17 Conditional Probability

17.1 Monty Hall Confusion

Remember how we said that the Monty Hall problem confused even professional mathematicians? Based on the work we did with tree diagrams, this may seem surprising—the conclusion we reached followed routinely and logically. How could this problem be so confusing to so many people?

Well, one flawed argument goes as follows: let's say the contestant picks door A. And suppose that Carol, Monty's assistant, opens door B and shows us a goat. Let's use the tree diagram 16.3 from Chapter 16 to capture this situation. There are exactly three outcomes where contestant chooses door A, and there is a goat behind door B:

$$(A, A, B), \ (A, A, C), \ (C, A, B). \tag{17.1}$$

These outcomes have respective probabilities 1/18, 1/18, 1/9.

Among those outcomes, switching doors wins only on the last outcome, (C, A, B). The other two outcomes *together* have the *same* 1/9 probability as the last one So in this situation, the probability that we win by switching is the *same* as the probability that we lose. In other words, in this situation, switching isn't any better than sticking!

Something has gone wrong here, since we know that the actual probability of winning by switching in 2/3. The mistaken conclusion that sticking or switching are equally good strategies comes from a common blunder in reasoning about how probabilities change given some information about what happened. We have asked for the probability that one event, [win by switching], happens, *given* that another event, [pick A AND goat at B], happens. We use the notation

$$\Pr\big[[\text{win by switching}] \mid [\text{pick A AND goat at B}]\big]$$

for this probability which, by the reasoning above, equals 1/2.

17.1.1 Behind the Curtain

A "given" condition is essentially an instruction to focus on only some of the possible outcomes. Formally, we're defining a new sample space consisting only of some of the outcomes. In this particular example, we're given that the player chooses door A and that there is a goat behind B. Our new sample space therefore consists

solely of the three outcomes listed in (17.1). In the opening of Section 17.1, we calculated the conditional probability of winning by switching given that one of these outcome happened, by weighing the 1/9 probability of the win-by-switching outcome, (C, A, B), against the $1/18 + 1/18 + 1/9$ probability of the three outcomes in the new sample space.

$$\Pr\big[[\text{win by switching}] \mid [\text{pick A AND goat at B}]\big] = \Pr\big[(C, A, B) \mid \{(C, A, B), (A, A, B), (A, A, C)\}\big]$$
$$+ \frac{\Pr[(C, A, B)]}{\Pr[\{(C, A, B), (A, A, B), (A, A, C)\}]} = \frac{1/9}{1/18 + 1/18 + 1/9} = \frac{1}{2}.$$

There is nothing wrong with this calculation. So how come it leads to an incorrect conclusion about whether to stick or switch? The answer is that this was the wrong thing to calculate, as we'll explain in the next section.

17.2   Definition and Notation

The expression $\Pr\big[X \mid Y\big]$ denotes the probability of event X, given that event Y happens. In the example above, event X is the event of winning on a switch, and event Y is the event that a goat is behind door B and the contestant chose door A. We calculated $\Pr\big[X \mid Y\big]$ using a formula which serves as the definition of conditional probability:

Definition 17.2.1. Let X and Y be events where Y has nonzero probability. Then

$$\Pr\big[X \mid Y\big] ::= \frac{\Pr[X \cap Y]}{\Pr[Y]}.$$

The conditional probability $\Pr\big[X \mid Y\big]$ is undefined when the probability of event Y is zero. To avoid cluttering up statements with uninteresting hypotheses that conditioning events like Y have nonzero probability, we will make an implicit assumption from now on that all such events have nonzero probability.

Pure probability is often counterintuitive, but conditional probability can be even worse. Conditioning can subtly alter probabilities and produce unexpected results in randomized algorithms and computer systems as well as in betting games. But Definition 17.2.1 is very simple and causes no trouble—provided it is properly applied.

