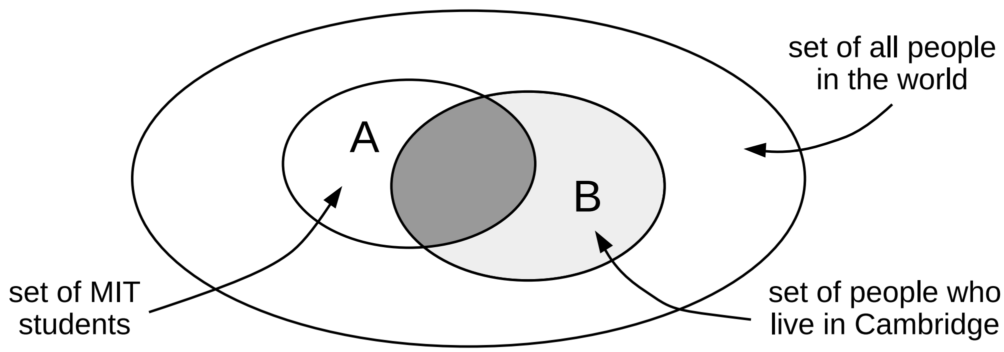
Tom Leighton, Marten van Dijk, and Brooke Cowan

Conditional Probability

Suppose that we pick a random person in the world. Everyone has an equal chance of being selected. Let A be the event that the person is an MIT student, and let B be the event that the person lives in Cambridge. What are the probabilities of these events? Intuitively, we're picking a random point in the big ellipse shown below and asking how likely that point is to fall into region A or B:



The vast majority of people in the world neither live in Cambridge nor are MIT students, so events A and B both have low probability. But what is the probability that a person is an MIT student, *given* that the person lives in Cambridge? This should be much greater— but what it is exactly?

What we're asking for is called a ***conditional probability***; that is, the probability that one event happens, given that some other event definitely happens. Questions about conditional probabilities come up all the time:

- What is the probability that it will rain this afternoon, given that it is cloudy this morning?
- What is the probability that two rolled dice sum to 10, given that both are odd?
- What is the probability that I'll get four-of-a-kind in Texas No Limit Hold 'Em Poker, given that I'm initially dealt two queens?

There is a special notation for conditional probabilities. In general, $\Pr(A \mid B)$ denotes the probability of event A, given that event B happens. So, in our example, $\Pr(A \mid B)$ is the probability that a random person is an MIT student, given that he or she is a Cambridge resident.

How do we compute $\Pr(A \mid B)$? Since we are *given* that the person lives in Cambridge, we can forget about everyone in the world who does not. Thus, all outcomes outside event B are irrelevant. So, intuitively, $\Pr(A \mid B)$ should be the fraction of Cambridge residents that are also MIT students; that is, the answer should be the probability that the person is in set $A \cap B$ (darkly shaded) divided by the probability that the person is in set B (lightly shaded). This motivates the definition of conditional probability:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

If $\Pr(B) = 0$, then the conditional probability $\Pr(A \mid B)$ is undefined.

Probability is generally counterintuitive, but conditional probability is the worst! Conditioning can subtly alter probabilities and produce unexpected results in randomized algorithms and computer systems as well as in betting games. Yet, the mathematical definition of conditional probability given above is very simple and should give you no trouble— provided you rely on formal reasoning and not intuition.

1 The Halting Problem

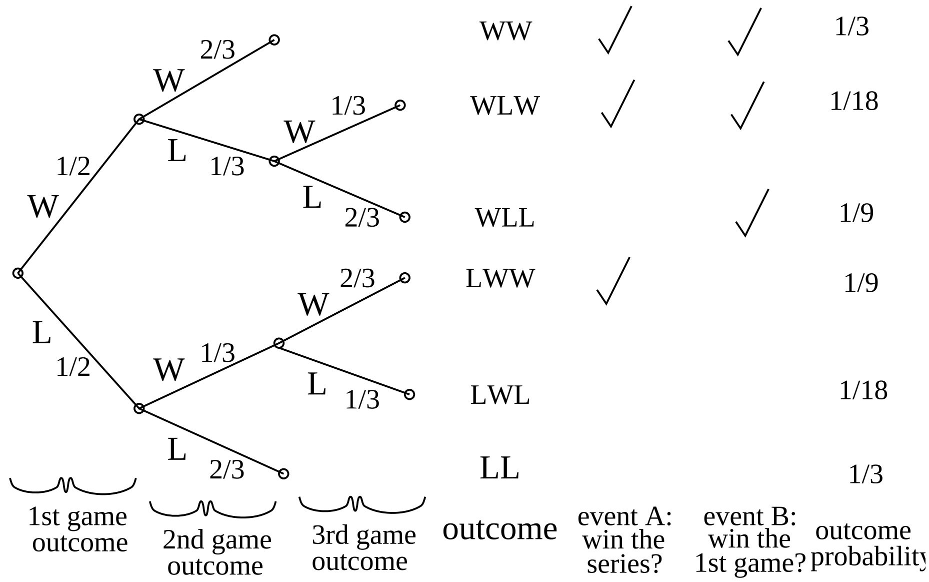
The *Halting Problem* is the canonical undecidable problem in computation theory that was first introduced by Alan Turing in his seminal 1936 paper. The problem is to determine whether a Turing machine halts on a given blah, blah, blah. Anyway, *much more importantly*, it is the name of the MIT EECS department's famed C-league hockey team.

