

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 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, prove, or 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 the 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, or spin a roulette wheel

- Watch the stock market, the weather, or 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*, which is 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 Super Bowl, at the time of writing, the 2014–2015 season has yet to start and the highest odds are on the Jacksonville Jaguars at 100 to 1. This means that the bookmaker estimates a probability of less than 1% that the Jaguars will win, an estimate based on plenty of team data and football statistics.

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 9 months of the year and slides over to your neighbor the other 3, 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,

4 COMPUTING PROBABILITIES: RIGHT WAYS AND WRONG WAYS

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.”

THE PROBABILIST’S TOYS AND LANGUAGE

Probabilists love to play with coins and dice. In a platonic sense. We like the *idea* of tossing coins and rolling dice as experiments that have equally likely outcomes. Suppose that a family with four children is chosen at random. What is the probability that all four are girls? A coin-tossing analogy would be to ask for the probability to get four heads when a coin is tossed four times. Many probability problems can be illustrated by coin tossing, but this would quickly become boring, so we introduce variation by also rolling dice, spinning roulette wheels, picking balls from urns, or drawing from decks of cards. Dice, roulette, and card games are also interesting in their own right, and you will find a chapter on gambling later in the book. Of course. Probability without gambling is like beer without bubbles.

Probability is the art of being certain of how uncertain you are. The statement “the probability to get heads is $1/2$ ” is a precise statement. It tells you that you are as likely to get heads as you are to get tails. Another way to think about probabilities is in terms of average long-term behavior. In this case, if you toss the coin repeatedly, in the long run you will get roughly 50% heads and 50% tails. Of this you can be certain. What you cannot be certain of is how the next toss will come up.

Probabilists use special terminology. For example, we often refer to a situation where there is uncertainty as an “experiment.” This situation

could be an actual experiment such as tossing a coin or rolling a die, but also something completely different such as following the stock market or watching the Wimbledon final. An experiment results in an *outcome* such as “heads,” “6,” “Volvo went up,” or “Björn Borg won” (those were the days). A group of outcomes is called an *event*. In plain language, an event is something that can happen in an experiment. It can be a single outcome (roll 6) or a group of outcomes (roll an odd number). The mathematical description of an event is that it is a *subset* of the set of all possible outcomes, and mathematicians would describe outcomes as *elements* of this set. Probabilists use the words “outcome” and “event” to emphasize the connection with things that happen in reality. In formulas, we denote events by uppercase letters and use the letter “P” to denote probability. The mathematical expression $P(A)$ should thus be read “the probability of (the event) A.” We may also talk about the probability of a statement rather than an event. However, it is mere language; the verbal description of an event is, of course, a statement.

The set of all possible outcomes is called the *sample space*.¹ Sometimes there is more than one choice of sample space. For example, suppose that you toss two coins and ask for the probability that you get two heads. As the number of heads can be 0, 1, or 2, you might be tempted to take these three numbers as the sample space and conclude that the probability to get two heads is $1/3$. However, if you repeated this experiment, you would notice after a while that you tend to get two heads less than one-third of the tosses. The problem is that your sample space consists of three outcomes that are *not equally likely*. Let us distinguish between the two coins by painting one red and the other blue. There are then four equally likely outcomes: both show heads; the red shows heads and the blue shows tails; the red shows tails and the blue shows heads; and both show tails. In a more convenient notation, our sample space consists of the four equally likely outcomes HH, HT, TH, and TT. One out of four gives two heads, and the correct probability is $1/4$. See Figure 1.1 for an illustration of the four equally likely outcomes.

Here is a similar problem. If you roll two dice, what is the probability that the sum of the two equals eight? First note that the sum of two dice

¹The term “sample space” was coined by mathematician and Austro-Hungarian fighter pilot Richard von Mises. That is, he coined the German term *Merkmahlraum* (label space), which appears in his 1931 book with the impressive German title *Wahrscheinlichkeitsrechnung* (probability calculus).

6 COMPUTING PROBABILITIES: RIGHT WAYS AND WRONG WAYS

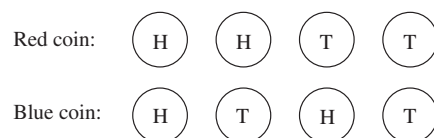


Figure 1.1 The four equally likely outcomes when you toss two coins.

can be any of the numbers 2, 3, ..., 12 but that these are not equally likely. To find the equally likely outcomes, we need to distinguish between the two dice, for example, by pretending that they have different colors, red and blue, just like we did with the two coins above, and consider 36 possible outcomes. As the sum can be 8 by adding $2 + 6$, $3 + 5$, or $4 + 4$, we might first think that there are 3 possibilities out of 36 to get sum 8, but we also need to distinguish, for example, between the cases “blue die equals 2 and red die equals 6” on the one hand and “blue die equals 6 and red die equals 2” on the other. If we make this distinction, we realize that there are five ways to get sum 8 and the probability is $5/36$. See Figure 1.2 for an illustration of the sample space of 36 equally likely outcomes and the event that the sum equals 8.

Here is another example of a similar nature. Consider a randomly chosen family with three children. What is the probability that they have exactly one daughter? There can be 0, 1, 2, or 3 girls, but you know by now that these are not equally likely. Instead, distinguish the kids by birth

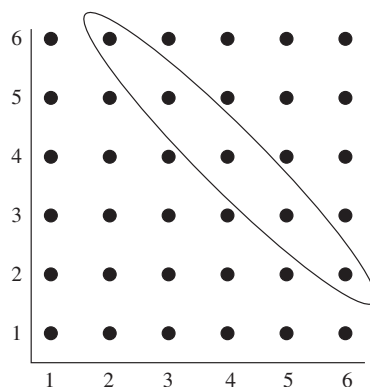


Figure 1.2 The sample space of 36 equally likely outcomes when you roll two dice. The event that the sum equals eight is marked; note that it consists of five outcomes because there are two ways to get 2 and 6 as well as 3 and 5 but only one way to get 4 and 4.

order so that, for example, BGB means that the first child is a boy, the second a girl, and the third a boy. The eight equally likely outcomes are as follows:

BBB, BBG, BGB, GBB, BGG, GBG, GGB, GGG

We're on easy street now; just note that three of the eight outcomes have one girl, and the probability of exactly one girl is therefore $3/8$. Now consider a randomly chosen girl who has two siblings. What is the probability that she has no sisters? This situation looks similar. If she has no sisters, this means that her family has three children, exactly one of whom is a girl and we just saw that the probability of this is $3/8$. Convinced? You should not be. This situation is different because we are not choosing a *family* with three children; we are choosing a *girl* who belongs to such a family. Thus, the outcome BBB is impossible. Is the probability then $3/7$? Think about this for a while before you read on.

I hope you answered no. We need a completely new sample space that also accounts for the chosen girl. If we denote her by an asterisk, the 12 equally likely outcomes are as follows:

BBG*, BG*B, G*BB, BG*G, BGG*, G*BG

GBG*, G*GB, GG*B, G*GG, GG*G, GGG*

and the probability that she has no sisters is $3/12=1/4$. Note how the previous outcomes are now split up according to how many girls they contain. The one with three girls, GGG, is split up into three equally likely outcomes because either of the three girls may be the chosen one. The probabilities that we have computed show that 37.5% of three-children families have exactly one daughter and 25% of girls from three-children families have no sisters.

What is the probability that all three children are of the same gender? Consider the following *faulty* argument: Two children must always be of the same gender. Whatever this gender is, the third child is equally likely to be of this gender or not, and thus the probability that all three are of the same gender is $1/2$. This example is a variant of a coin-tossing problem given by the British nobleman and amateur scientist Sir Francis Galton (about whom you will learn more in chapters to come) in 1894 to illustrate the dangers of sloppy thinking. Use our first sample space to discover the error, and argue that the correct probability is $1/4$.

8 COMPUTING PROBABILITIES: RIGHT WAYS AND WRONG WAYS

Let us next consider an old gambling problem that goes along the same lines. I have three dice and offer you even odds to play the following game: The dice are rolled, and their sum is computed. If the sum is 9, you win. If it is 10, I win. If it is neither, I roll again. Is this game fair?

There are six ways in which the sum can be 9:

$$1 + 2 + 6, 1 + 3 + 5, 1 + 4 + 4, 2 + 2 + 5, 2 + 3 + 4, 3 + 3 + 3$$

and likewise there are six ways to get sum 10:

$$1 + 3 + 6, 1 + 4 + 5, 2 + 2 + 6, 2 + 3 + 5, 2 + 4 + 4, 3 + 3 + 4$$

It sure looks like the game is fair, but beware, in the long run, I would slowly but surely win your money. But why?

Before you decide to play, you need to first identify the equally likely outcomes. And just like in the case of the two dice earlier, it is helpful to imagine that the three dice have three different colors, for example, red, green, and blue. If we list the dice in this order, the equally likely outcomes are (1,1,1), (1,1,2), (1,2,1), (2,1,1), (2,2,1), and so on until (6,6,6); a moment's thought reveals that there are $6 \times 6 \times 6 = 216$ of them. Let us look at one of the ways to get sum 9, $1 + 4 + 4$. This sum corresponds to three of the equally likely outcomes: (1,4,4), (4,1,4), and (4,4,1). If we instead consider $1 + 2 + 6$, this corresponds to six outcomes: (1,2,6), (1,6,2), (2,1,6), (2,6,1), (6,1,2), and (6,2,1). In general, if all three dice show different numbers, this can occur in six ways; if two show the same number, this can occur in three ways; and if all three are the same, this can only occur in one way.

Now count above to realize that 27 outcomes give sum 10 and only 25 give sum 9. The tie-breaker is the last outcome: There is only one way to combine $3 + 3 + 3$ but three ways to combine $3 + 3 + 4$; see Figure 1.3 for an illustration. Thus, out of the 52 outcomes that give a winner, I win in 27, or about 52%, and you win in the remaining 25, or 48%. Not a big difference, but it would be enough to make a living (some venture capital needed).

I mentioned that this problem is an old one. It was in fact solved almost 400 years ago by the great astronomer and telescope builder Galileo after being approached by a group of gambling Florentine noblemen. It is amusing to imagine how the world's most brilliant scientist of his time spent time helping people with their gambling problems. Good thing

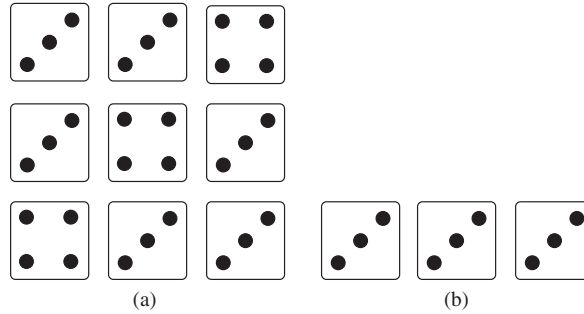


Figure 1.3 Three ways to get sum 10 from two 3s and a 4 (a), but only one way to get sum 9 from three 3s (b).

for Einstein that there were no casinos in Atlantic City in the 1930s; his Princeton office might have been flooded by gamblers having spent the last of their money on a bus ticket, desperate for help from the genius.

We are often interested in more than one event. For example, suppose that people are chosen for an opinion poll and asked about their smoking habits and political sympathies. Consider one selected person. Let us denote the event that she is a smoker by S and the event that she is a Republican by R . We can then make up new events. The event that she is a smoker *and* a Republican is a new event, which we write as “ S and R .” The event that she is a smoker *or* a Republican is another new event, written as “ S or R .” It is important to know that we by “ S or R ” mean “smoker or Republican *or both*.” This definition of “or” is typical in mathematics, logic, and computer science. In daily language, it is often emphasized by using the expression “and/or” to distinguish from what math people call the *exclusive or*, which only permits one of the two, like in the phrase “You want fries *or* onion rings with that?”

