18 Conditional Probability

18.1 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 17.3 from Chapter 17 to capture this situation. There are exactly three outcomes where contestant chooses door A, and there is a goat behind door B:

$$(A, A, B),\ (A, A, C),\ (C, A, B). \tag{18.1}$$

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

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

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

18.1.1 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 (18.1). In the opening of Section 18.1, 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.

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

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.

18.2 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 18.2.1. 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 18.2.1 is very simple and causes no trouble—provided it is properly applied.

18.2.1 What went wrong

So if everything in the opening Section 18.1 is mathematically sound, why does it seem to contradict the results that we established in Chapter 17? 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:

$$\begin{aligned}
&\Pr\big[[\text{win by switching}] \mid [\text{pick A AND Carol opens B}]\big] \\
&= \Pr\big[(C, A, B) \mid \{(C, A, B), (A, A, B)\}\big] + \frac{\Pr[(C, A, B)]}{\Pr[\{(C, A, B), (A, A, B)\}]} \\
&= \frac{1/9}{1/9 + 1/18} = \frac{2}{3},
\end{aligned}$$

which is exactly what we already deduced from the tree diagram 17.2 in Section 17.2.

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 individual. 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 prejudice the jury against Simpson but would lack any probitive value, so the discussion should be barred.

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 issue we want to illustrate is that conditional probability is used and misused all the time, and even experts under public scrutiny make mistakes.

18.3 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 $\Pr[A \mid B]$.

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 18.1.

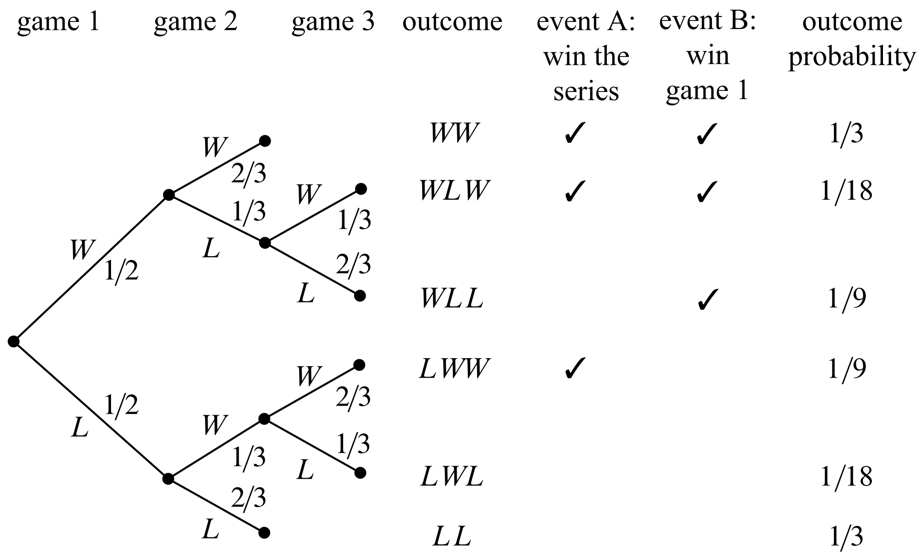


Figure 18.1 The tree diagram for computing the probability that the local team wins two out of three games given that they won the first game.

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 18.1.

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:

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

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

18.4 Why Tree Diagrams Work

We've now settled into a routine of solving probability problems using tree diagrams, but we have not really explained why they work. The explanation is that the probabilities that we've been recording on the edges of tree diagrams are actually conditional probabilities.

For example, look at 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 18.2.1 of conditional probability which could be written in a form called the *Product Rule* for conditional probabilities:

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

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

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

$$\begin{aligned}
&\Pr[\text{win first game} \cap \text{win second game}] \\
&\qquad = \Pr[\text{win first game}] \cdot \Pr\left[\text{win second game} \mid \text{win first game}\right] \\
&\qquad = \frac{1}{2} \cdot \frac{2}{3}.
\end{aligned}$$

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)**.**

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

An n-event version of the Rule is given in Problem 18.1, but its form should be clear from the three event version.

18.4.1 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

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

18.4.2 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 naive intuitive expectation. 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."

18.4.3 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 $\Pr[A \mid B]$, 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 18.2.

Step 4: Compute Event Probabilities
From Definition 18.2.1, we have

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

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

[1] The statistics in this example are roughly based on actual medical data, but have been altered somewhat to simplify the calculations.

Step 1: Find the Sample Space

The sample space is found with the tree diagram in Figure 18.2.

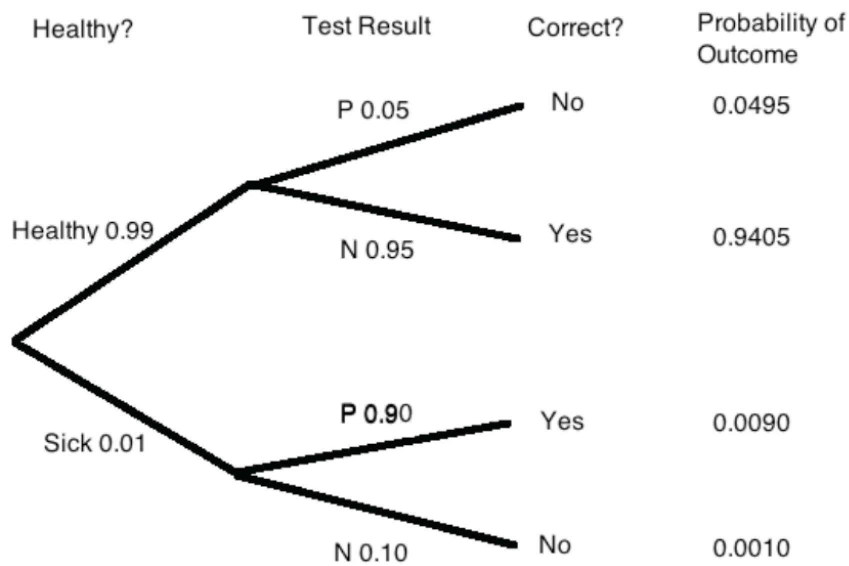


Figure 18.2 The tree diagram for a breast cancer test.

test doesn't tell us much. To see why percent accuracy is no guarantee of value, 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.

18.4.4 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 positive 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.

18.4.5 *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 judgements 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 18.1:

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

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

Theorem 18.4.1 (Bayes' Rule)**.**

$$\Pr\left[B \mid A\right] = \frac{\Pr\left[A \mid B\right] \cdot \Pr[B]}{\Pr[A]} \tag{18.2}$$

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 (18.2). ■

18.4.6 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 18.4.5 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 9.5.1 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, 100 times. So if N actually is composite, then the probability that 100 repetitions of the probabilistic test do not discover this is at most:

$$\left(\frac{1}{4}\right)^{100}.$$

If the test remained inconclusive after 100 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 18.9.

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.

18.5 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 18.5.1 (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*

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

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)**.**

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

The generalization of these rules to n disjoint events is a routine exercise (Problems 18.3 and 18.4).

18.5.1 Conditioning on a Single Event

The probability rules that we derived in Section 17.5.2 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:

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

This is easy to verify by plugging in the Definition 18.2.1 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.

$$\Pr[A \mid B \cup C] = \Pr[A \mid B] + \Pr[A \mid C] - \Pr[A \mid B \cap C]. \tag{18.3}$$

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 $\Pr[A \mid B]$ and $\Pr[A \mid C]$ are both close to 1. For example,

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

[2]Problem 18.14 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 (18.3) is 1.8, while the left-hand side is a probability which can be at most 1—actually, it is 18/19.

18.6 Simpson's Paradox

In 1973, a famous university was investigated for gender discrimination [7]. The investigation was prompted by evidence that, at first glance, appeared definitive: in 1973, 44% of male applicants to the school's graduate programs were accepted, but only 35% of female applicants were admitted.

However, this data turned out to be completely misleading. Analysis of the individual departments, showed not only that few showed significant evidence of bias, but also that among the few departments that *did* show statistical irregularities, most were slanted *in favor of women*. This suggests that if there was any sex discrimination, then it was against men!

Given the discrepancy in these findings, it feels like someone must be doing bad math—intentionally or otherwise. But the numbers are not actually inconsistent. In fact, this statistical hiccup is common enough to merit its own name: *Simpson's Paradox* occurs when multiple small groups of data all exhibit a similar trend, but that trend reverses when those groups are aggregated. To explain how this is possible, let's first clarify the problem by expressing both arguments in terms of conditional probabilities. For simplicity, suppose that there are only two departments EE and CS. Consider the experiment where we pick a random candidate. Define the following events:

- $A ::=$ the candidate is admitted to his or her program of choice,
- $F_{EE} ::=$ the candidate is a woman applying to the EE department,
- $F_{CS} ::=$ the candidate is a woman applying to the CS department,
- $M_{EE} ::=$ the candidate is a man applying to the EE department,
- $M_{CS} ::=$ the candidate is a man applying to the CS department.