In a best-of-three tournament, the Halting Problem wins the first game with probability $\frac{1}{2}$. In subsequent games, their probability of winning is determined by the outcome of the previous game. If the Halting Problem won the previous game, then they are envigorated by victory and win the current game with probability $\frac{2}{3}$. If they lost the previous game, then they are demoralized by defeat and win the current game with probablity only $\frac{1}{3}$. What is the probability that the Halting Problem wins the tournament, given that they win the first game?

1.1 Solution to the Halting Problem

This is a question about a conditional probability. Let A be the event that the Halting Problem 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 below, followed by an explanation of its construction and use.



Step 1: Find the Sample Space

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

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

Step 2: Define Events of Interest

The event that the Halting Problem wins the whole tournament is:

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

And the event that the Halting Problem wins the first game is:

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

The outcomes in these events are indicated with checkmarks in the tree diagram.

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 Halting Problem 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 Halting Problem wins the tournament, given that they win the first game:

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

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

1.2 Why Tree Diagrams Work

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

The answer involves conditional probabilities. In fact, the probabilities that we've been recording on the edges of tree diagrams *are* conditional probabilities. For example, consider the uppermost path in the tree diagram for the Halting Problem, which corresponds to the outcome WW. The first edge is labeled 1/2, which is the probability that the Halting Problem wins the first game. The second edge is labeled 2/3, which is the probability that the Halting Problem wins the second game, *given* that they won the first— that's 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. But why can we multiply edge probabilities to get outcome probabilities? For example, we concluded that:

$$\begin{aligned}
\Pr(WW) &= \frac{1}{2} \cdot \frac{2}{3} \\
&= \frac{1}{3}
\end{aligned}$$

Why is this correct?

The answer goes back to the definition of conditional probability. Rewriting this in a slightly different form gives the ***Product Rule*** for probabilities:

Rule (Product Rule for 2 Events)**.** *If* $\Pr(A_2) \neq 0$*, then:*

$$\Pr(A_1 \cap A_2) = \Pr(A_1) \cdot \Pr(A_2 \mid A_1)$$

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

$$\begin{aligned} &\Pr(\text{win first game} \cap \text{win second game}) \\ &\qquad = \Pr(\text{win first game}) \cdot \Pr(\text{win second game} \mid \text{win first game}) \\ &\qquad = \frac{1}{2} \cdot \frac{2}{3} \end{aligned}$$

So the Product Rule is the formal justification for multiplying edge probabilities to get outcome probabilities!

To justify multiplying edge probabilities along longer paths, we need a more general form the Product Rule:

Rule (Product Rule for n Events)**.** *If* $\Pr(A_1 \cap \ldots \cap A_{n-1}) \neq 0$, *then:*

$$\Pr(A_1 \cap \ldots \cap A_n) = \Pr(A_1) \cdot \Pr(A_2 \mid A_1) \cdot \Pr(A_3 \mid A_1 \cap A_2) \cdots \Pr(A_n \mid A_1 \cap \ldots \cap A_{n-1})$$

Let's interpret this big formula in terms of tree diagrams. Suppose we want to compute the probability that an experiment traverses a particular root-to-leaf path of length n. Let A_i be the event that the experiment traverses the i-th edge of the path. Then $A_1 \cap \ldots \cap A_n$ is the event that the experiment traverse the whole path. The Product Rule says that the probability of this is the probability that the experiment takes the first edge times the probability that it takes the second, *given* it takes the first edge, times the probability it takes the third, *given* it takes the first two edges, and so forth. In other words, the probability of an outcome is the product of the edge probabilities along the corresponding root-to-leaf path.