17.2.1   What went wrong

So if everything in the opening Section 17.1 is mathematically sound, why does it seem to contradict the results that we established in Chapter 16? The problem is a

common one: *we chose the wrong condition.* In our initial description of the scenario, we learned the location of the goat when Carol opened door B. But when we defined our condition as "the contestant opens A and the goat is behind B," we included the outcome (A, A, C) in which Carol opens door C! The correct conditional probability should have been "what are the odds of winning by switching given the contestant chooses door A and Carol opens door B." By choosing a condition that did not reflect everything known. we inadvertently included an extraneous outcome in our calculation. With the correct conditioning, we still win by switching 1/9 of the time, but the smaller set of known outcomes has smaller total probability:

$$\Pr[\{(A, A, B), (C, A, B)\}] = \frac{1}{18} + \frac{1}{9} = \frac{3}{18}.$$

The conditional probability would then be:

$$\Pr\big[[\text{win by switching}] \mid [\text{pick A AND Carol opens B}]\big] = \Pr\big[(C, A, B) \mid \{(C, A, B), (A, A, B)\}\big]$$

$$+ \frac{\Pr[(C, A, B)]}{\Pr[\{(C, A, B), (A, A, B)\}]} = \frac{1/9}{1/9 + 1/18} = \frac{1}{2}.$$

which is exactly what we already deduced from the tree diagram 16.2 in the previous chapter.

The O. J. Simpson Trial

In an opinion article in the *New York Times*, Steven Strogatz points to the O. J. Simpson trial as an example of poor choice of conditions. O. J. Simpson was a retired football player who was accused, and later acquitted, of the murder of his wife, Nicole Brown Simpson. The trial was widely publicized and called the "trial of the century." Racial tensions, allegations of police misconduct, and new-at-the-time DNA evidence captured the public's attention. But Strogatz, citing mathematician and author I.J. Good, focuses on a less well-known aspect of the case: whether O. J.'s history of abuse towards his wife was admissible into evidence.

The prosecution argued that abuse is often a precursor to murder, pointing to statistics indicating that an abuser was as much as ten times more likely to commit murder than was a random indidual. The defense, however, countered with statistics indicating that the odds of an abusive husband murdering his wife were "infinitesimal," roughly 1 in 2500. Based on those numbers, the actual relevance of a history of abuse to a murder case would appear limited at best. According to the defense, introducing that history would make the jury hate Simpson but would lack any probitive value. Its discussion should be barred as prejudicial.

In other words, both the defense and the prosecution were arguing conditional probability, specifically the likelihood that a woman will be murdered by her husband, given that her husband abuses her. But both defense and prosecution omitted a vital piece of data from their calculations: Nicole Brown Simpson *was* murdered. Strogatz points out that based on the defense's numbers and the crime statistics of the time, the probability that a woman was murdered by her abuser, given that she was abused *and* murdered, is around 80%.

Strogatz's article goes into more detail about the calculations behind that 80% figure. But the real point we wanted to make is that conditional probability is used and misused all the time, and even experts under public scrutiny make mistakes.

17.3   The Four-Step Method for Conditional Probability

In a best-of-three tournament, the local C-league hockey team wins the first game with probability 1/2. In subsequent games, their probability of winning is determined by the outcome of the previous game. If the local team won the previous game, then they are invigorated by victory and win the current game with probability 2/3. If they lost the previous game, then they are demoralized by defeat and win the current game with probability only 1/3. What is the probability that the

local team wins the tournament, given that they win the first game?

This is a question about a conditional probability. Let A be the event that the local team wins the tournament, and let B be the event that they win the first game. Our goal is then to determine the conditional probability $\Pr\left[A \mid B\right]$.