CS	2 men admitted out of 5 candidates	40%
	50 women admitted out of 100 candidates	50%
EE	70 men admitted out of 100 candidates	70%
	4 women admitted out of 5 candidates	80%
Overall	72 men admitted, 105 candidates	$\approx$ 69%
	54 women admitted, 105 candidates	$\approx$ 51%

Table 18.1 Hypothetical admission statistics where men are overall more to be admitted, but are less likely to be admitted into each department.

Assume that all candidates are either men or women, and that no candidate belongs to both departments. That is, the events F_{EE}, F_{CS}, M_{EE} and M_{CS} are all disjoint.

In these terms, the plaintiff's assertion—that a male candidate is more likely to be admitted to the university than a female—can be expressed by the following inequality:

$$\Pr\left[A \mid M_{EE} \cup M_{CS}\right] > \Pr\left[A \mid F_{EE} \cup F_{CS}\right].$$

The university's retort that *in any given department*, a male applicant is less likely to be admitted than a female can be expressed by a pair of inequalities:

$$\begin{aligned} \Pr\left[A \mid M_{EE}\right] &< \Pr\left[A \mid F_{EE}\right] \quad \text{and} \\ \Pr\left[A \mid M_{CS}\right] &< \Pr\left[A \mid F_{CS}\right]. \end{aligned}$$

We can explain how there could be such a discrepancy between university-wide and department-by-department admission statistics by supposing that the CS department is more selective than the EE department, but CS attracts a far larger number of woman applicants than EE.[3]. Table 18.1 shows some admission statistics contrived to highlight how the inequalities asserted by both the plaintiff and the university could both hold.

Initially, we and the plaintiffs both assumed that the overall admissions statistics for the university could only be explained by gender discrimination. The department by department statistics seems to belie the accusation of discrimination. But do they really?

Suppose we replaced "the candidate is a man/woman applying to the EE department," by "the candidate is a man/woman for whom an admissions decision was made during an odd-numbered day of the month," and likewise with CS and an even-numbered day of the month. Since we don't think the parity of a date is a

[3]At the actual university in the lawsuit, the "exclusive" departments more popular among women were those that did not require a mathematical foundation, such as English and education. Women's disproportionate choice of these careers reflects gender bias, but one which predates the university's involvement.

cause for the outcome of an admission decision, we would most likely dismiss the “coincidence” that on both odd and even dates, women are more frequently admitted. Instead we would judge, based on the overall data showing women less likely to be admitted, that gender bias against women *was* an issue in the university.

Bear in mind that it would be the *same numerical data* that we would be using to justify our different conclusions in the department-by-department case and the even-day-odd-day case. We interpreted the same numbers differently based on our implicit causal beliefs, specifically that departments matter and date parity does not. It is circular to claim that the data corroborated our beliefs that there is or is not discrimination. Rather, our interpretation of the data correlation depended on our beliefs about the causes of admission in the first place.[4] This example highlights a basic principle in statistics that people constantly ignore: *never assume that correlation implies causation*.

18.7 Independence

Suppose that we flip two fair coins simultaneously on opposite sides of a room. Intuitively, the way one coin lands does not affect the way the other coin lands. The mathematical concept that captures this intuition is called *independence*.

Definition 18.7.1. An event with probability 0 is defined to be independent of every event (including itself). If $\Pr[B] \neq 0$, then event A is independent of event B iff

$$\Pr\left[A \mid B\right] = \Pr[A]. \tag{18.4}$$

In other words, A and B are independent if knowing that B happens does not alter the probability that A happens, as is the case with flipping two coins on opposite sides of a room.

Potential Pitfall

Students sometimes get the idea that disjoint events are independent. The *opposite* is true: if $A \cap B = \emptyset$, then knowing that A happens means you know that B does not happen. Disjoint events are *never* independent—unless one of them has probability zero.

[4] These issues are thoughtfully examined in *Causality: Models, Reasoning and Inference*, Judea Pearl, Cambridge U. Press, 2001.

18.7.1 Alternative Formulation

Sometimes it is useful to express independence in an alternate form which follows immediately from Definition 18.7.1:

Theorem 18.7.2. *A is independent of B if and only if*

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]. \tag{18.5}$$

Notice that Theorem 18.7.2 makes apparent the symmetry between A being independent of B and B being independent of A:

Corollary 18.7.3. *A is independent of B iff B is independent of A.*

18.7.2 Independence Is an Assumption

Generally, independence is something that you *assume* in modeling a phenomenon. For example, consider the experiment of flipping two fair coins. Let A be the event that the first coin comes up heads, and let B be the event that the second coin is heads. If we assume that A and B are independent, then the probability that both coins come up heads is:

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

In this example, the assumption of independence is reasonable. The result of one coin toss should have negligible impact on the outcome of the other coin toss. And if we were to repeat the experiment many times, we would be likely to have $A \cap B$ about 1/4 of the time.

On the other hand, there are many examples of events where assuming independence isn't justified. For example, an hourly weather forecast for a clear day might list a 10% chance of rain every hour from noon to midnight, meaning each hour has a 90% chance of being dry. But that does *not* imply that the odds of a rainless day are a mere $0.9^{12} \approx 0.28$. In reality, if it doesn't rain as of 5pm, the odds are higher than 90% that it will stay dry at 6pm as well—and if it starts pouring at 5pm, the chances are much higher than 10% that it will still be rainy an hour later.

Deciding when to *assume* that events are independent is a tricky business. In practice, there are strong motivations to assume independence since many useful formulas—such as equation (18.5)—only hold if the events are independent. But you need to be careful: we'll describe several famous examples where mistaken assumptions of independence led to trouble. This problem gets even trickier when there are more than two events in play.

18.8 Mutual Independence

We have defined what it means for two events to be independent. What if there are more than two events? For example, how can we say that the flips of n coins are all independent of one another? A set of events is said to be *mutually independent* if the probability of each event in the set is the same no matter which of the other events has occurred. This is equivalent to saying that for any selection of two or more of the events, the probability that all the selected events occur equals the product of the probabilities of the selected events.

For example, four events E_1, E_2, E_3, E_4 are mutually independent if and only if all of the following equations hold:

$$\begin{aligned}
\Pr[E_1 \cap E_2] &= \Pr[E_1] \cdot \Pr[E_2] \\
\Pr[E_1 \cap E_3] &= \Pr[E_1] \cdot \Pr[E_3] \\
\Pr[E_1 \cap E_4] &= \Pr[E_1] \cdot \Pr[E_4] \\
\Pr[E_2 \cap E_3] &= \Pr[E_2] \cdot \Pr[E_3] \\
\Pr[E_2 \cap E_4] &= \Pr[E_2] \cdot \Pr[E_4] \\
\Pr[E_3 \cap E_4] &= \Pr[E_3] \cdot \Pr[E_4] \\
\Pr[E_1 \cap E_2 \cap E_3] &= \Pr[E_1] \cdot \Pr[E_2] \cdot \Pr[E_3] \\
\Pr[E_1 \cap E_2 \cap E_4] &= \Pr[E_1] \cdot \Pr[E_2] \cdot \Pr[E_4] \\
\Pr[E_1 \cap E_3 \cap E_4] &= \Pr[E_1] \cdot \Pr[E_3] \cdot \Pr[E_4] \\
\Pr[E_2 \cap E_3 \cap E_4] &= \Pr[E_2] \cdot \Pr[E_3] \cdot \Pr[E_4] \\
\Pr[E_1 \cap E_2 \cap E_3 \cap E_4] &= \Pr[E_1] \cdot \Pr[E_2] \cdot \Pr[E_3] \cdot \Pr[E_4]
\end{aligned}$$

The generalization to mutual independence of n events should now be clear.

18.8.1 DNA Testing

Assumptions about independence are routinely made in practice. Frequently, such assumptions are quite reasonable. Sometimes, however, the reasonableness of an independence assumption is not so clear, and the consequences of a faulty assumption can be severe.

Let’s return to the O. J. Simpson murder trial. The following expert testimony was given on May 15, 1995:

Mr. Clarke: When you make these estimations of frequency—and I believe you touched a little bit on a concept called independence?

Dr. Cotton: Yes, I did.

Mr. Clarke: And what is that again?

Dr. Cotton: It means whether or not you inherit one allele that you have is not—does not affect the second allele that you might get. That is, if you inherit a band at 5,000 base pairs, that doesn't mean you'll automatically or with some probability inherit one at 6,000. What you inherit from one parent is [independent of] what you inherit from the other.

Mr. Clarke: Why is that important?

Dr. Cotton: Mathematically that's important because if that were not the case, it would be improper to multiply the frequencies between the different genetic locations.

Mr. Clarke: How do you—well, first of all, are these markers independent that you've described in your testing in this case?

Presumably, this dialogue was as confusing to you as it was for the jury. Essentially, the jury was told that genetic markers in blood found at the crime scene matched Simpson's. Furthermore, they were told that the probability that the markers would be found in a randomly-selected person was at most 1 in 170 million. This astronomical figure was derived from statistics such as:

- 1 person in 100 has marker A.
- 1 person in 50 marker B.
- 1 person in 40 has marker C.
- 1 person in 5 has marker D.
- 1 person in 170 has marker E.