The event that the selected individual is not a Republican is simply written as “not R .” The event that she is neither a Republican nor a smoker can be expressed in two different ways. One way is to negate that she is either, which gives “not (R or S).” The other way is to negate each separately and put them together: “(not R) and (not S).” We have argued for the following equality between events:

$$\text{not } (R \text{ or } S) = (\text{not } R) \text{ and } (\text{not } S)$$

10 COMPUTING PROBABILITIES: RIGHT WAYS AND WRONG WAYS

The parentheses are there to make it clear to what “not” refers. In a similar way,

$$\text{not (R and S)} = (\text{not R}) \text{ or } (\text{not S})$$

Make sure that you understand these little exercises in logic; we will make use of them later.

THE PROBABILIST’S RULE BOOK

Probabilities can be expressed as fractions, decimal numbers, or percentages. If you toss a coin, the probability to get heads is $1/2$, which is the same as 0.5, which is the same as 50%. There are no rules for when to use which notation, and you will see examples of all three in this book. In daily language, proper fractions are often used and often expressed, for example, as “one in ten” instead of $1/10$ (“one-tenth”). This is also natural when you deal with equally likely outcomes. Decimal numbers are more common in technical and scientific reporting when probabilities are calculated from data. Percentages are also common in daily language and often with “chance” replacing “probability.” Meteorologists, for example, typically say things like “there is a 20% chance of rain.” The phrase “the probability of rain is 0.2” means the same thing. When we deal with probabilities from a theoretical viewpoint, we always think of them as numbers between 0 and 1, not as percentages.

Regardless of how probabilities are expressed, they must follow certain rules. One such rule that is easy to understand is that a probability can never be a negative number. The lowest possible probability is 0, meaning that we are dealing with something that just does not happen. There is no point in trying to emphasize this further by letting the probability be -0.3 or -5 .² A related rule is that a probability can never be more than 1 (or 100%). If the probability is 1 (or 100%), we are describing something that we are absolutely certain about. Of course you can still say that you

²I do not know how familiar you are with negative numbers, but to mathematicians they are as natural as air and water. Here is the world’s funniest math joke: A biologist, a physicist, and a mathematician are sitting at a sidewalk cafe watching a house across the street. After a while two people enter the house. A little later, three people exit. “Reproduction,” says the biologist. “Measurement error,” says the physicist. “Hmm,” says the mathematician, “if a person enters the house it will be empty again.”

are 200% certain that the Texas Rangers will win the World Series, but nobody outside Dallas will take you seriously.

The next rule is that the probability that something does *not* occur can be computed as one minus the probability that it does occur. In a formula,

$$P(\text{not } A) = 1 - P(A)$$

Also easy to accept. The probability not to get 6 when you roll a die is $5/6$, which is also equal to $1 - 1/6$. If the chance of rain is 20%, then the chance that it does not rain is 80%. In all its simplicity, this rule turns out to be surprisingly useful. In fact, in his excellent book *Taking Chances: Winning with Probability*, British probabilist John Haigh names it probability's Trick Number One.

In the world of gambling, probabilities are often expressed by *odds*. To say that the odds are 4:1 against the event A means that it is four times as likely that A does not occur than that it occurs. We get the equation $P(\text{not } A) = 4 \times P(A)$, which has the solution $P(A) = 1/5$ and $P(\text{not } A) = 4/5$. As bookmakers are in the business to make a living, offering odds of 4:1 in reality means that they think that the probability of A is *less* than $1/5$.

Another rule. Let A and B be events such that whenever A occurs, B must also occur. Then $P(A)$ is less than (or equal to) $P(B)$, and the mathematical notation for this is $P(A) \leq P(B)$. For an example, let A be the event to roll a 6 and B the event to roll an even number. Whenever A occurs, B must also occur. However, B can occur without A occurring if you roll 2 or 4. In particular, the composition of two events is always less probable than each individual event. What I mean is that $P(A \text{ and } B)$ is always less than both $P(A)$ and $P(B)$, regardless of what A and B are.

As an example of the rule from the last paragraph, let us consider Mrs. Boudreaux and Mrs. Thibodeaux who are chatting over their fence when the new neighbor walks by. He is a man in his sixties with shabby clothes and a distinct smell of cheap whiskey. Mrs. B, who has seen him before, tells Mrs. T that he is a former Louisiana state senator. Mrs. T finds this very hard to believe. "Yes," says Mrs. B, "he is a former state senator who got into a scandal long ago, had to resign, and started drinking." "Oh," says Mrs. T, "that sounds more likely." "No," says Mrs. B, "I think you mean less likely."

Strictly speaking, Mrs. B is right. Consider the following two statements about the shabby man: "He is a former state senator" and "He is

a former state senator who got into a scandal long ago, had to resign, and started drinking.” It is tempting to think that the second is more likely because it gives a more exhaustive explanation of the situation at hand. However, this reason is precisely why it is a *less* likely statement. Note that whenever somebody satisfies the second description, he must also satisfy the first but not vice versa. Thus, the second statement has a lower probability (from Mrs. T’s subjective point of view; Mrs. B of course knows who the man is). This example is a variant of examples presented in the book *Judgment under Uncertainty* by Economics Nobel laureate³ Daniel Kahneman and coauthors Paul Slovic and Amos Tversky. They show empirically how people often make similar mistakes when they are asked to choose the most probable among a set of statements. It certainly helps to know the rules of probability. A more discomfoting aspect is that the more you explain something in detail, the more likely you are to be wrong. If you want to be credible, be vague.

The final rule is the *addition rule*. It says that in order to get the probability that either of two events occur, you add the probabilities of the two individual events. This rule, however, only applies if the two events in question *cannot occur at the same time* (the technical term for such events is that they are *mutually exclusive*). In a formula:

$$P(A \text{ or } B) = P(A) + P(B)$$

For example, roll a die and consider the events A: to get 6 and B: to get an odd number. These events qualify as mutually exclusive because you cannot get both 6 and an odd number in the same roll. It is “same roll” that is important here; of course you can get 6 in one roll and an odd number in the next. By the formula above, the probability to get 6 or an odd number in the same roll is $1/6 + 3/6 = 4/6$.

In his bestseller *Innumeracy*, John Allen Paulos tells the story of how he once heard a local weatherman claim that there was a 50% chance of rain on Saturday and a 50% chance of rain on Sunday and thus a 100% chance of rain during the weekend. Clearly absurd, but what is the error? Faulty use of the addition rule! As a rainy Saturday does not exclude a rainy

³But I want to point out that the Economics prize is not a “true” Nobel prize in the sense that it was not mentioned in Alfred Nobel’s will. The prize was first awarded in 1969, and its official name is “The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel.” Just so that you know.

Sunday, we here have two events that can both occur the same weekend. In cases like this one, there is a modified version of the addition rule that says that you first add the two probabilities as before and then subtract the probability that *both* events occur. In a formula, it looks as follows:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Note that if A and B cannot occur at the same time, then $P(A \text{ and } B) = 0$ and we have the first addition rule as a special case. If we let A denote the event that it rains on Saturday and B the event that it rains on Sunday, the event “A and B” describes the case in which it rains both days. To get the probability of rain over the weekend, we now add 50% and 50%, which gives 100%, but we must then subtract the probability that it rains both days. Whatever this is, it is certainly more than 0, so we end up with something less than 100%, just like common sense tells us that we should. I just wonder what the weatherman would have said if the chances of rain had been 75% each day.

Let us also check the formula in a dice example. If you roll two dice, what is the probability to get at least one 4? Here, the relevant events are A: 4 on the first die and B: 4 on the second die. The event to get at least one 4 is then the event “A or B,” and in Figure 1.4, you can check directly that

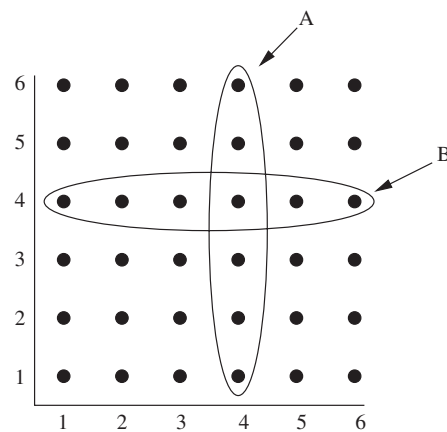


Figure 1.4 The sample space of 36 equally likely outcomes for rolling two dice. The events “4 on first die” and “4 on second die” are marked, and you may note that there are 6 outcomes in each event, 11 outcomes that are in at least one event, and 1 outcome that is in both.

14 COMPUTING PROBABILITIES: RIGHT WAYS AND WRONG WAYS

this equals $11/36$. Also, $P(A) = 6/36$, $P(B) = 6/36$, and $P(A \text{ and } B) = 1/36$ because there is only one outcome that gives 4 on both dice. As $6 + 6 - 1 = 11$, the formula is valid.

Whenever probabilities are assigned, this must be done in a way such that none of the rules are violated. Ask a friend how likely he thinks it is that it will rain Saturday, Sunday, both days, and at least one of the days, respectively. You will then get four probabilities that must satisfy the rules that we have discussed above. For example, somebody may think that rain on Saturday is pretty likely, say 70%, and the same for Sunday. Rain both days? Well, maybe 50%. For the last probability, let's say 80%. But this assignment of probabilities violates the addition rule because 80 is not equal to $70 + 70 - 50 = 90$. Somebody else might come up with the following probabilities (same order): 70%, 60%, 80%, and 50%. These do satisfy the addition rule but suffer from another problem. Can you tell which? (Hint: Mrs. Boudreaux could.)

Let us keep thinking about weekend weather. Suppose that both Saturday and Sunday each have probability 0.5 to get rain and that the probability is p that it rains both days (we now think of probabilities as numbers between 0 and 1, not percentages). What is the range of possible values of p ? How does the probability of rain during the weekend depend on p ?

If we let A and B be the events "rain on Saturday" and "rain on Sunday" respectively, then a rainy weekend is the event " A or B ," and because $p = P(A \text{ and } B)$, we get the equation

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = 1 - p$$

As p must be less than both $P(A)$ and $P(B)$, it cannot be more than 0.5. If p is 0, then $P(A \text{ or } B) = 1$ and the rainy weekend is a fact. As p ranges from 0 to 0.5, the probability of a rainy weekend decreases from 1 to 0.5. Why? It has to do with how likely rainy Saturdays and Sundays are to come in pairs. Think of a year, which has 52 weekends. On average, we expect to get rain 26 Saturdays and 26 Sundays. If p is 0, this means that if it rains on a Saturday, it *never* rains on the following Sunday and if it does not rain on Saturday, it *always* rains on Sunday. Thus, the 26 rainy Saturdays and 26 rainy Sundays must be spread over the year so that they never come in pairs. The only way to do this is to let every weekend have exactly one rainy day. As p gets bigger, rainy days are more likely to come in pairs, and the extreme case is when $p = 0.5$. Then *all* rainy days come in pairs and the year has half of its weekends rainy and the other half dry.

Here is an exercise for you. Change the probabilities a little, and let $P(A) = 0.6$ and $P(B) = 0.7$, and let p again denote $P(A \text{ and } B)$. Explain why p must be between 0.3 and 0.6.