We can tackle conditional probability questions just like ordinary probability problems: using a tree diagram and the four step method. A complete tree diagram is shown in Figure 17.1.
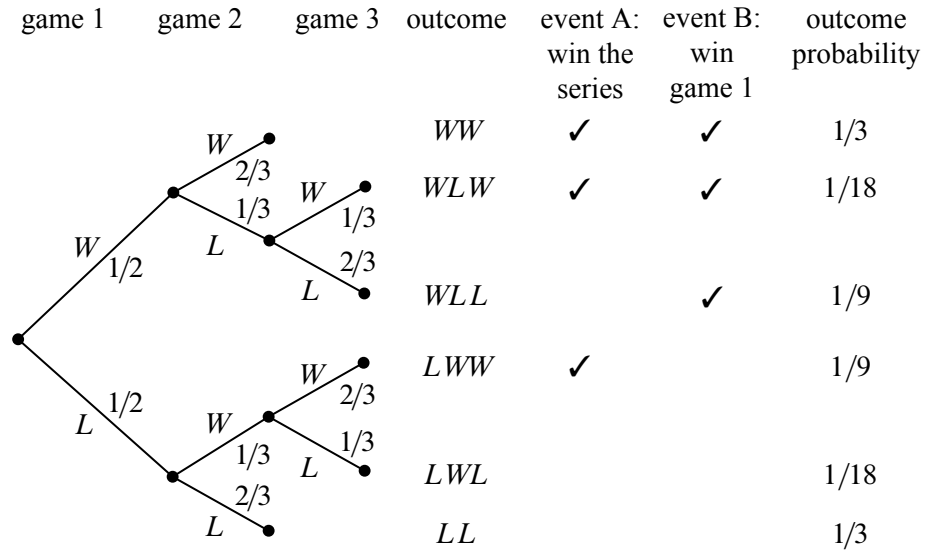
game 1	game 2	game 3	outcome	event A: win the series	event B: win game 1	outcome probability
			WW	✓	✓	1/3
			WLW	✓	✓	1/18
			WLL		✓	1/9
			LWW	✓		1/9
			LWL			1/18
			LL			1/3

Figure 17.1 The tree diagram for computing the probability that the local team wins two out of three games given that they won the first game.

Step 1: Find the Sample Space

Each internal vertex in the tree diagram has two children, one corresponding to a win for the local team (labeled W) and one corresponding to a loss (labeled L). The complete sample space is:

$$\mathcal{S} = \{WW, WLW, WLL, LWW, LWL, LL\}.$$

Step 2: Define Events of Interest

The event that the local team wins the whole tournament is:

$$T = \{WW, WLW, LWW\}.$$

And the event that the local team wins the first game is:

$$F = \{WW, WLW, WLL\}.$$

The outcomes in these events are indicated with check marks in the tree diagram in Figure 17.1.

Step 3: Determine Outcome Probabilities

Next, we must assign a probability to each outcome. We begin by labeling edges as specified in the problem statement. Specifically, the local team has a 1/2 chance of winning the first game, so the two edges leaving the root are each assigned probability 1/2. Other edges are labeled 1/3 or 2/3 based on the outcome of the preceding game. We then find the probability of each outcome by multiplying all probabilities along the corresponding root-to-leaf path. For example, the probability of outcome WLL is:

$$\frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{1}{9}.$$

Step 4: Compute Event Probabilities

We can now compute the probability that the local team wins the tournament, given that they win the first game:

$$
\begin{aligned}
\Pr\left[A \mid B\right] &= \frac{\Pr[A \cap B]}{\Pr[B]} \\
&= \frac{\Pr[\{WW, WLW\}]}{\Pr[\{WW, WLW, WLL\}]} \\
&= \frac{1/3 + 1/18}{1/3 + 1/18 + 1/9} \\
&= \frac{7}{9}.
\end{aligned}
$$

We're done! If the local team wins the first game, then they win the whole tournament with probability 7/9.

17.4   Why Tree Diagrams Work

We've now settled into a routine of solving probability problems using tree diagrams. But we've left a big question unaddressed: mathematical justification behind those funny little pictures. Why do they work?