Then these numbers were multiplied to give the probability that a randomly-selected person would have all five markers:

$$\begin{aligned}\Pr[A \cap B \cap C \cap D \cap E] &= \Pr[A] \cdot \Pr[B] \cdot \Pr[C] \cdot \Pr[D] \cdot \Pr[E] \\ &= \frac{1}{100} \cdot \frac{1}{50} \cdot \frac{1}{40} \cdot \frac{1}{5} \cdot \frac{1}{170} = \frac{1}{170{,}000{,}000}.\end{aligned}$$

The defense pointed out that this assumes that the markers appear mutually independently. Furthermore, all the statistics were based on just a few hundred blood samples.

After the trial, the jury was widely mocked for failing to "understand" the DNA evidence. If you were a juror, would *you* accept the 1 in 170 million calculation?

18.8.2 Pairwise Independence

The definition of mutual independence seems awfully complicated—there are so many selections of events to consider! Here's an example that illustrates the subtlety of independence when more than two events are involved. Suppose that we flip three fair, mutually-independent coins. Define the following events:

- A_1 is the event that coin 1 matches coin 2.
- A_2 is the event that coin 2 matches coin 3.
- A_3 is the event that coin 3 matches coin 1.

Are A_1, A_2, A_3 mutually independent?

The sample space for this experiment is:

$$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Every outcome has probability $(1/2)^3 = 1/8$ by our assumption that the coins are mutually independent.

To see if events A_1, A_2 and A_3 are mutually independent, we must check a sequence of equalities. It will be helpful first to compute the probability of each event A_i:

$$\begin{aligned}\Pr[A_1] &= \Pr[HHH] + \Pr[HHT] + \Pr[TTH] + \Pr[TTT] \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}.\end{aligned}$$

By symmetry, $\Pr[A_2] = \Pr[A_3] = 1/2$ as well. Now we can begin checking all the equalities required for mutual independence:

$$\begin{aligned}\Pr[A_1 \cap A_2] &= \Pr[HHH] + \Pr[TTT] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} \\ &= \Pr[A_1]\Pr[A_2].\end{aligned}$$

By symmetry, $\Pr[A_1 \cap A_3] = \Pr[A_1] \cdot \Pr[A_3]$ and $\Pr[A_2 \cap A_3] = \Pr[A_2] \cdot \Pr[A_3]$ must hold also. Finally, we must check one last condition:

$$\begin{aligned}\Pr[A_1 \cap A_2 \cap A_3] &= \Pr[HHH] + \Pr[TTT] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4} \\ &\neq \frac{1}{8} = \Pr[A_1]\Pr[A_2]\Pr[A_3].\end{aligned}$$

The three events A_1, A_2 and A_3 are not mutually independent even though any two of them are independent! This not-quite mutual independence seems weird at first, but it happens. It even generalizes:

Definition 18.8.1. A set $A_1, A_2, \ldots,$ of events is *k-way independent* iff every set of k of these events is mutually independent. The set is *pairwise independent* iff it is 2-way independent.

So the events A_1, A_2, A_3 above are pairwise independent, but not mutually independent. Pairwise independence is a much weaker property than mutual independence.

For example, suppose that the prosecutors in the O. J. Simpson trial were wrong and markers A, B, C, D and E are only *pairwise* independently. Then the probability that a randomly-selected person has all five markers is no more than:

$$\begin{aligned}\Pr[A \cap B \cap C \cap D \cap E] &\leq \Pr[A \cap E] = \Pr[A] \cdot \Pr[E] \\ &= \frac{1}{100} \cdot \frac{1}{170} = \frac{1}{17{,}000}.\end{aligned}$$

The first line uses the fact that $A \cap B \cap C \cap D \cap E$ is a subset of $A \cap E$. (We picked out the A and E markers because they're the rarest.) We use pairwise independence on the second line. Now the probability of a random match is 1 in 17,000—a far cry from 1 in 170 million! And this is the strongest conclusion we can reach assuming only pairwise independence.

On the other hand, the 1 in 17,000 bound that we get by assuming pairwise independence is a lot better than the bound that we would have if there were no independence at all. For example, if the markers are dependent, then it is possible that

everyone with marker E has marker A,

everyone with marker A has marker B,

everyone with marker B has marker C, and

everyone with marker C has marker D.

In such a scenario, the probability of a match is

$$\Pr[E] = \frac{1}{170}.$$

So a stronger independence assumption leads to a smaller bound on the probability of a match. The trick is to figure out what independence assumption is reasonable. Assuming that the markers are *mutually* independent may well *not* be reasonable unless you have examined hundreds of millions of blood samples. Otherwise, how would you know that marker D does not show up more frequently whenever the other four markers are simultaneously present?

18.9 Probability versus Confidence

Let's look at some other problems like the breast cancer test of Section 18.4.2, but this time we'll use more extreme numbers to highlight some key issues.

18.9.1 Testing for Tuberculosis

Let's suppose we have a really terrific diagnostic test for tuberculosis (TB): if you have TB, the test is *guaranteed* to detect it, and if you don't have TB, then the test will report that correctly 99% of the time!

In other words, let "*TB*" be the event that a person has TB, "*pos*" be the event that the person tests positive for TB, so "$\overline{pos}$" is the event that they test negative. Now we can restate these guarantees in terms of conditional probabilities:

$$\Pr\left[pos \mid TB\right] = 1, \tag{18.6}$$

$$\Pr\left[\overline{pos} \mid \overline{TB}\right] = 0.99. \tag{18.7}$$

This means that the test produces the correct result at least 99% of the time, regardless of whether or not the person has TB. A careful statistician would assert:[5]

Lemma. *You can be 99%* confident *that the test result is correct.*

Corollary 18.9.1. *If you test positive, then*

> ***either** you have TB **or** something very unlikely (probability 1/100) happened.*

Lemma 18.9.1 and Corollary 18.9.1 may *seem* to be saying that

False Claim. *If you test positive, then the probability that you have TB is* 0.99*.*

But this would be a mistake.

To highlight the difference between confidence in the test diagnosis versus the probability of TB, let's think about what to do if you test positive. Corollary 18.9.1

[5] Confidence is usually used to describe the probability that a statistical estimations of some quantity is correct (Section 20.4.3). We are trying to simplify the discussion by using this one concept to illustrate standard approaches to both hypothesis testing and estimation.

In the context of hypothesis testing, statisticians would normally distinguish the "false positive" probability, in this case the probability 0.01 that a healthy person is incorrectly diagnosed as having TB, and call this the *significance* of the test. The "false negative" probability would be the probability that person with TB is incorrectly diagnosed as healthy; it is zero. The *power* of the test is one minus the false negative probability, so in this case the power is the highest possible, namely, one.

seems to suggest that it's worth betting with high odds that you have TB, because it makes sense to bet against something unlikely happening—like the test being wrong. But having TB actually turns out to be *a lot less likely* than the test being wrong. So the either-or of Corollary 18.9.1 is really an either-or between something happening that is *extremely* unlikely—having TB—and something that is only *very* unlikely—the diagnosis being wrong. You're better off betting against the extremely unlikely event: it is better to bet the diagnosis is wrong.

So some knowledge of the probability of having TB is needed in order to figure out how seriously to take a positive diagnosis, even when the diagnosis is given with what seems like a high level of confidence. We can see exactly how the frequency of TB in a population influences the importance of a positive diagnosis by actually calculating the probability that someone who tests positive has TB. That is, we want to calculate $\Pr\left[TB \mid pos\right]$, which we do next.

18.9.2 Updating the Odds

Bayesian Updating

A standard way to convert the test probabilities into outcome probabilities is to use Bayes Theorem (18.2). It will be helpful to rephrase Bayes Theorem in terms of "odds" instead of probabilities.

If H is an event, we define the *odds* of H to be

$$\text{Odds}(H) ::= \frac{\Pr[H]}{\Pr[\overline{H}]} = \frac{\Pr[H]}{1 - \Pr[H]}.$$

For example, if H is the event of rolling a four using a fair, six-sided die, then

$$\Pr[\text{roll four}] = 1/6, \text{ so}$$
$$\text{Odds}(\text{roll four}) = \frac{1/6}{5/6} = \frac{1}{5}.$$

A gambler would say the odds of rolling a four were "one to five," or equivalently, "five to one *against*" rolling a four.

