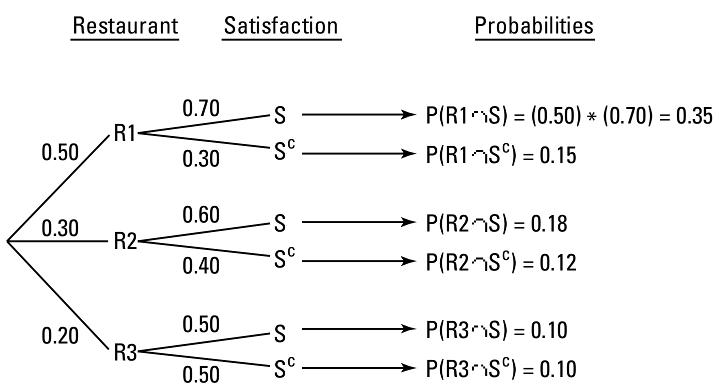


Figure 3-13: Probabilities of satisfaction for the three restaurants on the tree diagram.

It helps if you set up all the joint probabilities for each of the branches right away before you try to answer the question. Using the multiplication rule (see Chapter 2), you multiply the stage-one probability by the stage-two-given-stage-one probability for each branch. All the joint probabilities are shown in Figure 3-13.

Putting probabilities together to find the total probability

The example problem presented at the beginning of this section calls for the probability that a customer who eats at a restaurant in this town will be satisfied. The customer's satisfaction can happen in one of three ways:

- The customer goes to Restaurant 1 and is satisfied $(R1 \cap S)$.
- The customer goes to Restaurant 2 and is satisfied $(R2 \cap S)$.
- The customer goes to Restaurant 3 and is satisfied $(R3 \cap S)$.

These events don't share any common elements, so you can find the probability of their union by adding up all their probabilities, using the addition rule (see Chapter 2). This gives you P(S), the probability that a customer who eats at a restaurant in the town will be satisfied.

Each of the probabilities needed represents a pathway on the tree:

- The top pathway represents customers going to Restaurant 1 and coming away satisfied.
- Pathway three (from the top) represents customers going to Restaurant 2 and coming away satisfied.
- Pathway five (from the top) represents customers going to Restaurant 3 and coming away satisfied.

These probabilities, respectively, are $0.50 * 0.70 = 0.35$, $0.30 * 0.60 = 0.18$, and $0.20 * 0.50 = 0.10$. Adding the probabilities gives you $P(S) = 0.35 + 0.18 + 0.10 = 0.63$. So, the chance that a customer will be satisfied after eating at a restaurant in the town is 63 percent.

Finding the posterior probability with Bayes' Theorem

The *posterior probability* is a conditional probability of A given B where A actually occurs first. For example, suppose doctors have a hard time detecting a certain disease without a blood test. In reality, the patient either has the disease or doesn't have it. Given that the patient has the disease, doctors hope the probability of testing positive is high and the probability of testing negative is low. The people who design the blood test want to know the probability of testing positive given that the patient has the disease. But the doctor wants to know something different; she wants the probability that the patient has the disease given that he tests positive. The doctor wants the posterior probability — a probability found after the fact, in the opposite direction from how the data actually occurs.

Some probability problems give you the probability of B given A, but you have to find the conditional probability of A given B, written $P(A|B)$ — the probability that A happened at stage one given that B happened at stage two. To find this probability, for students and the doctors mentioned previously, you use *Bayes' Theorem,* which says that if you want $P(A|B)$ — the probability in the opposite order of what's given on the tree diagram — you do the following:

1. **Find the probability of the pathway that goes through A and B.**
2. **Divide by the total probability of all pathways that lead to B.**

For example, suppose (as you do in the previous sections for the example problem) that a customer has three restaurants to choose from in a certain town: Restaurant 1, Restaurant 2, and Restaurant 3. Previous data collection has shown that these restaurants get 50 percent, 30 percent, and 20 percent of the business, respectively. Suppose you know that 70 percent of the customers who dine at Restaurant 1 are satisfied (and 30 percent are not); 60 percent of the customers who dine at Restaurant 2 are satisfied (and 40 percent are not); and 50 percent of the customers who dine at Restaurant 3 are satisfied (and 50 percent are not). Now, suppose that you find out a customer is satisfied with his or her restaurant experience. You have two questions to answer:

- What's the chance that he or she ate at Restaurant 2, given that he or she is satisfied (the posterior probability)?
- Assuming that the customer is satisfied, which of the three restaurants was he or she more likely to have eaten at?