INDEPENDENCE, AIRPLANES, AND RUSSIAN PEASANTS

Plenty of random things happen in the world all the time, most of which have nothing to do with one another. If you toss a coin and I roll a die, the probability that you get heads is $1/2$ regardless of the outcome of my die. If there is a 20% chance of rain tomorrow, this does not change if a flu outbreak in Asia is reported. Changes in the U.S. stock market indexes have nothing to do with who wins the Wimbledon tennis tournament. Events that in this way are unrelated to each other are called *independent*. It is easy to compute the probability that two independent events both occur; simply multiply the probabilities of the two events. We call this computation the *multiplication rule* for probabilities, described in a formula as

$$P(A \text{ and } B) = P(A) \times P(B)$$

It works in two directions. If we can argue that two events are independent, then we can use the multiplication rule to compute the probability that both occur at the same time. Conversely, if we can show that the multiplication rule holds, then we can conclude that the events are independent. It can be argued at some length why this is true and we will just look at some simple examples to convince ourselves that formula and intuition agree. Let us do the first example above, that you toss a coin and I roll a die. There are 12 equally likely outcomes: (H,1), ..., (H,6), (T,1), ..., (T,6) in the obvious notation. What is now the probability that you toss heads and I roll a 6? Obviously $1/12$. The individual probabilities of heads and 6 are $1/2$ and $1/6$, respectively, and $1/2 \times 1/6$ equals $1/12$ indeed.

For another example, take a deck of cards, draw one card, and consider the two events, A: to get an ace, and H: to get hearts. Are these independent? Let us check whether the multiplication rule holds. The individual probabilities are

$$P(A) = 4/52 = 1/13$$

$$P(H) = 13/52 = 1/4$$

and the probability to get both A and H is the probability to get the ace of hearts, which is $1/52$, which is the product of $1/13$ and $1/4$. We have

$$P(A \text{ and } H) = P(A) \times P(H)$$

which means that A and H are independent. Now remove the two of spades from the deck, reshuffle, and consider the same two events as above. Are they still independent? They must be, right? After all, the two of spades has nothing to do with either aces or hearts. Let us compute the probabilities. There are now 51 cards, and we get

$$P(A) = 4/51$$

$$P(H) = 13/51$$

and $P(A \text{ and } H) = 1/51$. As $P(A \text{ and } H)$ is not equal to $P(A) \times P(H)$, we must conclude that the events are not independent anymore. What happened? Removing the two of spades changes the proportions of aces in the deck from $4/52$ to $4/51$, but not *within the suit of hearts* where it remains at $1/13 = 4/52$. Here is how you should think about independent events: *If one event has occurred, the probability of the other does not change.* In the card example, the probability of A is $4/51$ but changes to $1/13$ if the event H occurs.

Here is a question I often ask my students after I have introduced independence: If two events cannot occur at the same time, are they independent? At first you might think so. After all, they have nothing to do with each other, right? Wrong! They have a lot to do with each other. If one has occurred, we know for certain that the other cannot occur. The probability to roll a 6 is $1/6$, but if I tell you that the outcome is an odd number, the probability of a 6 drops down to 0. Think this through. It is important to understand independence.

There is a story that is sometimes told about the great Russian mathematician Andrey Nikolaevich Kolmogorov, among many other things the founder of the modern theory of probability. In Stalin's Soviet Union in the 1930s, the concept of independence did not fit well with the historical determinism of Marxist ideology. When questioned by a panel of ideologues about this possible heresy, Kolmogorov countered, "If the peasants pray for rain and it actually starts to rain, were their prayers answered?" The atheist ideologues had to confess that this must indeed be a case of

independent events and Kolmogorov lived a long and productive life until his death in 1987 at the age of 84.

In December 1992, a small passenger airplane crashed in a residential neighborhood near Bromma airport outside Stockholm in Sweden, causing no death or injury to any of the residents. Already disturbed by increasing traffic and expansion plans for the airport, the residents now got more reasons to worry. In an effort to calm people, the airport manager said in an interview on TV that statistically people should now feel safer because the probability to have another accident had become so much smaller than before. I was at the time a graduate student in Sweden, studying probability and statistics, and thought that it was amusing to hear both “statistically” and “probability” used in the same sentence in such a careless way. In youthful vigor, I immediately wrote a letter that was published in some leading Swedish newspapers, where I explained why the airport manager’s statement was incorrect. I also encouraged him to contact me so that I could recommend a good probability textbook. I never heard from him.

The airport manager’s error is common: He confuses the probability that something happens *twice* and the probability that something happens *again*. Toss a coin twice. What is the probability to get heads twice? One-fourth. Toss a coin until you get heads. What is the probability that you get heads again in the next toss? One-half, by independence. Replace the coin tosses with flights to and from Bromma Airport and the probability of tossing heads with the probability of having a crash, and you got him. His only possible defense would be that crashes are not independent, and that after such a crash, an investigation is started that may improve security. Perhaps. But first of all, that was not his argument. He believed that there was magic in the sheer probabilities. Second, even if there was such an investigation, it would not be likely to dramatically reduce the probability of another crash, which can occur for many different reasons. The events are not independent, but almost. Compare with the example above where the events “ace” and “hearts” are not independent when a card is drawn from a deck without the two of spades. The probability to get an ace is $4/51$, which is roughly 0.078, and the probability to get an ace if we know that the card is hearts is $1/13$, or roughly 0.077, not much different. The events are almost independent.

In a probability class, I once pointed out that even if you have just tossed nine heads in a row, the next toss is still equally likely to give

another head as it is to give tails. A student approached me after class and wondered how this could be possible. After all, aren't sequences of 10 consecutive heads pretty rare? The first reply is that a coin has no memory. When you start tossing a coin, would you need to know whether the coin has been tossed before and what it gave? Of course not. The student had no problem accepting this assertion but still insisted that if he was to toss a coin repeatedly, sequences of 10 consecutive heads would be very rare, which would contradict my claim. Although he is right that a sequence of 10 consecutive heads is pretty rare (it has probability $1/1,024$, less than one thousandth), this is irrelevant because I was talking about the probability to get heads once more after we had already gotten nine in a row. If he tossed his coin repeatedly in sequences of 10, he would start with nine consecutive heads about once every 512 times and about half of these would finish with yet another head in the 10th toss. Probability of 10 consecutive heads: $1/1,024$, probability of heads once more after nine consecutive heads to start with: $1/2$. Airport managers and college students are not alone. These types of mistakes are very common, and I will address them in more depth and detail in later chapters.

Suppose now that you have agreed to settle a dispute with cousin Joe by tossing a coin. The problem is that neither of you has any change. Joe suggests that you instead toss a bottle cap, which will count as heads if it lands with the top up, and tails otherwise. As you cannot assume that these are equally likely, is there any way in which fairness can be guaranteed?

You can suggest a trick invented by computer pioneer John von Neumann. Instead of tossing the cap once and observing heads or tails, the cap is tossed twice. If this gives the sequence HT, you win; if it gives TH, Joe wins. If it gives HH or TT, nobody wins and you start over. Suppose the probability of heads is some value p , not necessarily $1/2$. As the probability of tails is then $1 - p$, independence gives that the probability to get HT is $p \times (1 - p)$ and the probability to get TH is $(1 - p) \times p$, which is the same. The procedure is fair (but may take a while if p is very close to 0 or 1).

For independence of more than two events, the multiplication rule still applies. If A, B, and C are independent, then $P(A \text{ and } B) = P(A) \times P(B)$, and similar for the combinations A–C and B–C. Also, the probability that all three events occur is $P(A \text{ and } B \text{ and } C) = P(A) \times P(B) \times P(C)$. Things are a bit more complicated with three events. It is not enough that the events are independent two by two as the following example shows.

I will let you do it on your own. You toss a coin twice and consider the three events

- A: heads in first toss
- B: heads in second toss
- C: different in first and second toss

Show that the events are independent two by two but that C is not independent of the event “A and B” and that the multiplication rule fails for all three events. Note that A alone does not give any information about C, and neither does B alone. However, A and B in combination tells us that C cannot occur.

If you want to compute the probability that at least one of several independent events occur, Trick Number One from page 11 133 comes in handy. First compute the probability that *none* of the events occurs, and then subtract this probability from 1. For example, in the carnival game *chuck-a-luck* you roll three dice and win a prize if you get at least one 6. What is the probability that you win? The probability to roll 6 with one die is $1/6$, and as you have three attempts, you might think that you have a 50–50 chance. It is certainly true that three times $1/6$ equals $1/2$, but this is irrelevant to the problem. If you follow the advice I just gave, first compute the probability that none of the dice gives 6. By independence, this probability is

$$P(\text{no 6s}) = 5/6 \times 5/6 \times 5/6 = (5/6)^3$$

and we get

$$P(\text{at least one 6}) = 1 - (5/6)^3 \approx 0.42$$

and, as always in games that somebody wants you to pay money to play, you are more likely to lose than to win. What if there are instead four dice? Your chance to win is then $1 - (5/6)^4$, which is approximately 0.52 so with four dice you would have an edge.

Another example. An American roulette table has the numbers 1–36, plus 0 and 00. Thus, if you bet on a single number, your chance to win is $1/38$. How many rounds do you have to play if you want to have a

20 COMPUTING PROBABILITIES: RIGHT WAYS AND WRONG WAYS

50–50 chance to win at least once? Perhaps 19 rounds (half of 38)? Call the number n . By the same argument as above, we get the equation

$$P(\text{win at least once}) = 1 - (37/38)^n$$

For $n = 19$, this is only about 0.4. For $n = 25$, it is approximately equal to 0.49, and for $n = 26$, it is just above 0.5. You need to play 26 rounds. That 38 divided by 2 equals 19 is another example of something that is true but irrelevant.⁴ The number 19 arises in a different way though; If you instead bet on 19 different numbers in one round, you have a 50–50 chance to win. Of course, you can then only win once, whereas with successive bets on the same number, you can win many times. As we shall see later, in the long run, you lose just as much regardless of how you play. Unfortunately.

CONDITIONAL PROBABILITY, SWEDISH TV, AND BRITISH COURTS

If two events are not independent, they are called . . . get ready now . . . dependent. If two events are dependent, the probability of one changes with the knowledge of whether the other has occurred. The probability to roll a 6 is $1/6$. If I tell you that the outcome is an even number, you can rule out the outcomes 1, 3, and 5, and the probability to get 6 changes to $1/3$. We call this the *conditional probability* of getting 6 *given* that the outcome is even. I have mentioned that you can think of probabilities in terms of average long-term behavior. The same is true for conditional probabilities; you just ignore all outcomes that do not satisfy the condition. In the dice example I just gave, you would thus disregard all odd outcomes and count the proportion of 6s among the even outcomes, and this should stay close to one-third after a while. There is a multiplication rule that can be stated in terms of conditional probabilities. For any events A and B, the following is always true:

$$P(A \text{ and } B) = P(A) \times P(B \text{ given } A)$$

⁴Which reminds me of the marginally funny story about a man in a balloon who is lost and asks a man on the ground where he is. The man replies, “You are in a balloon.” “Just my luck,” says the balloonist, “asking a mathematician.” “How did you know I’m a mathematician?” asks the man, and the balloonist replies, “Your answer was correct but useless!”

In other words, to find the probability that both A and B occur, first find the probability of A, and then the conditional probability of B given that A occurred. In applications of the formula, it is up to you which event you want to call A and which to call B. Suppose that you draw two cards from a deck. What is the probability that both are aces? The probability that the first card is an ace is $4/52$, which is our $P(A)$. Given that the first card is an ace, there are now three aces left among the remaining 51 cards. The conditional probability of another ace is thus $3/51$, our $P(B \text{ given } A)$. Multiply the two to get the probability of two aces as $4/52 \times 3/51$, which is approximately 0.0045.

If you compare the two versions of the multiplication rule, you realize that independent events have the special property that $P(B) = P(B \text{ given } A)$; the unconditional and the conditional probabilities are the same. This observation makes sense. In the last example, suppose that instead of drawing two cards, you draw one card, put it back, reshuffle the deck, and draw again. Now what is the probability that you get two aces? In this case, the events to get an ace in the first and second draws *are* independent and the probability is $4/52 \times 4/52$, which is about 0.0059. (Why is it larger than before? Think about what happens if you draw three, four, or five cards and ask for three, four, or five aces.)