Odds are just another way to talk about probabilities. For example, saying the odds that a horse will win a race are "three to one against" means that a winning \$1 bet will return \$3 plus the \$1 bet initially. Three to one against winnning is the same as odds of one to three that the horse will win, which means the horse will win with probability $1/4$. In general,

$$\Pr[H] = \frac{\text{Odds}(H)}{1 + \text{Odds}(H)}.$$

Now suppose an event E offers some evidence about H. We now want to find the conditional probability of H given E. We can just as well find the odds of H given E,

$$\begin{aligned}
\text{Odds}(H \mid E) &::= \frac{\Pr[H \mid E]}{\Pr[\overline{H} \mid E]} \\
&= \frac{\Pr[E \mid H]\Pr[H]/\Pr[E]}{\Pr[E \mid \overline{H}]\Pr[\overline{H}]/\Pr[E]} && \text{(Bayes Theorem)} \\
&= \frac{\Pr[E \mid H]}{\Pr[E \mid \overline{H}]} \cdot \frac{\Pr[H]}{\Pr[\overline{H}]} \\
&= \text{Bayes-factor}(E, H) \cdot \text{Odds}(H),
\end{aligned}$$

where

$$\text{Bayes-factor}(E, H) ::= \frac{\Pr[E \mid H]}{\Pr[E \mid \overline{H}]}.$$

So to update the odds of H given the evidence E, we just multiply by Bayes Factor:

Lemma 18.9.2.

$$\text{Odds}(H \mid E) = \text{Bayes-factor}(E, H) \cdot \text{Odds}(H).$$

Odds for the TB test

The probabilities of test outcomes given in (18.6) and (18.7) are exactly what we need to find Bayes factor for the TB test:

$$\begin{aligned}
\text{Bayes-factor}(pos, TB) &= \frac{\Pr[pos \mid TB]}{\Pr[pos \mid \overline{TB}]} \\
&= \frac{1}{1 - \Pr[\overline{pos} \mid \overline{TB}]} \\
&= \frac{1}{1 - 0.99} = 100.
\end{aligned}$$

So testing positive for TB increases the odds you have TB by a factor of 100, which means a positive test is significant evidence supporting a diagnosis of TB. That seems good to know. But Lemma 18.9.2 also makes it clear that when a random person tests positive, we still can’t determine the odds they have TB unless we know what are the *odds of their having TB in the first place*, so let’s examine that.

In 2011, the United States Center for Disease Control got reports of 11,000 cases of TB in US. We can estimate that there were actually about 30,000 cases of TB that year, since it seems that only about one third of actual cases of TB get reported. The US population is a little over 300 million, which means

$$\Pr[TB] \approx \frac{30,000}{300,000,000} = \frac{1}{10,000}.$$

So the odds of TB are $1/9999$. Therefore,

$$\text{Odds}(TB \mid pos) = 100 \cdot \frac{1}{9,999} \approx \frac{1}{100}.$$

In other words, even if someone tests positive for TB at the 99% confidence level, the odds remain about 100 to one *against* their having TB. The 99% confidence level is not nearly high enough to overcome the relatively tiny probability of having TB.

18.9.3 Facts that are Probably True

We have figured out that if a random person tests positive for TB, the probability they have TB is about 1/100. Now if you personally happened to test positive for TB, a competent doctor typically would tell you that the probability that you have TB has risen from 1/10,000 to 1/100. But has it? Not really.

Your doctor should have not have been talking in this way about your particular situation. He should just have stuck to the statement that for *randomly chosen* people, the positive test would be right only one percent of the time. But you are not a random person, and whether or not you have TB is a fact about reality. The truth about your having TB may be *unknown* to your doctor and you, but that does not mean it has some probability of being true. It is either true or false, we just don't know which.

In fact, if you were worried about a 1/100 probability of having this serious disease, you could use additional information about yourself to change this probability. For example, native born residents of the US are about half as likely to have TB as foreign born residents. So if you are native born, "your" probability of having TB halves. Conversely, TB is twenty-five times more frequent among native born Asian/Pacific Islanders than native born Caucasions. So your probability of TB would increase dramatically if your family was from an Asian/Pacific Island.

The point is that the probability of having TB that your doctor reports to you depends on the probability of TB for a random person whom the doctor thinks is *like you*. The doctor has made a judgment about you based, for example, on what personal factors he considers relevant to getting TB, or how serious he thinks the consequences of a mistaken diagnosis would be. These are important medical judgments, but they are not mathematical. Different doctors will make different judgments about who is like you, and they will report differing probabilities. There is no "true" model of who you are, and there is no true individual probability of your having TB.

18.9.4 Extreme events

By definition, flipping a *fair* coin is equally likely to come up Heads or Tails. Now suppose you flip a fair coin one hundred times and get a Head every time. What do you think the odds are that the next flip will also be a Head?

If we make the usual assumption that the coin remains fair after one hundred flips, then by definition the official answer is that a Tail on the next flip is just as likely as another Head. But this belies what any sensible person would do, which is to bet heavily on the next flip being another Head.

How to make sense of this? To begin, let's recognize how absurd it is to wonder about what happens after one hundred flips of a fair coin all come up Heads, because the probability of this happening is unimaginably tiny. For example, the probability that just *fifty* flips of a fair coin come up Heads is 2^{-50}. We can try to make some sense of how small this number is with the observation that 2^{-50} is about equal to the probability that you will be struck by lightning while reading this paragraph. Ain't gonna happen.

The negligible probability that the first fifty flips of a fair coin will all be Heads, let alone that the first one hundred are Heads, simply undermines the credibility of the assumption that the coin is fair. Despite being told the coin is fair, we can't help but consider at least some remote possibility of a mistake: somehow the coin being flipped was not fair but rather was one that had a reasonable chance of flipping one hundred Heads. For example, if a biased coin had probability 0.99 of flipping a Head, then one hundred independent tosses will all come up Heads with probability a little more than one third.

So let's suppose that there are two coins, a fair one and a 0.99 biased one. One of these coins is randomly chosen, with the fair coin hugely favored: the biased coin will be chosen only with the extremely small "struck-by-lightning" probability 2^{-50}. The chosen coin is then flipped one hundred times.

Let E be the event of flipping one hundred heads and H be the event that the chosen coin is the biased one. Now

$$\begin{aligned}
\text{Odds}(H) &= \frac{2^{-50}}{1-2^{-50}} \approx 2^{-50},\\
\text{Bayes-factor}(E,H) &= \frac{\Pr[E \mid H]}{\Pr[E \mid \overline{H}]} = \frac{(99/100)^{100}}{2^{-100}} > 0.36 \cdot 2^{100},\\
\text{Odds}(H \mid E) &= \text{Bayes-factor}(E,H) \cdot \text{Odds}(H)\\
&> 0.36 \cdot 2^{100} \cdot 2^{-50} = 0.36 \cdot 2^{50}.
\end{aligned}$$

So after flipping one hundred Heads, the odds that the biased coin was chosen are overwhelming, namely more than $0.36 \cdot 2^{50}$, in which case the probability that the next flip will be a Head is 0.99. In other words, by assuming some tiny probability that the tossed coin was indeed mistakenly biased toward Heads, we can justify our intuition that after one hundred consecutive Heads, the next flip is very likely to be a Head.

Making an assumption about the probability that some unverified fact is true is known as the *Bayesian* approach to a hypothesis testing problem. By granting a tiny probability that the coin being flipped is biased, the Bayesian approach provides a reasonable justification for estimating that the odds of a Head on the next flip are ninety-nine to one in favor.

18.9.5 Confidence in the Next Flip

If we stick to confidence rather than probability, we don't need to make any Bayesian assumptions about the probability of a fair coin being chosen. We know that if one hundred Heads are flipped, then either the coin is biased, or else the fair coin produced one hundred Heads—something that virtually never happens. This means that when all one hundred flips come up Heads, we can be virtually 100% confident that the coin is biased and therefore 99% confident that the next flip will be a Head.

Problems for Section 18.4

Homework Problems

Problem 18.1.
The Conditional Probability Product Rule for n Events is

Rule.

$$\Pr[E_1 \cap E_2 \cap \ldots \cap E_n] = \Pr[E_1] \cdot \Pr\left[E_2 \mid E_1\right] \cdot \Pr\left[E_3 \mid E_1 \cap E_2\right] \cdots \\ \cdot \Pr\left[E_n \mid E_1 \cap E_2 \cap \ldots \cap E_{n-1}\right].$$

(a) Restate the Rule without elipses—that is, without "…"—by using "$\bigcap_{i=a}^{b}$" for suitable a, b.

(b) Prove it by induction.

Problems for Section 18.5

Practice Problems

Problem 18.2.
Dirty Harry places two bullets in random chambers of the six-bullet cylinder of his revolver. He gives the cylinder a random spin and says "Feeling lucky?" as he holds the gun against your heart.

(a) What is the probability that you will get shot if he pulls the trigger?

(b) Suppose he pulls the trigger and you don't get shot. What is the probability that you will get shot if he pulls the trigger a second time?

(c) Suppose you noticed that he placed the two shells next to each other in the cylinder. How does this change the answers to the previous two questions?

Problem 18.3.
State and prove a version of the Law of Total Probability that applies to disjoint events $E_1, \ldots, E_n$ whose union is the whole sample space.

Problem 18.4.
State and prove a version of Bayes Rule that applies to disjoint events $E_1, \ldots, E_n$ whose union is the whole sample space. You may assume the n-event Law of Total Probability, Problem 18.3.

Class Problems

Problem 18.5.
There are two decks of cards. One is complete, but the other is missing the ace of spades. Suppose you pick one of the two decks with equal probability and then select a card from that deck uniformly at random. What is the probability that you picked the complete deck, given that you selected the eight of hearts? Use the four-step method and a tree diagram.