The answer involves conditional probabilities. In fact, the probabilities that we've been recording on the edges of tree diagrams *are* conditional probabilities. For example, consider the uppermost path in the tree diagram for the hockey team problem, which corresponds to the outcome WW. The first edge is labeled 1/2,

which is the probability that the local team wins the first game. The second edge is labeled 2/3, which is the probability that the local team wins the second game, *given* that they won the first—a conditional probability! More generally, on each edge of a tree diagram, we record the probability that the experiment proceeds along that path, given that it reaches the parent vertex.

So we've been using conditional probabilities all along. For example, we concluded that:

$$\Pr[WW] = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}.$$

Why is this correct?

The answer goes back to Definition 17.2.1 of conditional probability which could be written in a form called the *Product Rule* for conditional probabilities:

Rule (Conditional Probability Product Rule: 2 Events)**.**

$$\Pr[E_1 \cap E_2] = \Pr[E_1] \cdot \Pr[E_2 \mid E_1].$$

Multiplying edge probabilities in a tree diagram amounts to evaluating the right side of this equation. For example:

$$\Pr[\text{win first game} \cap \text{win second game}]$$
$$= \Pr[\text{win first game}] \cdot \Pr[\text{win second game} \mid \text{win first game}]$$
$$= \frac{1}{2} \cdot \frac{2}{3}.$$

So the Conditional Probability Product Rule is the formal justification for multiplying edge probabilities to get outcome probabilities.

To justify multiplying edge probabilities along a path of length three, we need a rule for three events:

Rule (Conditional Probability Product Rule: 3 Events)**.**

$$\Pr[E_1 \cap E_2 \cap E_3] = \Pr[E_1] \cdot \Pr[E_2 \mid E_1] \cdot \Pr[E_3 \mid E_1 \cap E_2].$$

An *n*-event version of the Rule is given in Problem 17.1, but its form should be clear from the three event version.

17.4.1 Probability of Size-*k* Subsets

As a simple application of the product rule for conditional probabilities, we can use the rule to calculate the number of size-*k* subsets of the integers [1..*n*]. Of course we already know this number is $\binom{n}{k}$, but now the rule will give us a new derivation of the formula for $\binom{n}{k}$.

Let's pick some size-k subset, $S \subseteq [1..n]$, as a target. Suppose we choose a size-k subset at random, with all subsets of $[1..n]$ equally likely to be chosen, and let p be the probability that our randomly chosen equals this target. That is, the probability of picking S is p, and since all sets are equally likely to be chosen, the number of size-k subsets equals $1/p$.

So what's p? Well, the probability that the *smallest* number in the random set is one of the k numbers in S is k/n. Then, *given* that the smallest number in the random set is in S, the probability that the *second* smallest number in the random set is one of the remaining $k-1$ elements in S is $(k-1)/(n-1)$. So by the product rule, the probability that the *two* smallest numbers in the random set are both in S is

$$\frac{k}{n} \cdot \frac{k-1}{n-1}.$$

Next, given that the two smallest numbers in the random set are in S, the probability that the third smallest number is one of the $k-2$ remaining elements in S is $(k-2)/(n-2)$. So by the product rule, the probability that the *three* smallest numbers in the random set are all in S is

$$\frac{k}{n} \cdot \frac{k-1}{n-1} \cdot \frac{k-2}{n-2}.$$

Continuing in this way, it follows that the probability that *all k* elements in the randomly chosen set are in S, that is, the probabilty that the randomly chosen set equals the target, is

$$
\begin{aligned}
p &= \frac{k}{n} \cdot \frac{k-1}{n-1} \cdot \frac{k-2}{n-2} \cdots \frac{k-(k-1)}{n-(k-1)} \\
&= \frac{k \cdot (k-1) \cdot (k-1) \cdots 1}{n \cdot (n-1) \cdot (n-2) \cdots (n-(k-1))} \\
&= \frac{k!}{n!/(n-k)!} \\
&= \frac{k!(n-k)!}{n!}.
\end{aligned}
$$