Note that there is a certain symmetry here. If $P(B \text{ given } A)$ is different from $P(B)$, then $P(A \text{ given } B)$ is also different from $P(A)$. You may try to prove this rule from the multiplication formula above, noting that “A and B” is the same as “B and A.” Also note that the multiplication rule gives a way to compute the conditional probability if it is not obvious how to do so directly. Shuffling around the factors in the formula above gives the expression

$$P(B \text{ given } A) = \frac{P(A \text{ and } B)}{P(A)}$$

which will be useful to us later. Note the difference between $P(A \text{ and } B)$, which is the probability that *both* A and B occur, and $P(B \text{ given } A)$, which is the probability that B occurs *if we know that A has occurred*. These can be quite different. For example, choose an American at random. Let A be the event that you get somebody from Rhode Island and B the event that you get somebody of Portuguese descent (a *Luso-American*). Then, $P(A \text{ and } B)$ is the probability to get a Rhode Islander with Portuguese ancestry, which is about 0.03% (as of April 2014 there are about 100,000

22 COMPUTING PROBABILITIES: RIGHT WAYS AND WRONG WAYS

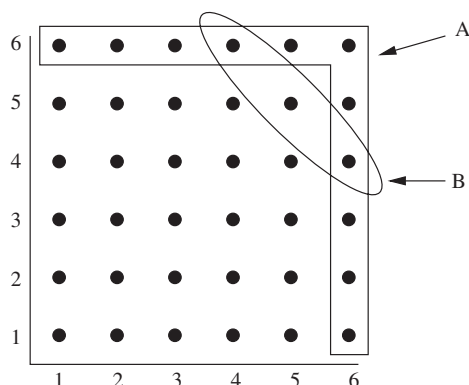


Figure 1.5 Rolling two dice. The events marked are A: to get at least one 6, and B: to get a sum equal to 10.

such individuals among the U.S. population of 316 million). The conditional probability $P(B \text{ given } A)$, on the other hand, is the probability that a Rhode Islander has Portuguese ancestry, and this is about 9.5% (100,000 out of about 1.05 million).

For a simple illustration of how to compute conditional probabilities, let us again turn to the experiment of rolling two dice. Let A be the event to get at least one 6 and let B be the event that the sum of the dice equals 10. See Figure 1.5 for the sample space with these two events. Two outcomes satisfy both A and B (“4 on first, 6 on second” and “6 on first, 4 on second”), and we thus have $P(A \text{ and } B) = 2/36$. In the figure you also see that $P(A) = 11/36$, and by the formula above, the conditional probability that the sum is 10 given that at least one die shows 6 is

$$P(B \text{ given } A) = \frac{P(A \text{ and } B)}{P(A)} = \frac{2/36}{11/36} = 2/11$$

and you can understand this intuitively: If you know that there is at least one 6, there are 11 possible outcomes, and because two of these have the sum equal to 10, the conditional probability to get sum 10 is $2/11$. It is nice to see that the formal computation and the intuitive reasoning agree. Provided that your intuition does not go agley, they always do.

In the early 1990s, a leading Swedish tabloid tried to create an uproar with the headline “Your ticket is thrown away!.” This was in reference to the popular Swedish TV show “Bingolotto” where people bought lottery

tickets and mailed them to the show. The host then, in live broadcast, drew one ticket from a large mailbag and announced a winner. Some observant reporter noticed that the bag contained only a small fraction of the hundreds of thousands tickets that were mailed. Thus the conclusion: Your ticket has most likely been thrown away!

Let us solve this quickly. Just to have some numbers, let us say that there are a total of 100,000 tickets and that 1,000 of them are chosen at random to be in the final drawing. If the drawing was from all tickets, your chance to win would be $1/100,000$. The way it is actually done, you need to both survive the first drawing to get your ticket into the bag and then get your ticket drawn from the bag. The probability to get your entry into the bag is $1,000/100,000$. The conditional probability to be drawn from the bag, given that your entry is in it, is $1/1,000$. Multiply to get $1/100,000$ once more. There were no riots in the streets.

Conditional probability can also explain why Mrs. T from page 11 made her statement “That sounds more likely.” She thought of a conditional probability without even knowing it. It was hard for her to believe that a former senator could be so shabby, but when she found out more about him, she found it easier to believe. Thus, in her mind, $P(B \text{ given } A)$ was larger than $P(B)$ (what are A and B?).

Misunderstanding probability can be more serious than upsetting Swedish TV viewers or making fun of Louisiana politicians. One famous case is that of Sally Clark. In 1999, a British jury convicted her of murdering two of her children who had died suddenly at the ages of 11 and 8 weeks, respectively. A famous pediatrician, Roy Meadow, called in as an expert witness claimed that the chance of having two cases of infant sudden death syndrome, or “cot deaths,” in the same family was 1 in 73 million. There was no physical or other evidence of murder, nor was there a motive. Most likely, the jury was so impressed with the seemingly astronomical odds against the incidents that they convicted. But where did the number come from? Data suggested that a baby born in a family similar to the Clarks faced a 1 in 8,500 chance of dying a cot death. Two cot deaths in the same family, it was argued, therefore had a probability of $1/8,500 \times 1/8,500$, which is roughly equal to $1/73,000,000$.

Did you spot the possible error? I hope you did. The computation assumes that successive cot deaths in the same family are independent events. This assumption is clearly questionable, and even a person without any medical expertise might suspect that genetic or environmental

factors play a role. Indeed, there is some evidence to this effect as it has been suggested that certain groups suffer higher rates of sudden infant death syndrome (SIDS), for example, males and African-Americans. The Royal Statistical Society issued the following statement questioning the independence assumption:

Not only was no such empirical justification provided in the case, but there are very strong a priori reasons for supposing that the assumption will be false. There may well be unknown genetic or environmental factors that predispose families to SIDS, so that a second case within the family becomes more likely.

If there is one cot death, what is the risk for the next child to face the same fate? Data are scarce and there is much uncertainty. There are studies that do not show an increased risk, and there are those that do. One such study puts the risk at a 10-fold increase, thus going from $1/8,500$ to $1/850$. For the sake of argument, let us assume that the probability of a second case of SIDS is as high as $1/100$. This is a conditional probability and the probability of having two cot deaths in the same family equals $1/8,500 \times 1/100$, which equals $1/850,000$. Now, this is still a small number, close to one-in-a-million, and might not have made the jurors judge differently. But what does the probability $1/850,000$ have to do with Sally's guilt? Is this the probability that she is innocent? Not at all. It is the probability that, under the assumption of innocence, two of her children die. Thus, the innocence is what we condition upon, not of what we compute the probability. There is a fundamental difference between $P(\text{evidence given innocence})$ and $P(\text{innocence given evidence})$ and confusion of the two is known as the *prosecutor's fallacy*. And even though the event of two cot deaths has a very low probability, so does the event of double infanticide, so the question becomes how large these probabilities are relative to each other. We will return to such calculations in Chapter 4. Sally spent 3 years in prison before her verdict was overturned, Roy Meadow was found guilty of "serious professional misconduct," and Sally Clark died in 2007 at the age of 42. A tragic case indeed.

Next, let us look at a paradox that is not usually presented as a probability problem. Your teacher tells the class there will be a surprise exam next week. On one day, Monday–Friday, you will be told in the morning that an exam is to be given on that day. You quickly realize that the exam will not be given on Friday; if it was, it would not be a surprise because it is the last possible day to get the exam. Thus, Friday is ruled

out, which leaves Monday–Thursday. But then Thursday is impossible also, now having become the last possible day to get the exam. Thursday is ruled out, but then Wednesday becomes impossible, then Tuesday, then Monday, and you conclude: There is no such thing as a surprise exam! But the teacher decides to give the exam on Tuesday, and come Tuesday morning, you are surprised indeed.

This problem, which is often also formulated in terms of surprise fire drills or surprise executions, is known by many names, for example, the “hangman’s paradox” or by serious philosophers as the “prediction paradox.” To resolve it, I find it helpful to treat it as a probability problem. Let us suppose that the day of the exam is chosen randomly among the 5 days of the week. Now start a new school week. What is the probability that you get the test on Monday? Obviously $1/5$ because this is the probability that Monday is chosen. If the test was not given on Monday, what is the probability that it is given on Tuesday? The probability that Tuesday is chosen to start with is $1/5$, but we are now asking for the *conditional* probability that the test is given on Tuesday, given that it was not given on Monday. As there are now 4 days left, this conditional probability is $1/4$. Similarly, the conditional probabilities that the test is given on Wednesday, Thursday, and Friday conditioned on that it has not been given thus far are $1/3$, $1/2$, and 1, respectively.

We could define the “surprise index” each day as the probability that the test is *not* given. On Monday, the surprise index is therefore 0.8, on Tuesday it has gone down to 0.75, and it continues to go down as the week proceeds with no test given. On Friday, the surprise index is 0, indicating absolute certainty that the test will be given that day. Thus, it is possible to give a surprise test but not in a way so that you are equally surprised each day, and it is never possible to give it so that you are surprised on Friday.

LIAR, LIAR

This entire section is devoted to a classic probability problem. It is easy to state but can lead to great confusion and frustration. It may strike you as a wee bit tedious, and if you feel that it fails to catch your interest, you can safely skip this section and proceed to the next without missing any vital information.

The problem is a typical example of how you sometimes need to stop and think about what you are asked to do before you do anything. It goes

like this: Adam, Bob, and Carol are each known to tell the truth with probability $1/3$ (independently of each other) and lie otherwise. If Adam denies that Bob confirms that Carol lies, what is the probability that Carol tells the truth?

Yikes. Let us first realize that we are here in fact asked for a conditional probability and name the two events of interest:

C: Carol tells the truth

A: Adam denies that Bob confirms that Carol lies

We are now asking for the conditional probability $P(C \text{ given } A)$. From page 21, we know that this can be computed as $P(C \text{ and } A)$ divided by $P(A)$, so let us first find $P(C \text{ and } A)$. Let us get rid of a double negation and rephrase A as

Adam says that Bob says that Carol tells the truth

which means that the combined event “C and A” can be written as

Carol tells the truth and Adam says that Bob says that Carol tells the truth

The question is now: For which combinations of lying and truth-telling among the three will this last event occur? First of all, Carol must tell the truth. What about the others? If Adam tells the truth when he confirms that Bob confirms Carol’s truth-telling, then Bob is also telling the truth. Thus, the combined event occurs if everybody tells the truth, and this has probability $1/3 \times 1/3 \times 1/3 = 1/27$.

What if Adam lies? Then Bob says that Carol lies, so he is also lying and the combination “Adam lies, Bob lies, Carol tells the truth” also makes the combined event occur. That combination has probability $2/3 \times 2/3 \times 1/3 = 4/27$. Adding this number to the $1/27$ from above and noting that no other combinations work, we conclude that

$$P(C \text{ and } A) = 5/27$$

For the event A alone, the two combinations above make it occur but there are more possibilities. For example, if Adam tells the truth and both

Table 1.1 Possible combinations of lying and truth-telling for the three individuals

Adam	T	T	T	T	L	L	L	L
Bob	T	T	L	L	T	T	L	L
Carol	T	L	T	L	T	L	T	L
A	yes	no	no	yes	no	yes	yes	no
C and A	yes	no	no	no	no	no	yes	no
Probability	1/27	2/27	2/27	4/27	2/27	4/27	4/27	8/27

Bob and Carol lies, the event A occurs. Why? Well, suppose that Adam truthfully confirms that Bob says: “Carol tells the truth.” If Bob lies, this means that Carol also lies and the event “Adam says that Bob says that Carol tells the truth” occurs. Table 1.1 gives all possible combinations of truth-telling (T) and lying (L), and whether the events occur. Note that for the event A to occur, we need an odd number of truth-tellers, and for the combined event “C and A” to occur, we in addition need Carol to be one of them.