2 *A Posteriori* Probabilities

Suppose that we turn the hockey question around: what is the probability that the Halting Problem won their first game, given that they won the series?

This seems like an absurd question! After all, if the Halting Problem won the series, then the winner of the first game has already been determined. Therefore, who won the first game is a question of fact, not a question 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 Halting Problem 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 The Halting Problem won the first game.

A conditional probability $\Pr(B \mid A)$ is called an ***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.

Mathematically, a posteriori probabilities are *no different* from ordinary probabilities; the distinction is only at a higher, philosophical level. Our only reason for drawing attention to them is to say, "Don't let them rattle you."

Let's return to the original problem. The probability that the Halting Problem won their first game, given that they won the series is $\Pr(B \mid A)$. We can compute this using the definition of conditional probability and our earlier tree diagram:

$$\begin{aligned}\Pr(B \mid A) &= \frac{\Pr(B \cap A)}{\Pr(A)} \\ &= \frac{1/3 + 1/18}{1/3 + 1/18 + 1/9} \\ &= \frac{7}{9}\end{aligned}$$

This answer is suspicious! In the preceding section, we showed that $\Pr(A \mid B)$ was also $7/9$. Could it be true that $\Pr(A \mid B) = \Pr(B \mid A)$ in general? Some reflection suggests this is unlikely. For example, the probability that I feel uneasy, given that I was abducted by aliens, is pretty large. But the probability that I was abducted by aliens, given that I feel uneasy, is rather small.

Let's work out the general conditions under which $\Pr(A \mid B) = \Pr(B \mid A)$. By the definition of conditional probability, this equation holds if an only if:

$$\frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A \cap B)}{\Pr(A)}$$

This equation, in turn, holds only if the denominators are equal or the numerator is 0:

$$\Pr(B) = \Pr(A) \qquad \text{or} \qquad \Pr(A \cap B) = 0$$

The former condition holds in the hockey example; the probability that the Halting Problem wins the series (event A) is equal to the probability that it wins the first game (event B). In fact, both probabilities are $1/2$.

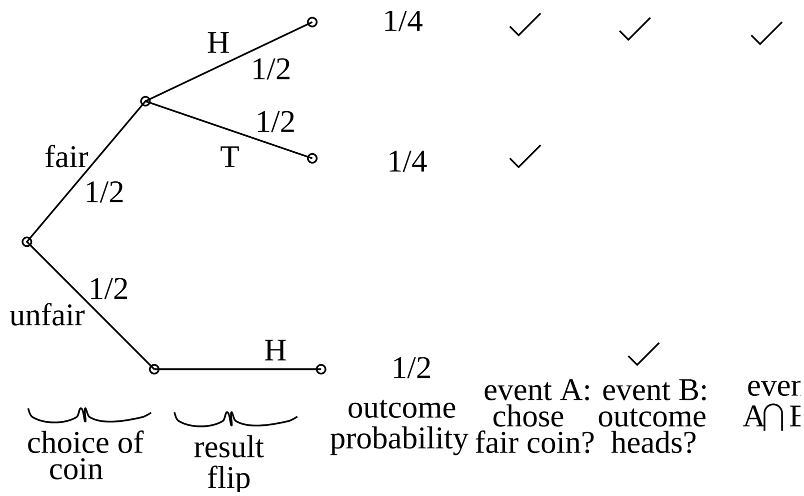
2.1 A Coin Problem

Suppose you have two coins. One coin is fair; that is, comes up heads with probability $1/2$ and tails with probability $1/2$. The other is a trick coin; it has heads on both sides, and so always comes up heads. Now you choose a coin at random so that you're equally likely to pick each one. If you flip the coin you select and get heads, then what is the probability that you flipped the fair coin?

This is another *a posteriori* problem since we want the probability of an event (that the fair coin was chosen) given the outcome of a later event (that heads came up). Intuition may fail us, but the standard four-step method works perfectly well.

Step 1: Find the Sample Space

The sample space is worked out in the tree diagram below.



Step 2: Define Events of Interest

Let A be the event that the fair coin was chosen. Let B the event that the result of the flip was heads. The outcomes in each event are marked in the figure. We want to compute $\Pr(A \mid B)$, the probability that the fair coin was chosen, given that the result of the flip was heads.