Problem 18.6.
Suppose you have three cards: A♡, A♠ and a jack. From these, you choose a random hand (that is, each card is equally likely to be chosen) of two cards, and let n be the number of aces in your hand. You then randomly pick one of the cards in the hand and reveal it.

(a) Describe a simple probability space (that is, outcomes and their probabilities) for this scenario, and list the outcomes in each of the following events:

1. $[n \geq 1]$, (that is, your hand has an ace in it),
2. A♡ is in your hand,
3. the revealed card is an A♡,
4. the revealed card is an ace.

(b) Then calculate $\Pr\left[n = 2 \mid E\right]$ for E equal to each of the four events in part (a). Notice that most, but *not all*, of these probabilities are equal.

Now suppose you have a deck with d distinct cards, a different kinds of aces (including an A♡), you draw a random hand with h cards, and then reveal a random card from your hand.

(c) Prove that $\Pr[\text{A}\heartsuit \text{ is in your hand}] = h/d$.

(d) Prove that

$$\Pr\left[n = 2 \mid \text{A}\heartsuit \text{ is in your hand}\right] = \Pr[n = 2] \cdot \frac{2d}{ah}. \tag{18.8}$$

(e) Conclude that

$$\Pr\left[n = 2 \mid \text{the revealed card is an ace}\right] = \Pr\left[n = 2 \mid \text{A}\heartsuit \text{ is in your hand}\right].$$

Problem 18.7.
A fair six-sided die is repeatedly thrown until a six appears.

A natural sample space modelling this situation is the set of finite strings of integers from one to six that end at the first occurrence of a six. That is, $\mathcal{S} ::= [1..5]^*6$.

For example, 256 is the outcome corresponding to successively throwing a two, a five and a six. The length-one string $6 \in \mathcal{S}$ is the outcome corresponding to six appearing on the first throw.

(a) What should Pr[256] be defined to be? ...Pr[6]? What about the probability of an arbitrary outcome $s \in \mathcal{S}$?

(b) Verify that $\mathcal{S}$ with the probabilities assigned in part (a) defines a probability space. What does this imply about the possibility of never throwing a six?

For any string $r \in [1..5]^*$, let F_r be the event that values of the initial throws are are the successive elements of r. Let V be the event that all the dice throws are eVen. That is, V is the event $\{2,4\}^*6$ that all throws are twos and fours until the first six.

(c) Suppose t is a string of twos and fours, that is, $t \in \{2,4\}^*$. Explain why

$$\Pr\left[V \mid F_t\right] = \Pr[V]. \tag{18.9}$$

(d) Explain why equation (18.9) implies that for $t \in \{2,4\}^*$,

$$\Pr\left[F_t \mid V\right] = \Pr[F_t].$$

Conclude that

$$\Pr\left[6 \mid V\right] = \frac{2}{3}, \tag{18.10}$$

(e) Given that all throws are even, the only possible first throws are two, four and six. Since the die is fair, these are all equally likely, so the probability $\Pr\left[6 \mid V\right]$ that the first throw is a six must be 1/3, contradicting equation (18.10)! Explain.[6]

(f) Conclude immediately from (18.10) that

$$\Pr[V] = \frac{1}{4}. \tag{18.11}$$

---

[6] If you're thrown by this, you are not alone. There are several websites devoted to explanations of this seductive problem. In fact, when it came up at the MIT Theory of Computation faculty lunch in April 2018, several attendees confidently defended this mistaken reasoning.

Problem 18.8.
There are three prisoners in a maximum-security prison for fictional villains: the Evil Wizard Voldemort, the Dark Lord Sauron, and Little Bunny Foo-Foo. The parole board has declared that it will release two of the three, chosen uniformly at random, but has not yet released their names. Naturally, Sauron figures that he will be released to his home in Mordor, where the shadows lie, with probability $2/3$.

A guard offers to tell Sauron the name of one of the other prisoners who will be released (either Voldemort or Foo-Foo). If the guard has a choice of naming either Voldemort or Foo-Foo (because both are to be released), he names one of the two with equal probability.

Sauron knows the guard to be a truthful fellow. However, Sauron declines this offer. He reasons that knowing what the guards says will reduce his chances, so he is better off not knowing. For example, if the guard says, "Little Bunny Foo-Foo will be released", then his own probability of release will drop to $1/2$ because he will then know that either he or Voldemort will also be released, and these two events are equally likely.

Dark Lord Sauron has made a typical mistake when reasoning about conditional probability. Using a tree diagram and the four-step method, **explain his mistake**. What is the probability that Sauron is released given that the guard says Foo-Foo is released?

Hint: Define the events S, F and "F" as follows:

$$\begin{aligned} \text{“}F\text{”} &= \text{Guard says Foo-Foo is released} \\ F &= \text{Foo-Foo is released} \\ S &= \text{Sauron is released} \end{aligned}$$

Problem 18.9.
Every Skywalker serves either the *light side* or the *dark side*.

- The first Skywalker serves the dark side.
- For $n \geq 2$, the n-th Skywalker serves the same side as the $(n-1)$-st Skywalker with probability $1/4$, and the opposite side with probability $3/4$.

Let d_n be the probability that the n-th Skywalker serves the dark side.

(a) Express d_n with a recurrence equation and sufficient base cases.

(b) Derive a simple expression for the generating function $D(x) ::= \sum_1^\infty d_n x^n$.

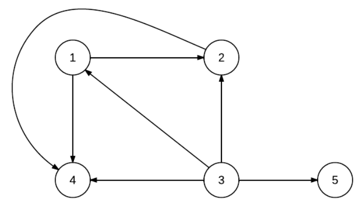


Figure 18.3 The DAG G_0

(c) Give a simple closed formula for d_n.

Problem 18.10. **(a)** For the directed acyclic graph (DAG) G_0 in Figure 18.3, a minimum-edge DAG with the same walk relation can be obtained by removing some edges. List these edges.

(b) List the vertices in a maximal chain in G_0.

Let G be the simple graph shown in Figure 18.4.

A directed graph $\overrightarrow{G}$ can be randomly constructed from G by assigning a direction to each edge independently with equal likelihood.

(c) What is the probability that $\overrightarrow{G} = G_0$?

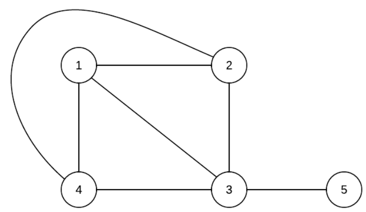


Figure 18.4 Simple graph G

Define the following events with respect to the random graph $\overrightarrow{G}$:

$$
\begin{aligned}
T_1 &::= \text{vertices } 2, 3, 4 \text{ are on a length three directed cycle,} \\
T_2 &::= \text{vertices } 1, 3, 4 \text{ are on a length three directed cycle,} \\
T_3 &::= \text{vertices } 1, 2, 4 \text{ are on a length three directed cycle,} \\
T_4 &::= \text{vertices } 1, 2, 3 \text{ are on a length three directed cycle.}
\end{aligned}
$$

(d) What are $\Pr[T_1], \Pr[T_1 \cap T_2], \Pr[T_1 \cap T_2 \cap T_3]$?

(e) $\overrightarrow{G}$ has the property that if it has a directed cycle, then it has a length three directed cycle. Use this fact to find the probability that $\overrightarrow{G}$ is a DAG.

Homework Problems

Problem 18.11.

There is a subject—naturally not *Math for Computer Science*—in which 10% of the assigned problems contain errors. If you ask a Teaching Assistant (TA) whether a problem has an error, then they will answer correctly 80% of the time, regardless of whether or not a problem has an error. If you ask a lecturer, he will identify whether or not there is an error with only 75% accuracy.

We formulate this as an experiment of choosing one problem randomly and asking a particular TA and Lecturer about it. Define the following events:

$$\begin{aligned}
E &::= [\text{the problem has an error}],\\
T &::= [\text{the TA says the problem has an error}],\\
L &::= [\text{the lecturer says the problem has an error}].
\end{aligned}$$

(a) Translate the description above into a precise set of equations involving conditional probabilities among the events E, T and L.

(b) Suppose you have doubts about a problem and ask a TA about it, and they tell you that the problem is correct. To double-check, you ask a lecturer, who says that the problem has an error. Assuming that the correctness of the lecturer's answer and the TA's answer are independent of each other, regardless of whether there is an error, what is the probability that there is an error in the problem?

(c) Is event T independent of event L (that is, $\Pr[T \mid L] = \Pr[T]$)? First, give an argument based on intuition, and then calculate both probabilities to verify your intuition.

Problem 18.12.
Suppose you repeatedly flip a fair coin until you see the sequence HTT or HHT. What is the probability you see the sequence HTT first?

Hint: Try to find the probability that HHT comes before HTT conditioning on whether you first toss an H or a T. The answer is not $1/2$.

Problem 18.13.
A 52-card deck is thoroughly shuffled and you are dealt a hand of 13 cards.

(a) If you have one ace, what is the probability that you have a second ace?

(b) If you have the ace of spades, what is the probability that you have a second ace? Remarkably, the answer is different from part (a).