We computed $P(C \text{ and } A)$ above, and by adding the probabilities of the “yes” entries in the table, we also get $P(A) = 13/27$. We can now compute the desired conditional probability as

$$P(C \text{ given } A) = \frac{P(C \text{ and } A)}{P(A)} = \frac{5/27}{13/27} = 5/13$$

Note that this probability is about 38.5%, slightly higher than the 33.3% that is the unconditional probability that Carol tells the truth. The fact that Adam says that Bob confirms Carol’s truth-telling makes us believe in her a little more, which might be a bit surprising because these guys are such a bunch of liars.

This problem is an old one. It was published by British astrophysicist Sir Arthur Eddington in his 1935 book *New Pathways in Science* and further explained by him in a 1935 article in the *The Mathematical Gazette*. He claims in turn to have learned about it in a 1919 after-dinner speech by his colleague A. C. D. Crommelin (who has a comet named after him). In Sir Arthur’s version, which I will state shortly, there is a fourth person also involved. Interestingly, the problem led to some controversy and different solutions were published. This discrepancy has to do with one

crucial assumption that I made above but did not explicitly state: If Adam lies when he says “Bob says that Carol tells the truth,” I interpreted this as meaning that Bob says that Carol lies, but it could also mean that Bob did not say anything at all. In fact, the whole solution rests on the following chain of assumptions that are not spelled out in the problem: First Carol says something that is either true or false. Next, Bob who knows whether she told the truth, says either “Carol tells the truth” or “Carol lies.” Finally, Adam says either “Bob says that Carol tells the truth” or “Bob says that Carol lies.” This interpretation also validates my rewriting to “get rid of a double negation.”

However, if we return to the original formulation “Adam denies that Bob confirms that Carol lies,” it might also be argued that Adam is asked the question, “Does Bob confirm that Carol lies?” and answers, “No.” If Adam lies, it means that Bob does indeed say, “Carol lies.” However, if Adam speaks the truth, this could mean that Bob denies that Carol lies, but it could also mean that Bob has not said anything at all in the matter. The latter was Sir Arthur’s interpretation of the problem. His only assumption was that Carol makes a statement that is either true or false, which led him to exclude only cases that are clearly inconsistent with the statement in the problem formulation (his interest in the problem in the first place was as an illustration of what he called the “exclusion method” in interpreting observational results in physics). In his view, the only cases inconsistent with the statement A are the cases $L-T-T$ and $L-L-L$ in the notation of the table above. The cases $T-T-L$ and $T-L-T$, which we excluded from A are included by Sir Arthur. He views all cases in which Adam tells the truth as consistent with A ; if Adam tells the truth, Sir Arthur argues, we simply cannot say anything about what Bob has said and there is no evidence against Carol speaking the truth. In his interpretation, $P(C \text{ and } A) = 7/27$ and $P(A) = 17/27$, which gives the final answer that Carol tells the truth with probability $7/17$.

In the December 1936 issue of the *Gazette*, two articles were published: one that agreed with Sir Arthur and one that disagreed. Of course there is no universally correct answer, only a correct answer relative to the assumptions that are made. With Sir Arthur’s interpretation, one must assume that Carol tells the truth when it cannot be proved that she lies. The interpretation depends also on what context we imagine. If these people are testifying in a court trial, it is reasonable to assume that they have all made statements and our interpretation is logical. If it is instead

an illustration of some principle in physics, Sir Arthur's interpretation may perhaps make sense, but I believe that most probabilists would agree that the only way to properly solve the problem is to make the assumptions that we have made. Still, I would not go as far as Warren Weaver, who in his modern classic *Lady Luck: The Theory of Probability* claims that Sir Arthur's solution involves a condition that is "rather ridiculous in character" (*Lady Luck* that came out in 1963 is by the way a wonderful book, to this day arguably the best nontechnical introduction to probability that there is). Of course, the complete set of assumptions in the liar problem should not be spelled out in the problem formulation. That would clutter the problem and reduce the chance of interesting conflicts between feisty British astrophysicists. Here is Sir Arthur's original formulation:

If A, B, C, and D each speak the truth once in three times (independently), and A affirms that B denies that C declares that D is a liar, what is the probability that D was speaking the truth?

I will leave it as an exercise for you to solve it. Sir Arthur gave the answer 25/71. With our interpretation, the correct answer is 13/41.

CAR DEALERS AND COLOR BLINDNESS

Suppose that you buy a used car in a city where street flooding is a common problem. You know that roughly 5% of all used cars have been flood damaged and estimate that 80% of such cars will later develop serious engine problems, whereas only 10% of used cars that are not flood damaged develop the same problems. Of course, no used car dealer worth his salt would let you know whether your car has been flood damaged, so you must resort to probability calculations. What is the probability that your car will later run into trouble?

You might think about this problem in terms of proportions. Out of every 1,000 cars sold, 50 are previously flood damaged, and of those, 80%, or 40 cars, develop problems. Among the 950 that are not flood damaged, we expect 10%, or 95 cars, to develop the same problems. Hence, we get a total of $40 + 95 = 135$ cars out of a thousand, and the probability of future problems is 13.5%.

If you solved the problem in this way, congratulations. You have just used the *law of total probability*, which is one of the most useful rules that

we have in probability. If we restate everything in terms of probabilities, we have done the calculation

$$P(\text{engine problems}) = 0.05 \times 0.80 + 0.95 \times 0.10 = 0.135$$

which means that we considered the two different cases “flood damaged” and “not flood damaged” separately and then combined the two to get our probability. For those of you who like math formulas, here is a general formula for two events, A and B:

$$P(B) = P(B \text{ given } A) \times P(A) + P(B \text{ given (not } A)) \times P(\text{not } A)$$

It is Sunday evening down at the local pub. You and your two colleagues Albert and Betsy have met for a pint and start discussing the bus you catch to work every morning, which tends to be late about 40% of the time. You decide to see who can best predict whether the bus will be late or on time for the entire next week. Each of you will suggest a sequence of five Ls and Ts (for “Late” and “on Time”). As you know that the bus is late with probability 0.4 each day, you decide to generate a sequence at random by choosing L with probability 0.4 and T with probability 0.6, five times. Your friend Albert thinks along the same lines but does not want to run the risk of getting too many Ls, so he decides on a sequence with two Ls and three Ts and chooses their positions randomly. Betsy notes that each day the bus is more likely than not to be on time and simply suggests the sequence TTTTT (which prompts Albert to shake his head and sigh “women” in his pint, “of course it will not be on time every day”). Who is most likely to guess the entire week correctly?

Let us compute the probability to guess correctly for a single day. If your guess is T, you are correct if the bus is on time, which has probability 0.6. Thus, Betsy has this probability every day, and Albert has it 3 days and 0.4 the other 2 days. You, with your more complicated strategy, are correct if you guess T and the bus is on time or if you guess L and the bus is late. The law of total probability gives that you are correct with probability $0.6 \times 0.6 + 0.4 \times 0.4 = 0.52$. With independence between the 5 days of the week, you have probability $0.52^5 \approx 0.038$ to be correct. Albert’s probability is $0.6^3 \times 0.4^2 \approx 0.035$ (this number is the same regardless of which 2 days he chooses for his two Ls) and Betsy’s probability is $0.6^5 \approx 0.078$. With probabilities 3.5%, 3.8%, and 7.8%, neither of you has a very good chance to get it right, but Betsy definitely has an edge. To somewhat save

his face, Albert can claim that he is more likely than Betsy to get the *number* of late days right. The probability that there are 2 late days is about 0.35 (which is his 0.035 multiplied by 10; can you see why the relevant probability is computed in this way?). His problem is, of course, to find the right 2 days.

Genetics is a science where probability theory is extremely useful. Recall that genes occur in pairs where one copy is inherited from the mother and one from the father. Suppose that a particular gene has two different *alleles* (variants) called A and a . An individual can then have either of the three *genotypes* AA , Aa , and aa . If the parents both have genotype Aa , what is the probability that their child gets the same genotype?

We assume that each of the two gene copies from each parent is equally likely to be passed on to the child and that genes from the father and the mother are inherited independently. There are then the four equally likely outcomes illustrated in Figure 1.6, and the probability that the child also has genotype Aa is $1/2$ (order has no meaning here, so Aa and aA are the same). Each of the genotypes AA and aa has probability $1/4$. The square in the figure is an example of a *Punnett square* where each cell has probability $1/4$. With one gene and two alleles we get a 2-by-2 square; with more genes or more alleles, we get larger squares.

An allele is said to be *recessive* if it is required to exist in two copies to be expressed and *dominant* if one copy is enough. For example, the hereditary disease *cystic fibrosis* (CF) is caused by a recessive allele of a particular gene. Let us denote this allele by C and the healthy allele by H so that only individuals with genotype CC gets the disease. Individuals with genotype CH are *carriers*, that is, they have the disease-causing allele but are healthy. It is estimated that approximately 1 in 25 individuals are carriers (among people of central and northern European descent; it is

		From mother:	
		A	a
From father:	A	AA	Aa
	a	aA	aa

Figure 1.6 A Punnett square illustrating possible genotypes.

32 COMPUTING PROBABILITIES: RIGHT WAYS AND WRONG WAYS

much less common in other ethnic groups). Given this information, what is the probability that a newborn of healthy parents has CF? Call this event CF and let B be the event that both parents are carriers. Then

$$P(\text{CF}) = P(\text{CF given B}) \times P(\text{B}) + P(\text{CF given (not B)}) \times P(\text{not B})$$

where we can note that $P(\text{CF given (not B)}) = 0$ because both parents need to be carriers in order for the child to possibly acquire the disease. Assuming that the mother's and father's genotypes are independent, we get

$$P(\text{B}) = \frac{1}{25} \times \frac{1}{25} = \frac{1}{625}$$

and since the child will get the disease only if it inherits the *C* allele from each parent, we get

$$P(\text{CF given B}) = \frac{1}{4}$$

which gives

$$P(\text{CF}) = \frac{1}{625} \times \frac{1}{4} = \frac{1}{2500}$$

In other words, the *incidence* of CF among newborns is 1 in 2,500, or 0.04%.

Now consider a family with one child where we know that both parents are healthy, that the mother is a carrier of the disease allele and nothing is known about the father's genotype. What is the probability that the child neither is a carrier nor has the disease?

Let E be the event we are interested in. The mother's genotype is *CH*, and we condition on the father's genotype to obtain

$$\begin{aligned} P(\text{E}) &= P(\text{E given CH}) \times P(\text{CH}) + P(\text{E given HH}) \times P(\text{HH}) \\ &= \frac{1}{4} \times \frac{1}{25} + \frac{1}{2} \times \frac{24}{25} \approx 0.49 \end{aligned}$$

where we figured out the conditional probabilities with Punnett squares.

One example of a serious genetic disease is the *Tay–Sachs Disease* which usually leads to death in early childhood. The disease is *autosomal*, which means that the gene is located on one of the chromosomes that is

not a sex chromosome. The disease is also recessive. In the general population, about 1 in 250 individuals are carriers (the gene is more common among certain groups such as Ashkenazi Jews, French Canadians, and Cajuns). This seems pretty high, right? It is not a very well-known disease, yet, it is carried by well over a million Americans. Of course, unless you are genetically tested for it, you will never know you are a carrier. How common is the disease itself? Let us find the probability that a newborn baby has Tay–Sachs.