Step 2: Compute Outcome Probabilities

First, we assign probabilities to edges in the tree diagram. Each coin is chosen with probability 1/2. If we choose the fair coin, then head and tails each come up with probability 1/2. If we choose the trick coin, then heads comes up with probability 1. By the Product Rule, the probability of an outcome is the product of the probabilities on the corresponding root-to-leaf path. All of these probabilities are shown in the tree diagram.

Step 4: Compute Event Probabilities

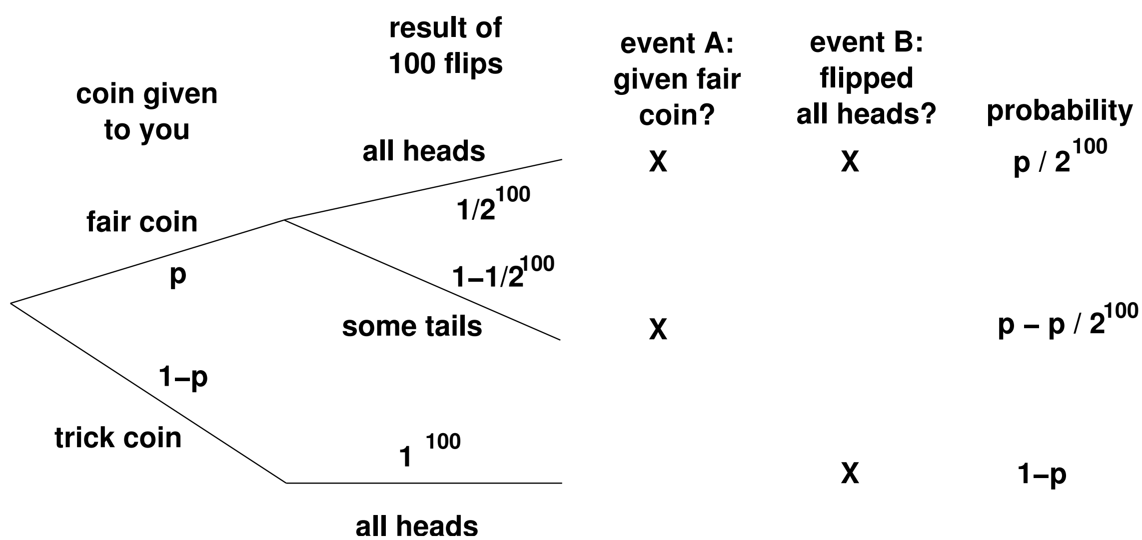
$$\begin{aligned}\Pr(A \mid B) &= \frac{\Pr(A \cap B)}{\Pr(B)} \\ &= \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{2}} \\ &= \frac{1}{3}\end{aligned}$$

The first equation uses the Product Rule. On the second line, we use the fact that the probability of an event is the sum of the probabilities of the outcomes it contains. The final line is simplification. The probability that the fair coin was chosen, given that the result of the flip was heads, is 1/3.

2.2 A Variant of the Two Coins Problem

Let's consider a variant of the two coins problem. Someone hands you either a fair coin or a trick coin with heads on both sides. You flip the coin 100 times and see heads every time. What can you say about the probability that you flipped the fair coin? Remarkably—nothing!

In order to make sense out of this outrageous claim, let's formalize the problem. The sample space is worked out in the tree diagram below. We do not know the probability that you were handed the fair coin initially— you were just given one coin or the other— so let's call that p.



Let A be the event that you were handed the fair coin, and let B be the event that you flipped 100 heads. Now, we're looking for $\Pr(A \mid B)$, the probability that you were handed the fair coin, given that you flipped 100 heads. The outcome probabilities are worked out in the tree diagram. Plugging the results into the definition of conditional probability gives:

$$\begin{aligned}\Pr(A \mid B) &= \frac{\Pr(A \cap B)}{\Pr(B)} \\ &= \frac{p/2^{100}}{1 - p + p/2^{100}} \\ &= \frac{p}{2^{100}(1-p) + p}\end{aligned}$$

This expression is very small for moderate values of p because of the 2^{100} term in the denominator. For example, if $p = 1/2$, then the probability that you were given the fair coin is essentially zero.

But we *do not know* the probability p that you were given the fair coin. And perhaps the value of p is *not* moderate; in fact, maybe $p = 1 - 2^{-100}$. Then there is nearly an even chance that you have the fair coin, given that you flipped 100 heads. In fact, maybe you were handed the fair coin with probability $p = 1$. Then the probability that you were given the fair coin is, well, 1!