So we have again shown the number of size-k subsets of $[1..n]$, namely $1/p$, is

$$\frac{n!}{k!(n-k)!}.$$

17.4.2   Medical Testing

Breast cancer is a deadly disease that claims thousands of lives every year. Early detection and accurate diagnosis are high priorities, and routine mammograms are

one of the first lines of defense. They're not very accurate as far as medical tests go, but they are correct between 90% and 95% of the time, which seems pretty good for a relatively inexpensive non-invasive test.[1] However, mammogram results are also an example of conditional probabilities having counterintuitive consequences. If the test was positive for breast cancer in you or a loved one, and the test is better than 90% accurate, you'd naturally expect that to mean there is better than 90% chance that the disease was present. But a mathematical analysis belies that gut instinct. Let's start by precisely defining how accurate a mammogram is:

- If you have the condition, there is a 10% chance that the test will say you do not have it. This is called a "false negative."

- If you do not have the condition, there is a 5% chance that the test will say you do. This is a "false positive."

17.4.3 Four Steps Again

Now suppose that we are testing middle-aged women with no family history of cancer. Among this cohort, incidence of breast cancer rounds up to about 1%.

Step 2: Define Events of Interest
Let A be the event that the person has breast cancer. Let B be the event that the test was positive. The outcomes in each event are marked in the tree diagram. We want to find $\Pr\left[A \mid B\right]$, the probability that a person has breast cancer, given that the test was positive.

Step 3: Find Outcome Probabilities
First, we assign probabilities to edges. These probabilities are drawn directly from the problem statement. By the Product Rule, the probability of an outcome is the product of the probabilities on the corresponding root-to-leaf path. All probabilities are shown in Figure 17.2.

Step 4: Compute Event Probabilities
From Definition 17.2.1, we have

$$\Pr\left[A \mid B\right] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{0.009}{0.009 + 0.0495} \approx 15.4\%.$$

So, if the test is positive, then there is an 84.6% chance that the result is incorrect, even though the test is nearly 95% accurate! So this seemingly pretty accurate test doesn't tell us much. To see why percent accuracy is no guarantee of value,

---

[1]The statistics in this example are roughly based on actual medical data, but have been rounded or simplified for illustrative purposes.

Step 1: Find the Sample Space

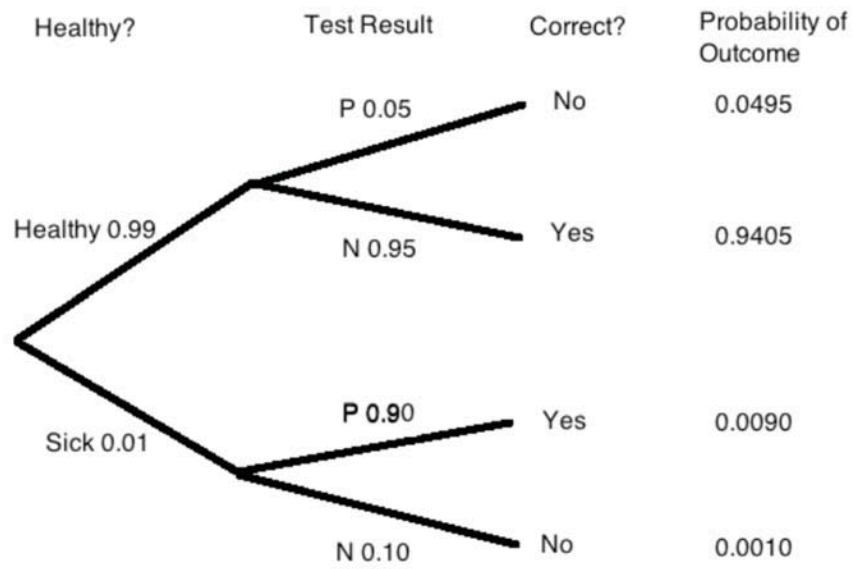The sample space is found with the tree diagram in Figure 17.2.