Problem 18.14.
Suppose $\Pr[\cdot] : \mathcal{S} \to [0, 1]$ is a probability function on a sample space $\mathcal{S}$ and let B be an event such that $\Pr[B] > 0$. Define a function $\Pr_B[\cdot]$ on outcomes $\omega \in \mathcal{S}$ by the rule:

$$\Pr_B[\omega] ::= \begin{cases} \Pr[\omega]/\Pr[B] & \text{if } \omega \in B, \\ 0 & \text{if } \omega \notin B. \end{cases} \tag{18.12}$$

(a) Prove that $\Pr_B[\cdot]$ is also a probability function on $\mathcal{S}$ according to Definition 17.5.2.

(b) Prove that

$$\Pr_B[A] = \frac{\Pr[A \cap B]}{\Pr[B]}$$

for all $A \subseteq \mathcal{S}$.

(c) Explain why the Disjoint Sum Rule carries over for conditional probabilities, namely,

$$\Pr\left[C \cup D \mid B\right] = \Pr\left[C \mid B\right] + \Pr\left[D \mid B\right] \qquad (C, D \text{ disjoint}).$$

Give examples of several further such rules.

Problem 18.15.
Professor Meyer has a deck of 52 randomly shuffled playing cards, 26 red, 26 black. He proposes the following game: he will repeatedly draw a card off the top of the deck and turn it face up so that you can see it. At any point while there are still cards left in the deck, you may choose to stop, and he will turn over the next card. If the turned up card is black you win, and otherwise you lose. Either way, the game ends.

Suppose that after drawing off some top cards without stopping, the deck is left with r red cards and b black cards.

(a) Show that if you choose to stop at this point, the probability of winning is $b/(r+b)$.

(b) Prove if you choose *not* to stop at this point, the probability of winning is still $b/(r+b)$, regardless of your stopping strategy for the rest of the game.

Hint: Induction on $r+b$.

Exam Problems

Problem 18.16.
Sally Smart just graduated from high school. She was accepted to three reputable colleges.

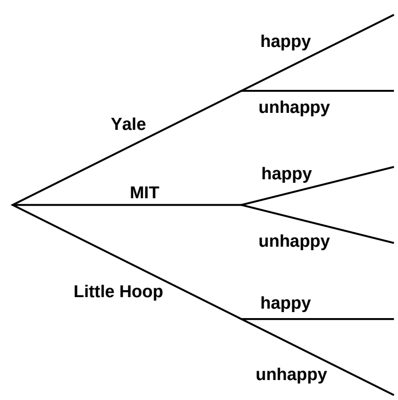
- With probability $4/12$, she attends Yale.
- With probability $5/12$, she attends MIT.

- With probability $3/12$, she attends Little Hoop Community College.

Sally is either happy or unhappy in college.

- If she attends Yale, she is happy with probability $4/12$.
- If she attends MIT, she is happy with probability $7/12$.
- If she attends Little Hoop, she is happy with probability $11/12$.

(a) A tree diagram to help Sally project her chance at happiness is shown below. On the diagram, fill in the edge probabilities, and at each leaf write the probability of the corresponding outcome.



(b) What is the probability that Sally is happy in college?

(c) What is the probability that Sally attends Yale, given that she is happy in college?

(d) Show that the event that Sally attends Yale **is not** independent of the event that she is happy.

(e) Show that the event that Sally attends MIT **is** independent of the event that she is happy.

Problem 18.17.
Here's a variation of Monty Hall's game: the contestant still picks one of three doors, with a prize randomly placed behind one door and goats behind the other two. But now, instead of always opening a door to reveal a goat, Monty instructs

Carol to *randomly* open one of the two doors that the contestant hasn't picked. This means she may reveal a goat, or she may reveal the prize. If she reveals the prize, then the entire game is *restarted*, that is, the prize is again randomly placed behind some door, the contestant again picks a door, and so on until Carol finally picks a door with a goat behind it. Then the contestant can choose to *stick* with his original choice of door or *switch* to the other unopened door. He wins if the prize is behind the door he finally chooses.

To analyze this setup, we define two events:

GP**:** The event that the contestant **g**uesses the door with the **p**rize behind it on his first guess.

OP**:** The event that the game is restarted at least once. Another way to describe this is as the event that the door Carol first **o**pens has a **p**rize behind it.

Give the values of the following probabilities:

(a) $\Pr[GP]$

(b) $\Pr\left[OP \mid \overline{GP}\right]$

(c) $\Pr[OP]$

(d) the probability that the game will continue forever

(e) When Carol finally picks the goat, the contestant has the choice of sticking or switching. Let's say that the contestant adopts the strategy of sticking. Let W be the event that the contestant wins with this strategy, and let $w ::= \Pr[W]$. Express the following conditional probabilities as simple closed forms in terms of w.

i) $\Pr\left[W \mid GP\right]$

ii) $\Pr\left[W \mid \overline{GP} \cap OP\right]$

iii) $\Pr\left[W \mid \overline{GP} \cap \overline{OP}\right]$

(f) What is the value of $\Pr[W]$?

(g) For any final outcome where the contestant wins with a “stick” strategy, he would lose if he had used a “switch” strategy, and vice versa. In the original Monty Hall game, we concluded immediately that the probability that he would win with a “switch” strategy was $1 - \Pr[W]$. Why isn't this conclusion quite as obvious for this new, restartable game? Is this conclusion still sound? Briefly explain.

Problem 18.18.
There are two decks of cards, the red deck and the blue deck. They differ slightly in a way that makes drawing the eight of hearts slightly more likely from the red deck than from the blue deck.

One of the decks is randomly chosen and hidden in a box. You reach in the box and randomly pick a card that turns out to be the eight of hearts. You believe intuitively that this makes the red deck more likely to be in the box than the blue deck.

Your intuitive judgment about the red deck can be formalized and verified using some inequalities between probabilities and conditional probabilities involving the events

$$R ::= \textbf{R}\text{ed deck is in the box},$$
$$B ::= \textbf{B}\text{lue deck is in the box},$$
$$E ::= \textbf{E}\text{ight of hearts is picked from the deck in the box.}$$

(a) State an inequality between probabilities and/or conditional probabilities that formalizes the assertion, “picking the eight of hearts from the red deck is more likely than from the blue deck.”

(b) State a similar inequality that formalizes the assertion “picking the eight of hearts from the deck in the box makes the red deck more likely to be in the box than the blue deck.”

(c) Assuming that initially each deck is equally likely to be the one in the box, prove that the inequality of part (a) implies the inequality of part (b).

(d) Suppose you couldn’t be sure that the red deck and blue deck initially were equally likely to be in the box. Could you still conclude that picking the eight of hearts from the deck in the box makes the red deck more likely to be in the box than the blue deck? Briefly explain.

Problem 18.19.
A flip of Coin 1 is x times as likely to come up Heads as a flip of Coin 2. A biased random choice of one of these coins will be made, where the probability of choosing Coin 1 is w times that of Coin 2.

(a) Restate the information above as equations between conditional probabilities involving the events

$$\begin{aligned} C1 &::= \text{Coin 1 was chosen,} \\ C2 &::= \text{Coin 2 was chosen,} \\ H &::= \text{the chosen coin came up Heads.} \end{aligned}$$

(b) State an inequality involving conditional probabilities of the above events that formalizes the assertion "Given that the chosen coin came up Heads, the chosen coin is more likely to have been Coin 1 than Coin 2."

(c) Prove that, given that the chosen coin came up Heads, the chosen coin is more likely to have been Coin 1 than Coin 2 iff

$$wx > 1.$$

Problem 18.20.
There is an unpleasant, degenerative disease called Beaver Fever which causes people to tell math jokes unrelentingly in social settings, believing other people will think they're funny. Fortunately, Beaver Fever is rare, afflicting only about 1 in 1000 people. Doctor Meyer has a fairly reliable diagnostic test to determine who is going to suffer from this disease:

- If a person will suffer from Beaver Fever, the probability that Dr. Meyer diagnoses this is 0.99.
- If a person will not suffer from Beaver Fever, the probability that Dr. Meyer diagnoses this is 0.97.

Let B be the event that a randomly chosen person will suffer **B**eaver Fever, and Y be the event that Dr. Meyer's diagnosis is "**Y**es, this person will suffer from Beaver Fever," with $\overline{B}$ and $\overline{Y}$ being the complements of these events.

(a) The description above explicitly gives the values of the following quantities. What are their values?

$$\Pr[B] \qquad \Pr\left[Y \mid B\right] \qquad \Pr\left[\overline{Y} \mid \overline{B}\right]$$

(b) Write formulas for $\Pr[\overline{B}]$ and $\Pr\left[Y \mid \overline{B}\right]$ solely in terms of the explicitly given quantities in part (a)—literally use their expressions, not their numeric values.

(c) Write a formula for the probability that Dr. Meyer says a person will suffer from Beaver Fever solely in terms of $\Pr[B]$, $\Pr[\overline{B}]$, $\Pr\left[Y \mid B\right]$ and $\Pr\left[Y \mid \overline{B}\right]$.

