1

Computing Probabilities: Right Ways and Wrong Ways

THE PROBABILIST

Whether you like it or not, probabilities rule your life. If you have ever tried to make a living as a gambler, you are painfully aware of this, but even those of us with more mundane life stories are constantly affected by these little numbers. Some examples from daily life where probability calculations are involved are the determination of insurance premiums, the introduction of new medications on the market, opinion polls, weather forecasts, and DNA evidence in courts. Probabilities also rule who you are. Did daddy pass you the X or the Y chromosome? Did you inherit grandma's big nose? And on a more profound level, quantum physicists teach us that *everything* is governed by the laws of probability. They toss around terms like the *Schrödinger wave equation* and *Heisenberg's uncertainty principle,* which are much too difficult for most of us to understand, but one thing they do mean is that the fundamental laws of physics can only be stated in terms of probabilities. And the fact that Newton's deterministic laws of physics are still useful can also be attributed to results from the theory of probabilities. Meanwhile, in everyday life, many of us use probabilities in our language and say things like "I'm 99% certain" or "There is a one-in-a-million chance" or, when something unusual happens, ask the rhetorical question "What are the odds?"

Some of us make a living from probabilities, by developing new theory and finding new applications, by teaching others how to use them, and occasionally by writing books about them. We call ourselves *probabilists*. In the universities, you find us in mathematics and statistics departments; there are no departments of probability. The terms “mathematician” and “statistician” are much more well known than “probabilist,” and we are a little bit of both but we don’t always like to admit it. If I introduce myself as a mathematician at a cocktail party, people wish they could walk away. If I introduce myself as a statistician, they do. If I introduce myself as a probabilist...well, most actually still walk away. They get upset that somebody who sounds like the Swedish Chef from the Muppet Show tries to impress them with difficult words. But some stay and give me the opportunity to tell them some of the things I will now tell you about.

Let us be etymologists for a while and start with the word itself, *probability.* The Latin roots are *probare*, which means to test, to prove, or to approve, and *habilis*, which means apt, skillful, able. The word “probable” was originally used in the sense “worthy of approval,” and its connection to randomness came later when it came to mean “likely” or “reasonable.” In my native Swedish, the word for probable is “sannolik,” which literally means “truthlike” as does the German word “wahrscheinlich.” The word “probability” still has room for nuances in the English language, and Merriam-Webster’s online dictionary lists four slightly different meanings. To us a probability is a number used to describe how likely something is to occur, and probability (without indefinite article) is the study of probabilities.

Probabilities are used in situations that involve randomness. Many clever people have thought about and debated what randomness really is, and we could get into a long philosophical discussion that could fill the rest of the book. Let’s not. The French mathematician Pierre-Simon Laplace (1749–1827) put it nicely: “Probability is composed partly of our ignorance, partly of our knowledge.” Inspired by Monsieur Laplace, let us agree that you can use probabilities whenever you are faced with uncertainty. You could:

- Toss a coin, roll a die, spin a roulette wheel
- Watch the stock market, the weather, the Super Bowl
- Wonder if there is an oil well in your backyard, if there is life on Mars, if Elvis is alive

These examples differ from each other. The first three are cases where the outcomes are equally likely. Each individual outcome has a probability that is simply one divided by the number of outcomes. The probability is $1/2$

to toss heads, $1/6$ to roll a 6, and $1/38$ to get the number 29 in roulette (an American roulette wheel has the numbers 1–36, 0, and 00). Pure and simple. We can also compute probabilities of groups of outcomes. For example, what is the probability to get an odd number when rolling a die? As there are three odd outcomes out of six total, the answer is $3/6 = 1/2$. These are examples of *classical probability*, the first type of probability problems studied by mathematicians, most notably, Frenchmen Pierre de Fermat and Blaise Pascal whose seventeenth century correspondence with each other is usually considered to have started the systematic study of probabilities. You will learn more about Fermat and Pascal later in the book.

The next three examples are cases where we must use data to be able to assign probabilities. If it has been observed that under current weather conditions it has rained about 20% of the days, we can say that the probability of rain today is 20%. This probability may change as more weather data are gathered and we can call it a *statistical probability*. As for the 2006 Super Bowl, I placed a bet on the Houston Texans that gave odds of 800 to 1, which means that the bookmaker assigned a probability of less than $1/800$ that the Texans would win. However he came to this conclusion, he must have used plenty of data other than that he once spent a summer in Houston and almost died of heatstroke.

The third trio of examples is different from the previous two in the sense that the outcome is already fixed; you just don't know what it is. Either there is an oil well or there isn't. Before you start drilling, you still want to have some idea of how likely you are to find oil and a geologist might tell you that the probability is about 75%. This percentage does not mean that the oil well is there nine months of the year and slides over to your neighbor the other three, but it does mean that the geologist thinks that your chances are pretty good. Another geologist may tell you the probability is 85%, which is a different number but means the same thing: Chances are pretty good. We call these *subjective probabilities*. In the case of a living Elvis, I suppose that depending on whom you ask you would get either 0% or 100%. I mean, who would say 25%? Little Richard?

Some knowledge about proportions may be helpful when assigning subjective probabilities. For example, suppose that your Aunt Jane in Pittsburgh calls and tells you that her new neighbor seems nice and has a job that "has something to do with the stars, astrologer or astronomer." Without having more information, what is the probability that the neighbor is an astronomer? As you have virtually no information, would you say 50%? Some people might. But you should really take into account that there are about four times as many astrologers as astronomers in the United States, so a probability of 20% is more realistic. Just because something is “either/or” does not mean it is “50–50.” Andy Rooney may have been more insightful than he intended when he stated his 50–50–90 rule: “Anytime you have a 50–50 chance of getting something right, there’s a 90% probability you’ll get it wrong.”