Figure 17.2   The tree diagram for a breast cancer test.

notice that there is a simple way to make a test that is 99% accurate: always return a negative result! This test gives the right answer for all healthy people and the wrong answer only for the 1% that actually have cancer. This 99% accurate test tells us nothing; the "less accurate" mammogram is still a lot more useful.

17.4.4 Natural Frequencies

That there is only about a 15% chance that the patient actually has the condition when the test say so may seem surprising at first, but it makes sense with a little thought. There are two ways the patient could test positive: first, the patient could have the condition and the test could be correct; second, the patient could be healthy and the test incorrect. But almost everyone is healthy! The number of healthy individuals is so large that even the mere 5% with false positive results overwhelm the number of genuinely positive results from the truly ill.

Thinking like this in terms of these "natural frequencies" can be a useful tool for interpreting some of the strange seeming results coming from those formulas. For example, let's take a closer look at the mammogram example.

Imagine 10,000 women in our demographic. Based on the frequency of the disease, we'd expect 100 of them to have breast cancer. Of those, 90 would have a positve result. The remaining 9,900 woman are healthy, but 5% of them—500, give or take—will show a false positive on the mammogram. That gives us 90 real positives out of a little fewer than 600 positives. An 85% error rate isn't so surprising after all.

17.4.5 *A Posteriori* Probabilities

If you think about it much, the medical testing problem we just considered could start to trouble you. You may wonder if a statement like "If someone tested positive, then that person has the condition with probability 18%" makes sense, since a given person being tested either has the disease or they don't.

One way to understand such a statement is that it just means that 15% of the people who test positive will actually have the condition. Any particular person has it or they don't, but a *randomly selected* person among those who test positive will have the condition with probability 15%.

But what does this 15% probability tell you if you *personally* got a positive result? Should you be relieved that there is less than one chance in five that you have the disease? Should you worry that there is nearly one chance in five that you do have the disease? Should you start treatment just in case? Should you get more tests?

These are crucial practical questions, but it is important to understand that they are not *mathematical* questions. Rather, these are questions about statistical judge-

ments and the philosophical meaning of probability. We'll say a bit more about this after looking at one more example of after-the-fact probabilities.

The Hockey Team in Reverse

Suppose that we turn the hockey question around: what is the probability that the local C-league hockey team won their first game, given that they won the series?

As we discussed earlier, some people find this question absurd. If the team has already won the tournament, then the first game is long since over. Who won the first game is a question of fact, not of probability. However, our mathematical theory of probability contains no notion of one event preceding another. There is no notion of time at all. Therefore, from a mathematical perspective, this is a perfectly valid question. And this is also a meaningful question from a practical perspective. Suppose that you're told that the local team won the series, but not told the results of individual games. Then, from your perspective, it makes perfect sense to wonder how likely it is that local team won the first game.

A conditional probability $\Pr\left[B \mid A\right]$ is called *a posteriori* if event B precedes event A in time. Here are some other examples of a posteriori probabilities:

- The probability it was cloudy this morning, given that it rained in the afternoon.

- The probability that I was initially dealt two queens in Texas No Limit Hold 'Em poker, given that I eventually got four-of-a-kind.

from ordinary probabilities; the distinction comes from our view of causality, which is a philosophical question rather than a mathematical one.