To answer the questions, you formalize Bayes' Theorem, set up a tree diagram, put the probabilities into the diagram, and find the probabilities that result, using Bayes' Theorem.

Formalizing Bayes' Theorem

Formally, Bayes' Theorem is given by $P\left(A_i|B\right)=\dfrac{P(A_i)*P\left(B|A_i\right)}{\sum_i P(A_i)*P\left(B|A_i\right)}$, where A_i is the particular event at stage one that has occurred, given that you end up at event B after stage two. The numerator is equal to $P(A \cap B)$, the probability that goes through A and B. The denominator is the total probability of all the pathways that lead to B.

TIP

As with the Law of Total Probability (see the section "Formalizing the Law of Total Probability"), you have to make sure you set up the branches and probabilities of your tree diagram properly, which makes finding problems with Bayes' Theorem much easier. The trick is remembering when to use Bayes' Theorem: When the sample space comes in stages and you want the probability of an event in the opposite order of which the problem gives it to you, bust out the Theorem.

Here's the big picture for the restaurant example I present at the beginning of this section. You know the individual satisfaction rates for each restaurant. You also know the percentage of business each restaurant gets, which represents the weight that each restaurant's satisfaction rate has in terms of the overall satisfaction rating for the town. Now, suppose you're the mayor of the town, and you run into someone in another state who claims to have eaten at a restaurant in your town and left satisfied. You secretly wonder which restaurant it was. How can you find out? By using Bayes' Theorem. In a way, Bayes' Theorem lets you be sneaky and work backward to answer your own questions. But before you launch into the calculations, take your time and map the information out.

Setting up the tree diagram and putting in the probabilities

The first step in solving the restaurant probability problem set up in the previous pages is to set up a tree diagram and the corresponding probabilities for the branches of the tree. The tree diagram you have to use for this problem is the same as Figure 3-13, because the basic information from the example problems is the same.

Answering Question 1: Finding the posterior probability

Suppose that a customer already ate at one of the three restaurants and is satisfied. This means that event S occurred at stage two of the process, but you don't know what happened at stage one (in other words, which restaurant the customer ate at). The question at hand is: Given that the customer is satisfied, what's the probability that the customer ate at Restaurant 2? In other words, you want the conditional probability $P(R2 \mid S)$. Using the definition of conditional probability (the probability of A given B equals the probability of A and B divided by the probability of B; see Chapter 2), you get the following: $P(R2|S) = \frac{P(R2 \cap S)}{P(S)}$.

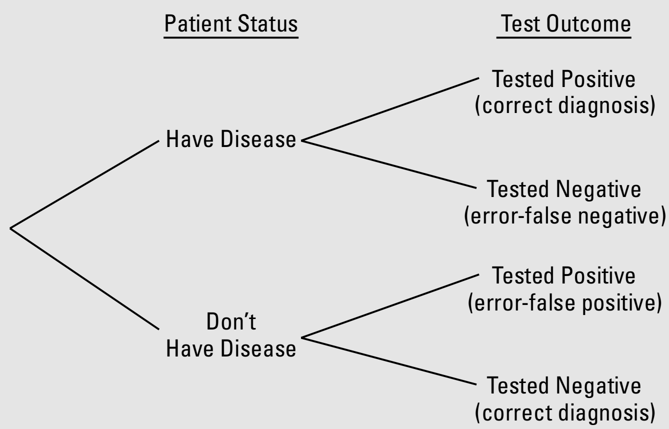
Notice that the denominator of the equation is P(S), which you find in the section "Putting probabilities together to find the total probability." You know it's P(S) because you know the customer is satisfied, so that becomes your new denominator. The numerator is $P(R2 \cap S)$, which is the probability that the customer went to Restaurant 2 and left satisfied. The numerator here is equal to the joint probability (of R2 and S) you find by multiplying P(R2) by $P(S \mid R2)$ — the two branches that make up the third pathway on the tree diagram.