First of all, both parents need to be carriers. If we assume that the genotype of the father is independent of that of the mother, there is a probability of $1/250 \times 1/250$, which is less than two-tenths of 1%. Second, if both parents are carriers, there is a 25% chance that the child gets the disease which we figure out by considering the Punnett square in Figure 1.6 with a representing the disease gene. Thus, the overall probability of a baby born with Tay–Sachs is $1/250 \times 1/250 \times 1/4$ which equals 1 in 250,000. Note that in this case, individuals who have the disease die long before they can reproduce so there are no other relevant Punnett squares. In general, if a serious autosomal recessive disorder (so serious that affected individuals do not reproduce) has a carrier frequency of 1 in n , the incidence of the disorder among newborns is 1 in $4n^2$. Examples of recessive disorders where affected individuals may (and often do) live long enough to reproduce are sickle cell anemia and the aforementioned CF.

Some genetic conditions are *sex-linked*, that is, the responsible gene is located on one of the sex chromosomes. The prime example of a sex-linked condition is *deuteranopia*, or red-green color blindness, the gene for which is located on the X chromosome. In order to avoid color blindness, it is enough to have one functioning gene and as females have two X chromosomes, whereas males have only one, the condition is more prevalent in men. See Figure 1.7 for examples of Punnett squares. In the one present in (a), neither parent is color blind but the mother is a carrier. The chromosome carrying the gene for color blindness is denoted by X^c , so X^cY is a color blind male, XY is a normal male, XX is a normal female, and X^cX is a female who is a carrier. In this case, there is a 25% chance that the parents get a color blind child.

In order for a female to be color blind, she has to have the genotype X^cX^c . If both parents are color blind, all children will also be color blind, and if the father is color blind and the mother is not, but is a carrier, there is a 50–50 chance of a color blind child, regardless of sex. See the square in

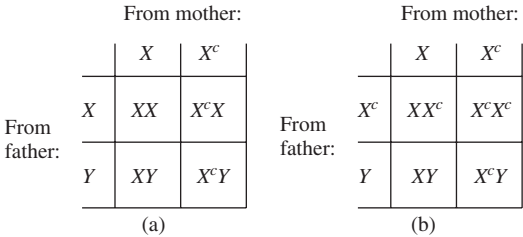


Figure 1.7 Punnett squares for color blindness.

Figure 1.7(b) for the latter situation. It is estimated that about 7% of men are color blind, thus, we can estimate that about 7% of X chromosomes carry the gene for color blindness. As a female needs two copies of this gene to be color blind, the probability of which is $0.07 \times 0.07 = 0.0049$, we would expect about 0.5% of women to be color blind and that is also what has been observed.

Because color blindness is a sex-linked X chromosome condition, for males it does not matter whether the father is color blind because they inherit his Y chromosome so what matters is what they inherit from the mother. Quite different for females; in order for a female to be color blind, her father *must* be color blind and her mother must either be color blind (in which case color blindness of the daughter is a certainty), or be a carrier (in which case there is a 50–50 chance). Thus, a color blind man can have a daughter who is not color blind, and if she has a son, there is a 50% chance that he is color blind. Thus, a man cannot inherit an X chromosome from his father, but he can inherit it from his grandfather by way of his mother. In this sense, color blindness may “skip a generation.” I shall leave it an open question whether the same is true for an Arkansas accent as is claimed by George Costanza in the Seinfeld episode “The Muffin Tops.”

SHUTTLECOCKS AND SPAGHETTI WESTERNS

The law of total probability also works with more than two events. Suppose that Ann and Bob play a game of tennis and that Ann is about to serve at deuce, which means that whoever first gets two points ahead wins the game. Suppose that Ann wins a point with probability $2/3$. What is the probability that she wins the game?

Not so easy to figure out directly. In fact, there is an unlimited number of ways in which Ann can win. She can win two straight points. She can win a point, lose a point, and then win two straight points. She can win, lose, win, . . . , lose, and then win two straight points. Probabilists do not fear infinite sums, and it is possible to find the probability that Ann wins the game by computing and adding the probabilities of all these cases. Feel free to try it for yourself. I will, however, demonstrate a more elegant way. Let us consider three distinct cases:

Case I: Ann wins the next two points

Case II: Bob wins the next two points

Case III: They win a point each, in any order

Now use the law of total probability with the three cases I, II, and III to get the formula

$$\begin{aligned} P(\text{Ann wins}) &= P(\text{Ann wins in I}) \times P(\text{I}) + P(\text{Ann wins in II}) \times P(\text{II}) \\ &\quad + P(\text{Ann wins in III}) \times P(\text{III}) \end{aligned}$$

What are now all these probabilities? We assume independence between consecutive points, which gives that the cases have probabilities

$$P(\text{I}) = 2/3 \times 2/3 = 4/9$$

$$P(\text{II}) = 1/3 \times 1/3 = 1/9$$

$$P(\text{III}) = 4/9$$

where the last probability is computed by adding the first two and subtracting the sum from one. So far so good. Now for the conditional probabilities. The first two are easy; clearly $P(\text{Ann wins in I}) = 1$ and $P(\text{Ann wins in II}) = 0$. But what about the third? In case III, the players are back at deuce, so the probability that Ann wins is now precisely the probability we asked for in the first place. Are we stuck?

No, in fact we are almost done! As $P(\text{Ann wins})$ and $P(\text{Ann wins in III})$ are equal but unknown, let us call this unknown number p and plug it in

above together with the known probabilities:

$$\begin{aligned} p &= 1 \times 4/9 + 0 \times 1/9 + p \times 4/9 \\ &= 4/9 + p \times 4/9 \end{aligned}$$

This formula is now an equation for p that is easy to solve and has the solution $p = 4/5$. The probability that Ann wins is $4/5$. Instead of computing the probability directly, we found three different cases where two could be dealt with explicitly and the third brought us back to the beginning, giving a simple equation for the unknown probability. Wasn't that cool? A general formula if Ann wins a point with probability w is given by

$$P(\text{Ann wins}) = \frac{w^2}{w^2 + (1 - w)^2}$$

which you can try to deduce on your own using the same idea as above. See Figure 1.8 for an illustration of the different cases in a *tree diagram*.

Let us finish by another racket sport problem, this time regarding badminton. In the United States, this sport is mostly considered a backyard game, and if you go by the badminton courts in a college gym, about 90% of the players are Asian, the rest being Scandinavian, with the odd Brit, German, or New Zealander tossed in the mix. However, in August 2005, Howard Bach and Tony Gunawan made history by winning the

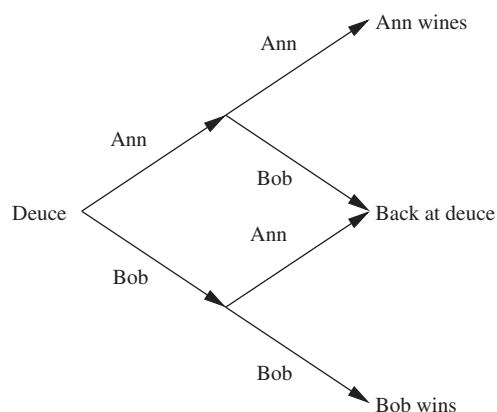


Figure 1.8 The different scenarios when Ann and Bob start from deuce.

men's doubles gold medal in the world championships, the first U.S. players ever to become world champions. It may be surprising to many that badminton is actually the fastest racket sport. The shuttlecock can reach top speeds of 200 mph, which is not bad for a bunch of goose feathers stuck into a cork.

In badminton you can only score when you serve. The exchange of shots is called a "rally"; thus, if you win a rally as server, you score a point. If you win a rally as receiver, the score is unchanged but you get to serve and the opportunity to score. Suppose that Ann and Bob are equally strong players so that Ann wins a rally against Bob with probability $1/2$ regardless of who serves (a reasonable assumption in badminton, but would, of course, not be so in tennis where the server has a big advantage). What is the probability that Ann scores the next point if she is the server?

We will use the same idea as in the tennis example. This time, the three cases are as follows:

Case I: Ann wins the rally

Case II: Ann loses the next two rallies

Case III: Ann loses the rally and wins the following rally.

In case I, Ann scores the point; in case II, Bob scores the point; and in case III, they are back where they started with Ann serving again, no point scored yet. The cases have probabilities $P(I) = 1/2$, $P(II) = 1/4$, and $P(III) = 1/4$, and the first two conditional probabilities are $P(\text{Ann wins in I}) = 1$ and $P(\text{Ann wins in II}) = 0$. The third conditional probability, $P(\text{Ann wins in III})$, is equal to the original probability $P(\text{Ann wins})$, so we denote this unknown number by p and get the equation

$$\begin{aligned} p &= 1 \times 1/2 + 0 \times 1/4 + p \times 1/4 \\ &= 1/2 + p \times 1/4 \end{aligned}$$

which has the solution $p = 2/3$. Thus, the server has a notable advantage and this is even more pronounced if one player is a little better than the other. For example, if Ann has probability 0.55 to win a rally, her probability to win the next point if she is currently the server goes up to 0.73. This phenomenon may explain why seemingly extreme set scores such as

15–3 or 15–4 are not uncommon in badminton tournaments. The general formula this time is given by

$$P(\text{Ann wins}) = \frac{w}{w + (1 - w)^2}$$

where w is the probability that Ann wins a rally. Again, I leave it as an exercise for you to deduce the formula.

The shuttlecock may be fast, but bullets are faster, and even faster is Clint Eastwood's draw of the revolver in the classic 1966 Sergio Leone movie *The Good, the Bad, and the Ugly*. I hope you have seen it and can remember the intense scene toward the end where the three title characters, also known as "Blondie," "Angel Eyes," and "Tuco," stand in a cemetery, guns in holsters, ready to draw. Let us interfere slightly with the script and assume that Blondie always hits his target, Angel Eyes hits with probability 0.9, and Tuco with probability 0.5. Let us also suppose that they take turns in shooting, that whomever is shot at shoots next (unless he is hit), and that Tuco starts. What strategy maximizes his probability of survival?

It seems obvious that he should try to kill Blondie who is the better shot of the other two. Indeed, if he shoots Angel Eyes, Blondie will kill him for sure, so that is not a good strategy. How likely is Tuco to survive with the better strategy? Let S be the event that Tuco survives and let H be the event that he hits his target. There is a 50% chance that Tuco fails to hit and kill Blondie and in that case, Blondie gets the next shot and kills Angel Eyes who is a better shot than Tuco. Tuco then gets one chance to kill Blondie and has a 50% chance to succeed, thus an overall 25% chance of surviving in this case. If Tuco succeeds to kill Blondie, a shootout between Angel Eyes and Tuco takes place until somebody gets killed. Formally, we apply the law of total probability to get the formula

$$\begin{aligned} P(S) &= P(S \text{ given } H) \times P(H) + P(S \text{ given (not } H)) \times P(\text{not } H) \\ &= 0.25 + P(S \text{ given (not } H)) \times 0.5 \end{aligned}$$

where we need to find $P(S \text{ given (not } H))$, which is the probability that Tuco survives a shootout with Angel Eyes (who gets the first shot). Let us ease the notation and rename the event that Tuco survives the shootout T . Now let $p = P(T)$ and consider the three cases