(d) Write a formula solely in terms of the expressions given in part (a) for the probability that a person will suffer Beaver Fever given that Doctor Meyer says they will. Then calculate the numerical value of the formula.

Suppose there was a vaccine to prevent Beaver Fever, but the vaccine was expensive or slightly risky itself. If you were sure you were going to suffer from Beaver Fever, getting vaccinated would be worthwhile, but even if Dr. Meyer diagnosed you as a future sufferer of Beaver Fever, the probability you actually will suffer Beaver Fever remains low (about 1/32 by part (d)).

In this case, you might sensibly decide not to be vaccinated—after all, Beaver Fever is not *that* bad an affliction. So the diagnostic test serves no purpose in your case. You may as well not have bothered to get diagnosed. Even so, the test may be useful:

(e) Suppose Dr. Meyer had enough vaccine to treat 2% of the population. If he randomly chose people to vaccinate, he could expect to vaccinate only 2% of the people who needed it. But by testing everyone and only vaccinating those diagnosed as future sufferers, he can expect to vaccinate a much larger fraction people who were going to suffer from Beaver Fever. Estimate this fraction.

Problem 18.21.

Suppose that *Let's Make a Deal* is played according to slightly different rules and with a red goat and a blue goat. There are three doors, with a prize hidden behind one of them and one goat behind each of the others. No doors are opened until the contestant makes a final choice to stick or switch. The contestant is allowed to pick a door and ask a certain question that the host then answers honestly. The contestant may then stick with their chosen door, or switch to either of the other doors.

(a) If the contestant asks "is there is a goat behind one of the unchosen doors?" and the host answers "yes," is the contestant more likely to win the prize if they stick, switch, or does it not matter? Clearly identify the probability space of outcomes and their probabilities you use to model this situation. What is the contestant's probability of winning if he uses the best strategy?

(b) If the contestant asks "is the *red* goat behind one of the unchosen doors?" and the host answers "yes," is the contestant more likely to win the prize if they stick, switch, or does it not matter? Clearly identify the probability space of outcomes and their probabilities you use to model this situation. What is the contestant's probability of winning if he uses the best strategy?

Problem 18.22.
You are organizing a neighborhood census and instruct your census takers to knock on doors and note the gender of any child that answers the knock. Assume that there are two children in every household, that a random child is equally likely to be a girl or a boy, and that the two children in a household are equally likely to be the one that opens the door.

A sample space for this experiment has outcomes that are triples whose first element is either B or G for the gender of the elder child, whose second element is either B or G for the gender of the younger child, and whose third element is E or Y indicating whether the *e*lder child or *y*ounger child opened the door. For example, (B, G, Y) is the outcome that the elder child is a boy, the younger child is a girl, and the girl opened the door.

(a) Let O be the event that a girl opened the door, and let T be the event that the household has two girls. List the outcomes in O and the outcomes in T.

(b) What is the probability $\Pr[T \mid O]$, that both children are girls, given that a girl opened the door?

(c) What mistake is made in the following argument? (For example, perhaps there's an arithmetic mistake [where?], or an unjustified assumption [what?], or some other error entirely.) Please identify and explain the error in detail.

> If a girl opens the door, then we know that there is at least one girl in the household. The probability that there is at least one girl is
>
> $$1 - \Pr[\text{both children are boys}] = 1 - (1/2 \times 1/2) = 3/4.$$
>
> So,
>
> $$\begin{aligned}&\Pr[T \mid \text{there is at least one girl in the household}]\\ &= \frac{\Pr[T \cap \text{there is at least one girl in the household}]}{\Pr[\text{there is at least one girl in the household}]}\\ &= \frac{\Pr[T]}{\Pr[\text{there is at least one girl in the household}]}\\ &= (1/4)/(3/4) = 1/3.\end{aligned}$$
>
> Therefore, given that a girl opened the door, the probability that there are two girls in the household is 1/3.

Problem 18.23.
A guard is going to release exactly two of the three prisoners, Sauron, Voldemort, and Bunny Foo Foo, and he's equally likely to release any set of two prisoners.

(a) What is the probability that Voldemort will be released?

The guard will truthfully tell Voldemort the name of one of the prisoners to be released. We're interested in the following events:

V**:** Voldemort is released.

"F": The guard tells Voldemort that Foo Foo will be released.

"S": The guard tells Voldemort that Sauron will be released.

The guard has two rules for choosing whom he names:

- never say that Voldemort will be released,
- if both Foo Foo and Sauron are getting released, say "Foo Foo."

(b) What is $\Pr\left[V \mid \text{"}F\text{"}\right]$?

(c) What is $\Pr\left[V \mid \text{"}S\text{"}\right]$?

(d) Show how to use the Law of Total Probability to combine your answers to parts (b) and (c) to verify that the result matches the answer to part (a).

Problem 18.24.
We are interested in paths in the plane starting at $(0, 0)$ that go one unit right or one unit up at each step. To model this, we use a state machine whose states are $\mathbb{N} \times \mathbb{N}$, whose start state is $(0, 0)$, and whose transitions are

$$(x, y) \to (x + 1, y),$$
$$(x, y) \to (x, y + 1).$$

(a) How many length n paths are there starting from the origin?

(b) How many states are reachable in exactly n steps?

(c) How many states are reachable in at most n steps?

(d) If transitions occur independently at random, going right with probability p and up with probability $q ::= 1 - p$ at each step, what is the probability of reaching position (x, y)?

(e) What is the probability of reaching state (x, y) *given* that the path to (x, y) reached (m, n) before getting to (x, y)?

(f) Show that the probability that a path ending at (x, y) went through (m, n) is the same for all p.

Problems for Section 18.6

Practice Problems

Problem 18.25.
Define the events A, F_{EE}, F_{CS}, M_{EE}, and M_{CS} as in Section 18.6.

In these terms, the plaintiff in a discrimination suit against a university makes the argument that in both departments, the probability that a female is admitted is less than the probability for a male. That is,

$$\Pr\left[A \mid F_{EE}\right] < \Pr\left[A \mid M_{EE}\right] \quad \text{and} \tag{18.13}$$
$$\Pr\left[A \mid F_{CS}\right] < \Pr\left[A \mid M_{CS}\right]. \tag{18.14}$$

The university's defence attorneys retort that *overall*, a female applicant is *more* likely to be admitted than a male, namely, that

$$\Pr\left[A \mid F_{EE} \cup F_{CS}\right] > \Pr\left[A \mid M_{EE} \cup M_{CS}\right]. \tag{18.15}$$

The judge then interrupts the trial and calls the plaintiff and defence attorneys to a conference in his office to resolve what he thinks are contradictory statements of facts about the admission data. The judge points out that:

$$\begin{aligned}
&\Pr\left[A \mid F_{EE} \cup F_{CS}\right] \\
&= \Pr\left[A \mid F_{EE}\right] + \Pr\left[A \mid F_{CS}\right] && \text{(because } F_{EE} \text{ and } F_{CS} \text{ are disjoint)} \\
&< \Pr\left[A \mid M_{EE}\right] + \Pr\left[A \mid M_{CS}\right] && \text{(by (18.13) and (18.14))} \\
&= \Pr\left[A \mid M_{EE} \cup M_{CS}\right] && \text{(because } M_{EE} \text{ and } M_{CS} \text{ are disjoint)}
\end{aligned}$$

so

$$\Pr\left[A \mid F_{EE} \cup F_{CS}\right] < \Pr\left[A \mid M_{EE} \cup M_{CS}\right],$$

which directly contradicts the university's position (18.15)!

Of course the judge is mistaken; an example where the plaintiff and defence assertions are all true appears in Section 18.6. What is the mistake in the judge's proof?

Problems for Section 18.7

Practice Problems

Problem 18.26.
Outside of their hum-drum duties as Math for Computer Science Teaching Assistants, Oscar is trying to learn to levitate using only intense concentration and Liz is trying to become the world champion flaming torch juggler. Suppose that Oscar's probability of success is $1/6$, Liz's chance of success is $1/4$, and these two events are independent.

(a) If at least one of them succeeds, what is the probability that Oscar learns to levitate?

(b) If at most one of them succeeds, what is the probability that Liz becomes the world flaming torch juggler champion?

(c) If exactly one of them succeeds, what is the probability that it is Oscar?

Problem 18.27.
What is the smallest size sample space in which there are two independent events, neither of which has probability zero or probability one? Explain.

Problem 18.28.
Give examples of event A, B, E such that

(a) A and B are independent, and are also conditionally independent given E, but are not conditionally independent given $\overline{E}$. That is,

$$\Pr[A \cap B] = \Pr[A]\Pr[B],$$
$$\Pr\left[A \cap B \mid E\right] = \Pr\left[A \mid E\right]\Pr\left[B \mid E\right],$$
$$\Pr\left[A \cap B \mid \overline{E}\right] \neq \Pr\left[A \mid \overline{E}\right]\Pr\left[B \mid \overline{E}\right].$$

Hint: Let $\mathcal{S} = \{1, 2, 3, 4\}$.