Let's return to the original problem. The probability that the local team won their first game, given that they won the series is $\Pr\left[B \mid A\right]$. We can compute this using the definition of conditional probability and the tree diagram in Figure 17.1:

$$\Pr\left[B \mid A\right] = \frac{\Pr[B \cap A]}{\Pr[A]} = \frac{1/3 + 1/18}{1/3 + 1/18 + 1/9} = \frac{7}{9}.$$

In general, such pairs of probabilities are related by Bayes' Rule:

Theorem 17.4.1 (Bayes' Rule)**.**

$$\Pr\left[B \mid A\right] = \frac{\Pr\left[A \mid B\right] \cdot \Pr[B]}{\Pr[A]} \tag{17.2}$$

Proof. We have

$$\Pr\left[B \mid A\right] \cdot \Pr[A] = \Pr[A \cap B] = \Pr\left[A \mid B\right] \cdot \Pr[B]$$

by definition of conditional probability. Dividing by $\Pr[A]$ gives (17.2).   ■

17.4.6   Philosphy of Probability

Let's try to assign a probability to the event

$$[2^{6972607} - 1 \text{ is a prime number}]$$

It's not obvious how to check whether such a large number is prime, so you might try an estimation based on the density of primes. The Prime Number Theorem implies that only about 1 in 5 million numbers in this range are prime, so you might say that the probability is about $2 \cdot 10^{-8}$. On the other hand, given that we chose this example to make some philosophical point, you might guess that we probably purposely chose an obscure looking prime number, and you might be willing to make an even money bet that the number is prime. In other words, you might think the probability is 1/2. Finally, we can take the position that assigning a probability to this statement is nonsense because there is no randomness involved; the number is either prime or it isn't. This is the view we take in this text.

An alternate view is the *Bayesian* approach, in which a probability is interpreted as a *degree of belief* in a proposition. A Bayesian would agree that the number above is either prime or composite, but they would be perfectly willing to assign a probability to each possibility. The Bayesian approach is very broad in its willingness to assign probabilities to any event, but the problem is that there is no single "right" probability for an event, since the probability depends on one's initial beliefs. On the other hand, if you have confidence in some set of initial beliefs, then Bayesianism provides a convincing framework for updating your beliefs as further information emerges.

As an aside, it is not clear whether Bayes himself was Bayesian in this sense. However, a Bayesian would be willing to talk about the probability that Bayes was Bayesian.

Another school of thought says that probabilities can only be meaningfully applied to *repeatable processes* like rolling dice or flipping coins. In this *frequentist* view, the probability of an event represents the fraction of trials in which the event occurred. So we can make sense of the *a posteriori* probabilities of the C-league hockey example of Section 17.4.5 by imagining that many hockey series were played, and the probability that the local team won their first game, given that they won the series, is simply the fraction of series where they won the first game among all the series they won.

Getting back to prime numbers, we mentioned in Section 8.5.1 that there is a probabilistic primality test. If a number N is composite, there is at least a 3/4 chance that the test will discover this. In the remaining 1/4 of the time, the test is inconclusive. But as long as the result is inconclusive, the test can be run independently again and again up to, say, 1000 times. So if N actually is composite, then

the probability that 1000 repetitions of the probabilistic test do not discover this is at most:

$$\left(\frac{1}{4}\right)^{1000}.$$

If the test remained inconclusive after 1000 repetitions, it is still logically possible that N is composite, but betting that N is prime would be the best bet you'll ever get to make! If you're comfortable using probability to describe your personal belief about primality after such an experiment, you are being a Bayesian. A frequentist would not assign a probability to N's primality, but they would also be happy to bet on primality with tremendous confidence. We'll examine this issue again when we discuss polling and confidence levels in Section 19.5.

Despite the philosophical divide, the real world conclusions Bayesians and Frequentists reach from probabilities are pretty much the same, and even where their interpretations differ, they use the same theory of probability.