Case I: Angel Eyes hits

Case II: Angel Eyes misses, Tuco hits

Case III: Angel Eyes misses, Tuco misses

to obtain

$$p = P(T \text{ given I}) \times P(I) + P(T \text{ given II}) \times P(II) + P(T \text{ given III}) \times P(III)$$

where $P(I) = 0.9$, $P(II) = 0.1 \times 0.5 = 0.05$, $P(III) = 0.1 \times 0.5 = 0.05$, $P(T \text{ given I}) = 0$, and $P(T \text{ given II}) = 1$. To find $P(T \text{ given III})$, note that if both Angel Eyes and Tuco miss their shots, they start over from the beginning and hence $P(T \text{ given III}) = p$. This gives

$$p = 0.05 + 0.05 \times p$$

which gives $p = 0.05/0.95$, and with this strategy, Tuco has survival probability

$$P(S) = \frac{0.05}{0.95} \times 0.5 + 0.25 \approx 0.28$$

We noted it's bad for Tuco to try to kill Angel Eyes because if he succeeds, he faces certain death as Blondie shoots him. If he fails, Angel Eyes will try to kill Blondie to maximize his own probability of survival. If Angel Eyes fails, Blondie kills him for the same reason and Tuco again gets one last shot at Blondie. Tuco surviving this scenario has probability $0.5 \times 0.1 \times 0.5 = 0.025$. If Angel Eyes succeeds and kills Blondie, Tuco must again survive a shootout with Angel Eyes but this time, Tuco gets to start. By an argument similar to the one above, his probability to survive the shootout is

$$p = 0.5 + 0.05 \times p$$

which gives $p = 0.5/0.95$ and Tuco's survival probability is

$$P(S) = 0.025 + 0.5 \times 0.9 \times \frac{0.5}{0.95} \approx 0.26$$

We have seen that if Tuco tries to kill Blondie, he has a 28% chance of surviving, and if he tries to kill Angel Eyes, he has a slightly smaller chance

of 26% to survive. Notice, though, that Tuco really gains from missing his shot, letting the two better shots fight it out first. The smartest thing he can do is to miss on purpose! If he aims at Blondie and misses, Blondie kills Angel Eyes and Tuco gets one last shot at Blondie with a 50–50 chance to survive. An even better strategy is to aim at Angel Eyes, miss on purpose, and give Angel Eyes a chance to kill Blondie. If Angel Eyes fails, he is a dead man and Tuco gets one last shot at Blondie. If Angel Eyes succeeds, Tuco again needs to survive the shootout which, as we just saw, has probability $p = 0.05/0.95$ and his overall survival probability is

$$P(S) = 0.1 \times 0.5 + 0.9 \times \frac{0.05}{0.95} \approx 0.52$$

By missing on purpose, Tuco has a better than average chance of surviving, which is a much better deal than trying to hit Angel Eyes or Blondie. In his 1965 book *Fifty Challenging Problems in Probability*, Fredric Mosteller presents a similar problem and expresses concern over the possibly unethical dueling conduct to miss on purpose. In the case of Tuco, I think it's safe to assume we can disregard any ethical considerations.

COMBINATORICS, PASTRAMI, AND POETRY

Combinatorics is the mathematics of counting and something that shows up in many probability problems. The fundamental principle in combinatorics is the *multiplication principle*, which is easier to illustrate with examples than try to state formally. Let's do lunch. Suppose that a deli offers three kinds of bread, three kinds of cheese, four kinds of meat, and two kinds of mustard. How many different meat and cheese sandwiches can you make? First choose the bread. For each choice of bread, you then have three choices of cheese, which gives a total of $3 \times 3 = 9$ bread/cheese combinations (rye/swiss, rye/provolone, rye/cheddar, wheat/swiss, wheat/provolone, . . . you get the idea). Then, choose among the four kinds of meat, and finally between the two types of mustard or no mustard at all. You get a total of $3 \times 3 \times 4 \times 2 = 72$ different sandwiches. Suppose that you also have the choice of adding lettuce, tomato, or onion in any combination you want. This choice gives another $2 \times 2 \times 2 = 8$ combinations (you have the choice "yes" or

“no” three times) to combine with the previous 108, so the total is now $108 \times 8 = 864$.

That was the multiplication principle. In each step you have several choices, and to get the total number of combinations, multiply. It is fascinating how quickly the number of combinations grow. Just add one more type of bread, cheese, and meat, respectively, and the number of sandwiches becomes 1,920. It would take years to try them all for lunch.

Another example is to consider how many possible positions there are in chess after two moves. White starts and has 20 possible opening moves. For each of these, black also has 20 possible moves and there are thus $20 \times 20 = 400$ possible positions already after the first two moves (but only a few of these would ever show up in a serious game). After the two opening moves, the number of possible moves depends on the previous moves, but suffice it to say that the number of positions grows very rapidly. No wonder computers are better chess players than people (sorry chess players). A somewhat related example that I am sure you have heard is the tale of the king who agreed to award the inventor of chess by placing one grain of rice on the first square, two on the second, and keep on doubling until the board was full. The last square would then have $2 \times 2 \times \cdots \times 2 = 2^{63}$ grains of rice, which would make enough sushi to feed the entire world for many years.

To provide a link between probability and poetry, we turn to the French poet and novelist Raymond Queneau who in 1961 wrote a book called *One Hundred Thousand Billion Poems*. The book has 10 pages, and each page contains a sonnet, which has 14 lines. There are cuts between the lines so that each line can be turned separately, and because all lines have the same rhyme scheme and rhyme sounds, any such combination gives a readable sonnet. The number of sonnets that can be obtained in this way is thus 10^{14} , which is indeed a hundred thousand billion. Somebody has calculated that it would take about 200 million years of nonstop reading to get through them all. I would instead recommend Queneau’s hilarious *Exercises in Style* in which the same story is retold in 99 different styles and which can be read in an afternoon. One may wonder if Queneau was ever asked what he thought was his best work and replied “I don’t know. I haven’t read most of it.”

How does probability enter into all this? Here is an example. A Swedish license plate consists of three letters followed by three digits. What is the

probability that a randomly chosen such plate has no duplicate letters and no duplicate digits?

The first question is how many letters there are in the Swedish alphabet. Aren't there letters like å, ä, and ö? Yes, but these are not used for license plates. A few others are also not used, and the total number of available letters is 23. There is, therefore, a total of $23 \times 23 \times 23 \times 10 \times 10 \times 10$, which is approximately 12 million license plates (excluding typical Swedish vanity plates such as VIKING or I♥ABBA). To get a plate that has no duplicate letters, we can choose the first letter in any way we want, so for this we have 23 choices. The next letter cannot be the same as the first, so here we have 22 choices. Finally, the third letter cannot be equal to any of the first two, which gives 21 choices. Same for the digits: first 10, then 9, and then 8 choices. The number of plates with no duplicates is thus $23 \times 22 \times 21 \times 10 \times 9 \times 8$. Divide by the total number to get the probability

$$P(\text{no duplicate letters or digits}) = \frac{23 \times 22 \times 21 \times 10 \times 9 \times 8}{23 \times 23 \times 23 \times 10 \times 10 \times 10} \approx 0.63$$

For the license plates, the order is important. For example, ABC123 is different from BCA231. In some combinatorial problems, order is irrelevant, for example, those that have to do with poker. You are dealt a poker hand (5 cards from a regular deck of 52 cards). What is the probability of being dealt a *flush* (five cards in the same suit, not all five consecutive)?

First find the total number of different hands, and then the number of hands that give a flush. As there are 52 ways to choose the first card, 51 ways to choose the second, and so on down to 48 ways to choose the fifth, the multiplication principle tells us that there are $52 \times 51 \times \dots \times 48 = 311,875,200$ different ways to get your cards. But then we have taken order into account and, for example, distinguished between the sequences ($\spadesuit A, \diamondsuit A, \heartsuit 2, \clubsuit A, \clubsuit A$) and ($\clubsuit A, \heartsuit 2, \diamondsuit A, \spadesuit A, \heartsuit A$), and in poker, four aces are four aces regardless of order. As there are $5 \times 4 \times 3 \times 2 \times 1 = 120$ ways to rearrange any given five cards (five choices for the first, four for the second, and so on), we need to divide 311,875,200 by 120 to get the number 2,598,960. There are about 2.6 million different poker hands.

For the number of hands that give a flush, first consider a flush in some given suit, for example, hearts. For the first card, we then have 13 choices, for the second 12, and so on, and each such sequence of five cards can be rearranged in 120 ways, which gives $13 \times 12 \times 11 \times 10 \times 9 / 120 = 1,287$

hands that give five hearts. From this number we must subtract the 10 hands that have five consecutive hearts, because such a hand counts as a *straight flush*, not a mere flush. The subtraction leaves 1,277 hands with a flush in hearts. Finally, as there are four suits, the number of hands that give a flush in any suit is $4 \times 1,277 = 5,108$. The probability to be dealt a flush is therefore $5,108/2,598,960 \approx 0.002$. You are dealt a flush on average once every 500 hands.

Let us introduce some notation that is convenient to use. You may be familiar with the notation for the *factorial* of a number n :

$$n! = n \times (n - 1) \times \cdots \times 2 \times 1$$

Thus, $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$, $5! = 120$, and $6! = 720$. The exclamation mark is not intended to indicate surprise, but factorials do grow surprisingly quickly. For example, the total number of ways of rearranging a deck of cards is $52!$, which is an enormous number. Take out a deck and shuffle it well. Do you think that the particular order of the cards that you got has ever occurred before in a deck in the history of card playing? Most likely not. If all people on earth started shuffling cards and produced one shuffled deck every 10 seconds around the clock they would have to do this for about four million sextillion septillion years to even have a chance of producing all possible orders. That's a number with 51 digits. That's a long time.

We saw above that the number of possible poker hands, that is, the number of ways to choose five cards out of 52 is $52 \times 51 \times \cdots \times 48/5!$. In general, if we choose k out of n objects, there are $n \times (n - 1) \times \cdots \times (n - k + 1)/k!$ ways to do this (convince yourself!). We use the following special notation:

$$\binom{n}{k} = \frac{n \times (n - 1) \times \cdots \times (n - k + 1)}{k!}$$

which is a number that is read “ n choose k .” If the numerator looks messy to you, just remember that it has k factors. Thus, there are $\binom{52}{5}$ different poker hands, $\binom{13}{5} - 10$ hands that give a flush in hearts, and $4 \times (\binom{13}{5} - 10)$ hands that give you a flush in any suit. Check that this agrees with our calculations above. Also convince yourself of the

following identity:

$$\binom{n}{k} = \binom{n}{n-k}$$

which can come in handy in computations. For example, if you have to compute $\binom{10}{8}$ by hand, it is easier to instead compute $\binom{10}{2}$ (do it and you will see why). A quick argument for the formula is that each choice of k objects can also be done by setting aside the $n - k$ objects that you do *not* choose. As there are $\binom{n}{k}$ ways to do the first and $\binom{n}{n-k}$ ways to do the second, the two expressions must be the same. Needless to say, computations by hand are seldom done these days, and even fairly simple pocket calculators have functions to compute $\binom{n}{k}$. The formula is still good to know.

Here is a real-life problem that comes from the field of home care medicine. It has been observed that the risk of a drug interaction is about 6% for a patient who takes two medications and about 50% for a patient who takes five medications. What is the risk of a drug interaction for a patient who takes nine medications? First of all, with nine medications there are $\binom{9}{2} = 9 \times 8/2 = 36$ pairs of medications that can interact. We are now looking for the probability that at least one such pair leads to an interaction and Trick Number One comes in handy again. The probability that there is no interaction between any two medications is 0.94, and if we assume that pairs are independent of each other, the probability of no interaction is

$$P(\text{no interaction}) = 0.94^{36} \approx 0.11$$

and the risk of having an interaction is thus $1 - 0.11 = 0.89$, almost 90%. One can question whether the independence assumption is reasonable; if medication A interacts with medication B, perhaps it is also more likely to interact with other medications. To test this assumption, we can use the other piece of information given, that the risk is about 50% for those who take five medications. With five medications, there are $\binom{5}{2} = 5 \times 4/2 = 10$ pairs, and the risk of interaction is $1 - 0.94^{10} \approx 0.46$ under the independence assumption. The 46% is close enough to the observed 50% to motivate our assumption. This problem was kindly given to me by one anonymous reviewer of my book proposal. Thanks reviewer number three.