You know that the customer was satisfied at stage two, so he or she got there on pathway one (from the top), pathway three (from the top), or pathway five (from the top) on the tree diagram. The total probability of being satisfied at stage two is the sum of the probabilities on those branches (see the section "Putting probabilities together to find the total probability"). Now, you want the probability that the customer went to Restaurant 2 given that he or she left satisfied, so you take the probability of pathway three (the one representing the customer going to Restaurant 2 and leaving satisfied) and divide it by the total probability of S, P(S): $(0.30 * 0.60) \div ([0.50 * 0.70] + [0.30 * 0.60] + [0.20 * 0.50])$. So, you get $0.18 \div 0.63 = 0.286$, or 28.6 percent. The probability of the customer dining at Restaurant 2, given that he or she left satisfied, is 28.6 percent.

PROBABILITY IN ACTION

Using Bayes' Theorem for disease testing

Medical researchers often use Bayes' Theorem to determine how effective their tests are at detecting certain diseases. In order to check out their tests, they try them on people whom they know have the disease and on people whom they know don't have the disease. So, stage one of the tree diagram is "have the disease and don't have the disease." On each group, the test results are either positive or negative, which is the second stage of the tree diagram. See the following figure for the tree diagram representing this situation.



If you want to find the probability that a test gives the correct diagnosis, you can use the same idea as the Law of Total Probability and add the probabilities for the first pathway in the tree (person has the disease and tests positive) and the fourth pathway (person doesn't have the disease and tests negative). However, in the case of medical tests, testing positive isn't always a correct diagnosis. If you sum the probabilities for pathways one and three, you get the total probability of testing positive (whether you have the disease or not). This probability is helpful, too.

The big question, however, is how effective the test is at detecting the disease. Suppose someone tests positive for the disease — what's the chance that he or she actually has the disease? In probability notation, you want P(have the disease|tested positive). Notice that this is the opposite order from how you collect and organize the information on the tree diagram? That means you need Bayes' Theorem to solve it because it's a posterior probability. Looking at the previous figure, you take the probability of pathway one divided by the sum of the probabilities for pathway one and pathway three, because you know the person tested positive. The probability of having the disease is the probability of pathway one (having the disease and testing positive). So, you divide the probability of pathway one by the probability of testing positive (pathway one + pathway three).

Answering Question 2: Which restaurant did the satisfied customer most likely go to?

In the example involving the three restaurants and customer satisfaction, you need to answer the following question: Assuming the customer is satisfied, which of the three restaurants was he or she more likely to have eaten at? Answering this question amounts to figuring and comparing three probabilities: $P(R1 \mid S)$, $P(R2 \mid S)$, and $P(R3 \mid S)$ — the probabilities that the customer ate at Restaurant 1, 2, or 3, respectively, given that the customer was satisfied. You can find each of these probabilities by using Bayes' Theorem because they're all posterior probabilities.

You find $P(R2 \mid S)$ in the previous section: 0.286. Now, $P(R1 \mid S) = \frac{P(R1 \cap S)}{P(S)} = \frac{P(R1) * P(S|R1)}{P(S)} = (0.50 * 0.70) \div 0.63 = 0.35 \div 0.63 = 0.556$, by Bayes' Theorem. $P(R1 \mid S)$ is the probability of being on pathway one of the tree, divided by the sum of pathways one, three, and five (the total probability of being satisfied). Finally, the probability of the customer having gone to Restaurant 3 given that he or she left satisfied is the probability of being on pathway five of the tree, divided by $P(S)$, which comes out to $P(R3|S) = \frac{P(R3 \cap S)}{P(S)} = \frac{P(R3) * P(S|R3)}{P(S)} = (0.20 * 0.50) \div 0.63 = 0.10 \div 0.63 = 0.159$, by Bayes' Theorem.

The three posterior probabilities you find sum to one (subject to round-off error). So, after you find $P(R1 \mid S)$ and $P(R2 \mid S)$, you know that $P(R3 \mid S)$ is one minus the sum of those two probabilities because they're complements. R3 is the last probability you calculate because it's a complement of the others. Remembering this may take some time off your calculations.

Because $P(R1 \mid S) = 0.556$ is the highest of the three probabilities, you make your conclusion: Given that the customer left satisfied, he or she was most likely to have eaten at Restaurant 1. This makes sense because Restaurant 1 gets the most business and has the highest customer satisfaction rating.