A similar problem arises in polling before an election. A pollster picks a random American and asks his or her party affiliation. If this process is repeated many times, what can be said about the population as a whole? To clarify the analogy, suppose that the country contains only two people. There is either one Republican and one Democrat (like the fair coin), or there are two Republicans (like the trick coin). The pollster picks a random citizen 100 times, which is analogous to flipping the coin 100 times. Suppose that he picks a Republican every single time. We just showed that, even given this polling data, the probability that there is one citizen in each party could still be anywhere between 0 and 1!

What the pollster *can* say is that either:

1. Something earth-shatteringly unlikely happened during the poll.

2. There are two Republicans.

This is as far as probability theory can take us; from here, you must draw your own conclusions. Based on life experience, many people would consider the second possibility more plausible. However, if you are just *convinced* that the country isn't entirely Republican (say, because you're a citizen and a Democrat), then you might believe that the first possibility is actually more likely.

3 Medical Testing

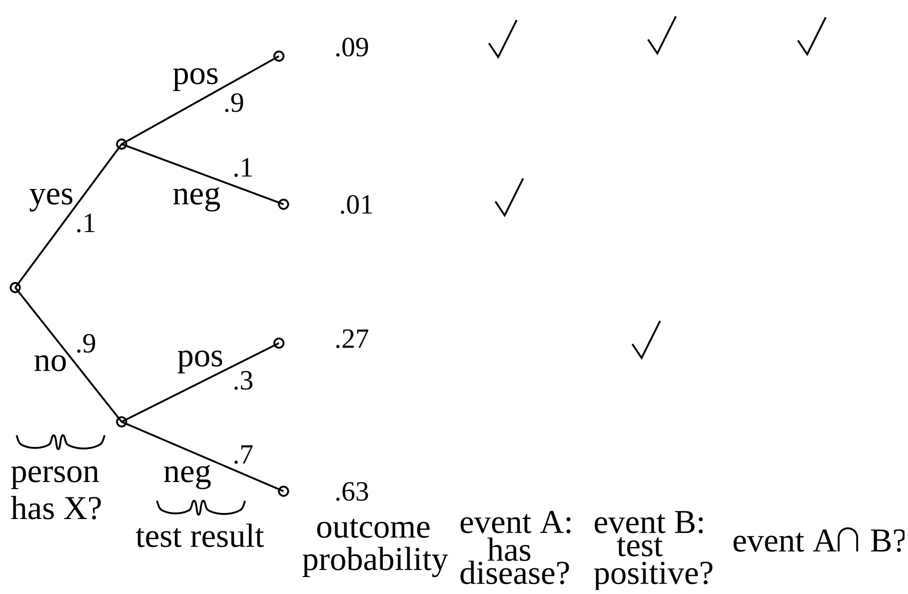
There is a deadly disease called X that has infected 10% of the population. There are no symptoms; victims just drop dead one day. Fortunately, there is a test for the disease. The test is not perfect, however:

- If you have the disease, there is a 10% chance that the test will say you do not. (These are called "false negatives".)
- If you do not have disease, there is a 30% chance that the test will say you do. (These are "false positives".)

A random person is tested for the disease. If the test is positive, then what is the probability that the person has the disease?

Step 1: Find the Sample Space

The sample space is found with the tree diagram below.



Step 2: Define Events of Interest

Let A be the event that the person has the disease. 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 disease X, 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 the figure.

Step 4: Compute Event Probabilities

$$\begin{aligned}\Pr(A \mid B) &= \frac{\Pr(A \cap B)}{\Pr(B)} \\ &= \frac{0.09}{0.09 + 0.27} \\ &= \frac{1}{4}\end{aligned}$$

If you test positive, then there is only a 25% chance that you have the disease!

This answer is initially surprising, but makes sense on reflection. There are two ways you could test positive. First, it could be that you are sick and the test is correct. Second, it could be that you are healthy and the test is incorrect. The problem is that almost everyone is healthy; therefore, most of the positive results arise from incorrect tests of healthy people!

We can also compute the probability that the test is correct for a random person. This event consists of two outcomes. The person could be sick and the test positive (probability 0.09), or the person could be healthy and the test negative (probability 0.63). Therefore, the test is correct with probability $0.09 + 0.63 = 0.72$. This is a relief; the test is correct almost three-quarters of the time.

But wait! There is a simple way to make the test correct 90% of the time: always return a negative result! This "test" gives the right answer for all healthy people and the wrong answer only for the 10% that actually have the disease. The best strategy is to completely ignore the test result!