17.5   The Law of Total Probability

Breaking a probability calculation into cases simplifies many problems. The idea is to calculate the probability of an event A by splitting into two cases based on whether or not another event E occurs. That is, calculate the probability of $A \cap E$ and $A \cap \overline{E}$. By the Sum Rule, the sum of these probabilities equals $\Pr[A]$. Expressing the intersection probabilities as conditional probabilities yields:

Rule 17.5.1 (Law of Total Probability: single event)**.**

$$\Pr[A] = \Pr\left[A \mid E\right] \cdot \Pr[E] + \Pr\left[A \mid \overline{E}\right] \cdot \Pr[\overline{E}].$$

For example, suppose we conduct the following experiment. First, we flip a fair coin. If heads comes up, then we roll one die and take the result. If tails comes up, then we roll two dice and take the sum of the two results. What is the probability that this process yields a 2? Let E be the event that the coin comes up heads, and let A be the event that we get a 2 overall. Assuming that the coin is fair, $\Pr[E] = \Pr[\overline{E}] = 1/2$. There are now two cases. If we flip heads, then we roll a 2 on a single die with probability $\Pr\left[A \mid E\right] = 1/6$. On the other hand, if we flip tails, then we get a sum of 2 on two dice with probability $\Pr\left[A \mid \overline{E}\right] = 1/36$. Therefore, the probability that the whole process yields a 2 is

$$\Pr[A] = \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{36} = \frac{7}{72}.$$

This rule extends to any set of disjoint events that make up the entire sample space. For example,

Rule (Law of Total Probability: 3-events)**.** *If E_1, E_2, and E_3 are disjoint and* $\Pr[E_1 \cup E_2 \cup E_3] = 1$*, then*

$$\Pr[A] = \Pr\left[A \mid E_1\right] \cdot \Pr[E_1] + \Pr\left[A \mid E_2\right] \cdot \Pr[E_2] + \Pr\left[A \mid E_3\right] \cdot \Pr[E_3].$$

This in turn leads to a three-event version of Bayes' Rule in which the probability of event E_1 given A is calculated from the "inverse" conditional probabilities of A given E_1, E_2, and E_3:

Rule (Bayes' Rule: 3-events)**.**

$$\Pr\left[E_1 \mid A\right] = \frac{\Pr\left[A \mid E_1\right] \cdot \Pr[E_1]}{\Pr\left[A \mid E_1\right] \cdot \Pr[E_1] + \Pr\left[A \mid E_2\right] \cdot \Pr[E_2] + \Pr\left[A \mid E_3\right] \cdot \Pr[E_3]}$$

The generalization of these rules to n disjoint events is a routine exercise (Problems 17.3 and 17.4).

17.5.1 Conditioning on a Single Event

The probability rules that we derived in Section 16.5.2 extend to probabilities conditioned on the same event. For example, the Inclusion-Exclusion formula for two sets holds when all probabilities are conditioned on an event C:

$$\Pr\left[A \cup B \mid C\right] = \Pr\left[A \mid C\right] + \Pr\left[B \mid C\right] - \Pr\left[A \cap B \mid C\right].$$

This is easy to verify by plugging in the Definition 17.2.1 of conditional probability.[2]

It is important not to mix up events before and after the conditioning bar. For example, the following is *not* a valid identity:

False Claim.

$$\Pr\left[A \mid B \cup C\right] = \Pr\left[A \mid B\right] + \Pr\left[A \mid C\right] - \Pr\left[A \mid B \cap C\right]. \qquad (17.3)$$

A simple counter-example is to let B and C be events over a uniform space with most of their outcomes in A, but not overlapping. This ensures that $\Pr\left[A \mid B\right]$ and $\Pr\left[A \mid C\right]$ are both close to 1. For example,

$$B ::= [0..9],$$
$$C ::= [10..18] \cup \{0\},$$
$$A ::= [1..18],$$

---

[2]Problem 17.14 explains why this and similar conditional identities follow on general principles from the corresponding unconditional identities.

so
$$\Pr[A \mid B] = \frac{9}{10} = \Pr[A \mid C].$$

Also, since 0 is the only outcome in $B \cap C$ and $0 \notin A$, we have

$$\Pr[A \mid B \cap C] = 0$$

So the right hand side of (17.3) is 1.8, while the left hand side is a probability which can be at most 1—actually, it is 18/19.

6.042J / 18.062J Mathematics for Computer Science
Spring 2015