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.

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.

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.