(b) A and B are conditionally independent given E, or given $\overline{E}$, but are not independent. That is,

$$\Pr\left[A \cap B \mid E\right] = \Pr\left[A \mid E\right]\Pr\left[B \mid E\right],$$
$$\Pr\left[A \cap B \mid \overline{E}\right] = \Pr\left[A \mid \overline{E}\right]\Pr\left[B \mid \overline{E}\right],$$
$$\Pr[A \cap B] \neq \Pr[A]\Pr[B].$$

Hint: Let $\mathcal{S} = \{1, 2, 3, 4, 5\}$.

Class Problems

Problem 18.29.
Event E is *evidence in favor* of event H when $\Pr\left[H \mid E\right] > \Pr[H]$, and it is *evidence against* H when $\Pr\left[H \mid E\right] < \Pr[H]$.

(a) Give an example of events A, B, H such that A and B are independent, both are evidence for H, but $A \cup B$ is evidence against H.

Hint: Let $\mathcal{S} = [1..8]$

(b) Prove E is evidence in favor of H iff $\overline{E}$ is evidence against H.

Problem 18.30.
Let G be a simple graph with n vertices. Let “$A(u, v)$” mean that vertices u and v are adjacent, and let “$W(u, v)$” mean that there is a length-two walk between u and v.

(a) Explain why $W(u, u)$ holds iff $\exists v.\ A(u, v)$.

(b) Write a predicate-logic formula defining $W(u, v)$ in terms of the predicate $A(.,.)$ when $u \neq v$.

There are $e ::= \binom{n}{2}$ possible edges between the n vertices of G. Suppose the actual edges of $E(G)$ are chosen with randomly from this set of e possible edges. Each edge is chosen with probability p, and the choices are mutually independent.

(c) Write a simple formula in terms of p, e and k for $\Pr[|E(G)| = k]$.

(d) Write a simple formula in terms of p and n for $\Pr[W(u, u)]$.

Let w, x, y and z be four distinct vertices.

Because edges are chosen mutually independently, if the edges that one event depends on are disjoint from the edges that another event depends on, then the two events will be mutually independent. For example, the events

$$A(w, y) \text{ AND } A(y, x)$$

and

$$A(w, z) \text{ AND } A(z, x)$$

are independent since the first event dependes on $\{\langle w—y\rangle, \langle y—x\rangle\}$, while the second event depends on $\{\langle w—z\rangle, \langle z—x\rangle\}$.

(e) Let

$$r ::= \Pr[\text{NOT}(W(w, x))], \tag{18.16}$$

where w and x are distinct vertices. Write a simple formula for r in terms of n and p.

Hint: Different length-two paths between x and y don't share any edges.

(f) Vertices x and y being on a three-cycle can be expressed simply as

$$A(x, y) \text{ AND } W(x, y).$$

Write a simple expression in terms of p and r for the probability that x and y lie on a three-cycle in G.

(g) Show that $W(w, x)$ and $W(y, z)$ may not be independent events. *Hint:* Just consider the case that $V(G) = \{w, x, y, z\}$ and $p = 1/2$.

Exam Problems

Problem 18.31. (a) Show that any total, symmetric, transitive relation is reflexive.

(b) Conclude that there are events A, B, C such that A is independent of B, B is independent of C, but C is not independent of A.

Problems for Section 18.8

Practice Problems

Problem 18.32.
Suppose A, B and C are mutually independent events, what about $A \cap B$ and $B \cup C$?

Class Problems

Problem 18.33.
Suppose you flip three fair, mutually independent coins. Define the following events:

- Let A be the event that *the first* coin is heads.
- Let B be the event that *the second* coin is heads.
- Let C be the event that *the third* coin is heads.
- Let D be the event that *an even number of* coins are heads.

(a) Use the four step method to determine the probability space for this experiment and the probability of each of A, B, C, D.

(b) Show that these events are not mutually independent.

(c) Show that they are 3-way independent.

Problem 18.34.
Let A, B, C be events. For each of the following statements, prove it or give a counterexample.

(a) If A is independent of B, then A is also independent of $\overline{B}$.

(b) If A is independent of B, and A is independent of C, then A is independent of $B \cap C$.

Hint: Choose A, B, C pairwise but not 3-way independent.

(c) If A is independent of B, and A is independent of C, then A is independent of $B \cup C$.

Hint: Part (b).

(d) If A is independent of B, and A is independent of C, and A is independent of $B \cap C$, then A is independent of $B \cup C$.

Problem 18.35.
Let A, B, C, D be events. Describe counterexamples showing that the following claims are false.

(a)

False Claim. *If A and B are independent given C, and are also independent given D, then A and B are independent given $C \cup D$.*

(b)

False Claim. *If A and B are independent given C, and are also independent given D, then A and B are independent given $C \cap D$.*

Hint: Choose A, B, C, D 3-way but not 4-way independent.

so A and B are not independent given $C \cap D$.

Homework Problems

Problem 18.36.
Describe events A, B and C that:

- satisfy the "product rule," namely,

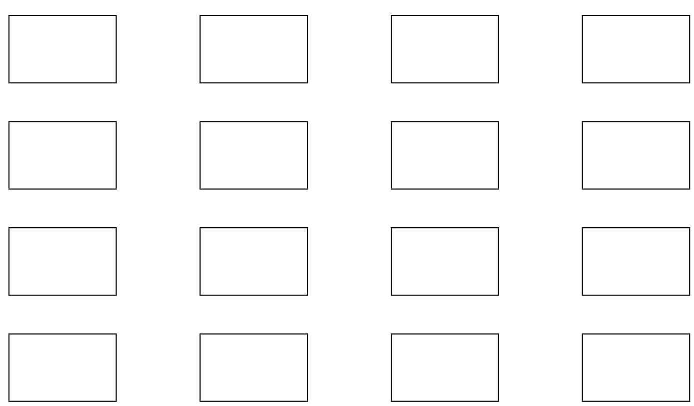
$$\Pr[A \cap B \cap C] = \Pr[A] \cdot \Pr[B] \cdot \Pr[C],$$

- and additionally, no two out of the three events are independent.

Hint: Choose A, B, C events over the uniform probability space on [1..6].

Exam Problems

Problem 18.37.
A classroom has sixteen desks in a 4×4 arrangement as shown below.



If two desks are next to each other, vertically or horizontally, they are called an *adjacent pair*. So there are three horizontally adjacent pairs in each row, for a total of twelve horizontally adjacent pairs. Likewise, there are twelve vertically adjacent pairs.

Boys and girls are assigned to desks mutually independently, with probability $p > 0$ of a desk being occupied by a boy and probability $q ::= 1 - p > 0$ of being occupied by a girl. An adjacent pair D of desks is said to have a *flirtation* when there is a boy at one desk and a girl at the other desk. Let F_D be the event that D has a flirtation.

(a) Different pairs D and E of adjacent desks are said to *overlap* when they share a desk. For example, the first and second pairs in each row overlap, and so do the second and third pairs, but the first and third pairs do not overlap. Prove that if D and E overlap, then F_D and F_E are independent events iff $p = q$.

(b) Find four pairs of desks D_1, D_2, D_3, D_4 and explain why $F_{D_1}, F_{D_2}, F_{D_3}, F_{D_4}$ are *not* mutually independent (even if $p = q = 1/2$).

Problems for Section 18.9

Problem 18.38.
An *International Journal of Pharmacological Testing* has a policy of publishing drug trial results only if the conclusion holds at the 95% confidence level. The editors and reviewers always carefully check that any results they publish came from a drug trial that genuinely deserved this level of confidence. They are also careful to check that trials whose results they publish have been conducted independently of each other.

The editors of the Journal reason that under this policy, their readership can be confident that at most 5% of the published studies will be mistaken. Later, the editors are embarrassed—and astonished—to learn that *every one* of the 20 drug trial results they published during the year was wrong. The editors thought that because the trials were conducted independently, the probability of publishing 20 wrong results was negligible, namely, $(1/20)^{20} < 10^{-25}$.

Write a brief explanation to these befuddled editors explaining what's wrong with their reasoning and how it could be that all 20 published studies were wrong.

Hint: xkcd comic: "significant" `xkcd.com/882/`

Practice Problems

Problem 18.39.
A somewhat reliable allergy test has the following properties:

- If you are allergic, there is a 10% chance that the test will say you are not.
- If you are not allergic, there is a 5% chance that the test will say you are.

(a) The test results are correct at what confidence level?

(b) What is the Bayes factor for being allergic when the test diagnoses a person as allergic?

(c) What can you conclude about the odds of a random person being allergic given that the test diagnoses them as allergic? Can you determine if the odds are better than even?

Suppose that your doctor tells you that because the test diagnosed you as allergic, and about 25% of people are allergic, the odds are six to one that you are allergic.

(d) How would your doctor calculate these odds of being allergic based on what's known about the allergy test?

(e) Another doctor reviews your test results and medical record and says your odds of being allergic are really much higher, namely thirty-six to one. Briefly explain how two conscientious doctors could disagree so much. Is there a way you could determine your actual odds of being allergic?