A more playful problem now. Roll a die six times. What is the probability that all six sides come up? At first glance, this problem does not seem to have anything to do with combinatorics, but we can translate it into a situation with balls and urns. Let us thus consider six urns, labeled 1 through 6 and six balls, numbered 1 through 6. Roll the die. If it shows 1 put ball number 1 in urn number 1; if it shows 2, put ball number 1 in urn number 2, and so on. Roll the die again, and do the same thing with ball number 2. After all six rolls, the six balls are distributed among the urns. The total number of ways in which this can be done is 6^6 by the multiplication principle (six choices of urn for each of the six balls). To get all different numbers, there are six choices for the first ball, five for the second, and so on; thus, a total of $6!$ ways. We have argued that the probability that all six sides come up is

$$P(\text{all six sides}) = 6!/6^6 \approx 0.015$$

A general formula for the probability that n balls distributed over n urns leaves no urn empty is thus

$$P(\text{no urn empty}) = n!/n^n$$

and even though the factorial in the numerator grows fast, it stands no chance against the denominator. The probability rapidly approaches 0 as n increases.

THE VON TRAPPS AND THE BINOMIAL DISTRIBUTION

Recall the problem on page 6 where we asked for the probability that a family with three children has exactly one daughter. By listing the eight possible outcomes and counting the cases with one daughter, we arrived at the solution $3/8$. Another way to solve this problem is to first note that by independence, each particular sequence of two boys and one girl, for example BBG, has probability $1/2 \times 1/2 \times 1/2 = 1/8$. As there are three such sequences (GBB, BGB, and BBG), we get the probability $3 \times 1/8 = 3/8$.

Let us now instead consider a family with seven children. What is the probability that they, like the von Trapps, have five daughters? There are now $2^7 = 128$ possible outcomes ranging from BBBB BBB to

GGGGGGG. It is tedious to list them all and count how many of them have five girls. Let us instead try the second approach. Each particular sequence with five girls and two boys, for example, GBGGBGG, has probability $(1/2)^7$, so the question is how many such sequences there are. Here is where combinatorics come in. The question becomes: In how many ways can we choose positions for the five Gs? The answer is $\binom{7}{5}$. Recall from the previous section that it is easier to compute this as $\binom{7}{2}$, which equals $7 \times 6/2 = 21$. The probability that a family with seven children has five daughters is thus

$$P(\text{five daughters}) = 21 \times (1/2)^7 \approx 0.16$$

We've done the von Trapps. Now you do the Jacksons. What is the probability that a family with nine children has three daughters? ⁵

Here is another problem that can be solved in the same way. If you roll a die 12 times, you expect to get on average two 6s but what is the probability to get exactly two 6s? First think of a particular sequence of 12 rolls with two 6s, for example, XX6X6XXXXXXXX, where "X" means "something else." By independence we multiply and get the probability $5/6 \times 5/6 \times 1/6 \times \cdots \times 5/6$, which we can also write as $(1/6)^2 \times (5/6)^{10}$. But this is the probability for any specified sequence with two 6s, so the question again is: How many sequences are there? And just like above, we need to choose positions, this time for two 6s in a sequence of 12 rolls. We get

$$P(\text{two 6s in twelve rolls}) = \binom{12}{2} \times (1/6)^2 \times (5/6)^{10} \approx 0.3$$

If we were to use the "old method" of counting in the sample space, we must first note that because 6 and X are not equally likely, we cannot use the sample space of the $2^{12} = 4,096$ outcomes ranging from XXXXXXXXXXXX to 666666666666. We have to break up each X into five different outcomes, combine all of these, and end up with a sample space with 6^{12} , a bit over two billion, equally likely outcomes. It is possible to proceed and solve the problem in this way, but I would not recommend it as a general method.

⁵For bonus points, name the daughter that is not Janet or La Toya.

Let us do the general formula now. Suppose that an experiment (such as giving birth or rolling a die) is repeated n times. We refer to each repetition of the experiment as a *trial*. Each trial results in a “success” with some probability p , independently of previous trials. The probability to get exactly k successes is then

$$P(k \text{ successes}) = \binom{n}{k} \times p^k \times (1 - p)^{n-k}$$

where k can be anything from 0 to n . So that the formula makes sense for $k = 0$ and $k = n$, the number $\binom{n}{0}$ is defined to be equal to 1, and any number raised to the power 0 is also defined to be 1. The probability to get 0 successes is then $(1 - p)^n$, and the probability to get all n successes is p^n . We say that the number of successes has a *binomial distribution* (the numbers $\binom{n}{k}$ are called *binomial coefficients* and you may be familiar with *Newton’s binomial theorem*). The numbers n and p are called the *parameters* of the binomial distribution. In the von Trapp example, we have $n = 7$ and $p = 1/2$; in the dice example, $n = 12$ and $p = 1/6$. In the problem with drug interactions on page 44, the number of interactions with nine medications has a binomial distribution with $n = 36$ and $p = 0.06$, and we computed the probability for $k = 0$.

Two assumptions are crucial for the binomial distribution. First, successive trials must be independent of each other, and second, the success probability p must be the same in each trial. Let me illustrate this in an example. Call a day “hot” if the high temperature is above 90 degrees, and suppose that the probability of a hot day in New Orleans in early July is 0.7. You now decide to count the number of hot days among

- (a) Each Fourth of July for the next 5 years
- (b) Each day in the first week of July next year
- (c) Each first day of the month next year

Does any of (a), (b), or (c) give you a binomial distribution for the number of hot days?

The answer is that only (a) gives a binomial distribution. The temperature on July 4th one year is certainly independent of the temperature on

July 4th another year, and it is reasonable to assume that our success probability 0.7 stays the same for another 5 years (global warming isn't *that* fast). Thus, you have a binomial distribution with parameters $n = 5$ and $p = 0.7$. In (b), the trials are not independent. If you have a hot day on July 1, you are more likely to get a hot day also on July 2 because there is a weather system in place that gives you hot temperatures. Thus, you do not have a binomial distribution in this case. The fact that an arbitrary day in early July is hot with probability 0.7 means that over the years, about 70% of early July days have been hot. These hot days have typically been concentrated to certain years though. Some years all or most days in early July were hot, and some years there were few, if any. Thus, on average, 7 out of 10 days are hot, but consecutive days are not independent.

In (c) finally, although it might be reasonable to assume that the temperature on the first of one month is independent of the temperature on the first of another month (weather systems don't usually stay around for that long), the problem is that the success probability changes. The probability of a hot day is, for example, far less than 0.7 in January, so you do not have a binomial distribution here either.

Recall the pub evening from page 30, but let us suppose that you instead use your strategies to guess the number of days the bus is late. The number of such late days has a binomial distribution with the parameters $n = 5$ (the number of work days in a week) and $p = 0.4$ (the probability that the bus is late on any given day). This time Betsy fares the worst. The probability that she is correct is the probability that the bus is never late, which corresponds to $k = 0$ in the formula and gives probability $0.6^5 \approx 0.08$. Albert is correct if the bus is late twice and the probability of this is

$$P(\text{the bus is late twice}) = \binom{5}{2} \times 0.4^2 \times 0.6^3 \approx 0.35$$

which also answers the question I posed on page 31. Your chances are again a bit more complicated to compute. As both your guess and the actual outcome have binomial distributions, we need to use the law of total probability. If you guess 0, which happens with probability $0.6^5 \approx 0.08$, you are correct if the bus is never late, which also happens with probability 0.08. The first term in the law of total probability is $0.08 \times 0.08 = 0.08^2$. For the next term, we need the binomial probability when $k = 1$, and this is $5 \times 0.4 \times 0.6^4 \approx 0.26$. This is the probability that you guess 1,

and in that case, you are correct if the actual number is also 1. Thus, square this probability and add to the previous: $0.08^2 + 0.26^2$. Continue in this way, compute and square each of the six binomial probabilities for k from 0 to 5, and then add them all. The result is the probability that you guess correctly and you may verify that it is about 0.25. Albert has been avenged.

Let us toss some more coins. If you toss four coins, the “typical” outcome is to get two heads and two tails. The probability of this outcome can be computed with the binomial distribution as

$$P(\text{two heads}) = \binom{4}{2} \times (1/2)^4 = 3/8$$

Likewise, the typical outcome when you toss six coins is three heads, when you toss eight coins, four heads, and so on. In general, how likely is it that you get the typical outcome, that is, equally many heads and tails when you toss an even number of coins? Suppose, thus, that you toss $2 \times n$ coins and ask for the probability to get n heads, in particular when n gets large. By the binomial distribution

$$P(n \text{ heads}) = \binom{2 \times n}{n} \times (1/2)^{2 \times n}$$

and it is not so easy to see where this heads for large n . There is, however, a nice approximation formula for factorials, called *Stirling's formula*, that comes in handy. This formula is quite technical, and I do not want to go into that kind of detail, so if you are interested, look up the formula on your own. It is pretty neat. Anyway, it turns out the approximate probability to get equally many heads and tails is

$$P(\text{equally many heads and tails in } 2 \times n \text{ tosses}) \approx 1/\sqrt{n \times \pi}$$

You may wonder what on earth the number π is doing in there. Isn't that the ratio of the circumference and the diameter of a circle, the famous 3.14? Yes indeed, but as anybody who has studied mathematics knows, the number π tends to pop up in the most unexpected situations.⁶ Better

⁶Other seemingly nonsequitur appearances of π are that if an integer is chosen at random, the probability that it is *square-free* (cannot be divided by any square such as 4, 9, ...) is $6/\pi^2$, and that if two integers are chosen at random, the probability that they are *relatively prime* (have no common divisors) is also $6/\pi^2$.

get used to it; you will see it again. Just for fun, let us try the formula for four coins, that is, $n = 2$. The approximation gives

$$P(\text{two heads and two tails}) \approx 1/\sqrt{2 \times \pi} \approx 0.40$$

and the exact answer we got above was $3/8$, in decimal notation 0.375. Not bad. Of course, this is about as small an n as we can have and the approximation formula only gets better when n is large, which is also when the formula is useful. Note that the probability $1/\sqrt{n \times \pi}$ goes to 0 as n increases. Thus, the typical outcome is actually very unlikely. It is only typical in the sense of an average; if the $2 \times n$ coins are tossed over and over, the average number of heads will be near n , but it will not happen often that we get *exactly* n heads. More about this issue later in the book.

Let me finish with a sports example. Is it easier for the underdog to win the Super Bowl or the World Series? The difference is that the Super Bowl is a single game, but the World Series is played in best of seven games. Which benefits the weaker team? Let us ignore all practical complications such as home field advantage and all kinds of unpredictable events and simply assume that each game is won by the underdog with probability p , independently of other games. In order to win the World Series, four games must be won and if this goal is achieved in less than seven games, the remaining games are not played. This gives the following four different ways to win: win four straight games; win three of the first four games and the fifth; win three of the first five games and the sixth; and win three of the first six games and the seventh. In each of these cases, the last game must be won and this has probability p . Moreover, three games must be won among the first three, four, five, or six games, and in these cases, the number of games won has a binomial distribution with n equal to 3, 4, 5, and 6, respectively. The probability that the underdog wins can now be computed as

$$P(\text{underdog wins}) = \sum_{n=3}^6 \binom{n}{3} \times p^3 \times (1-p)^{n-3} \times p$$

where we can try different values of p . In Table 1.2, the Super Bowl and World Series are compared for different winning probabilities for the underdog.

Table 1.2 Probability that the underdog wins the Super Bowl and the World Series

P(win single game)	50%	40%	30%	20%	10%
P(win Super Bowl)	50%	40%	30%	20%	10%
P(win World Series)	50%	29%	13%	3.3%	0.3%

Note that it is always easier for the underdog to win the Super Bowl than the World Series; the better team always benefits from playing more games. For example, a team that has a one in five shot to win a single game has only a 3.3% chance to win the World Series. We would expect more upsets in the Super Bowl than in the World Series, but I leave it to you to do the empirical investigation of historical championship data.

FINAL WORD

We are done with the introductory chapter. You are now armed with knowledge about probabilities and how to compute and interpret them. It's time to get to work.