There is a similar paradox in weather forecasting. During winter, almost all days in Boston are wet and overcast. Predicting miserable weather every day may be more accurate than really trying to get it right!

4 Conditional Probability Pitfalls

The remaining sections illustrate some common blunders involving conditional probability.

4.1 Carnival Dice

There is a gambling game called Carnival Dice. A player picks a number between 1 and 6 and then rolls three fair dice. The player wins if his number comes up on at least one die. The player loses if his number does not appear on any of the dice. What is the probability that the player wins? This problem sounds simple enough that we might try an intuitive lunge for the solution.

False Claim 1. *The player wins with probability* $\frac{1}{2}$.

Proof. Let A_i be the event that the i-th die matches the player's guess.

$$\begin{aligned}
\Pr(\text{win}) &= \Pr\left(A_1 \cup A_2 \cup A_3\right) \\
&= \Pr\left(A_1\right) \cup \Pr\left(A_2\right) \cup \Pr\left(A_3\right) \\
&= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\
&= \frac{1}{2}
\end{aligned}$$

□

The only justification for the second equality is that it looks vaguely reasonable; in fact, equality does not hold. Let's examine the expression $\Pr(A_1 \cup A_2 \cup A_3)$ to see exactly what is happening. Recall that the probability of an event is the sum of the probabilities of the outcomes it contains. Therefore, we could argue as follows:

$$\begin{aligned}
\Pr\left(A_1 \cup A_2 \cup A_3\right) &= \sum_{w \in A_1 \cup A_2 \cup A_3} \Pr\left(w\right) \\
&= \sum_{w \in A_1} \Pr\left(w\right) + \sum_{w \in A_2} \Pr\left(w\right) + \sum_{w \in A_3} \Pr\left(w\right) \\
&= \Pr\left(A_1\right) + \Pr\left(A_2\right) + \Pr\left(A_3\right)
\end{aligned}$$

This argument is valid provided that the events A_1, A_2, and A_3 are *disjoint*; that is, there is no outcome in more than one event. If this were not true for some outcome, then a term would be duplicated when we split the one sum into three. Subject to this caveat, the argument generalizes to prove the following theorem:

Theorem 2. *Let $A_1, A_2, \ldots A_n$ be disjoint events. Then:*

$$\Pr\left(A_1 \cup A_2 \cup \ldots \cup A_n\right) = \Pr\left(A_1\right) + \Pr\left(A_2\right) + \ldots + \Pr\left(A_n\right)$$

We can evaluate the probability of a union of events that are not necessarily disjoint using a theorem analogous to Inclusion-Exclusion. Here is the special case for a union of three events.

Theorem 3. *Let A_1, A_2, and A_3 be events, not necessarily disjoint. Then:*

$$\begin{aligned}
\Pr\left(A_1 \cup A_2 \cup A_3\right) = {} & \Pr\left(A_1\right) + \Pr\left(A_2\right) + \Pr\left(A_3\right) \\
& - \Pr\left(A_1 \cap A_2\right) - \Pr\left(A_1 \cap A_3\right) - \Pr\left(A_2 \cap A_3\right) \\
& + \Pr\left(A_1 \cap A_2 \cap A_3\right)
\end{aligned}$$

We can use this theorem to compute the real chance of winning at Carnival Dice. The probability that one die matches the player's guess is $1/6$. The probability that two dice both match the player's guess is 1/36 by the Product Rule. Similarly, the probability that all three dice match is 1/216. Plugging these numbers into the preceding theorem gives:

$$\begin{aligned}\Pr(\text{win}) &= \frac{1}{6}+\frac{1}{6}+\frac{1}{6}\\ &\quad -\frac{1}{36}-\frac{1}{36}-\frac{1}{36}\\ &\quad +\frac{1}{216}\\ &\approx 42\%\end{aligned}$$

These are terrible odds for a gambling game; you'd be much better off playing roulette, craps, or blackjack!

4.2 Other Identities

There is a close relationship between computing the size of a set and computing the probability of an event. Theorem 3 is one example; the probability of a union of events and the cardinality of a union of sets are computed using similar formulas.

In fact, all of the methods we developed for computing sizes of sets carry over to computing probabilities. This is because a probability space is just a weighted set; the sample space is the set and the probability function assigns a weight to each element. Earlier, we were counting the number of items in a set. Now, when we compute the probability of an event, we are just summing the weights of items. We'll see many examples of the close relationship between probability and counting over the next few weeks.

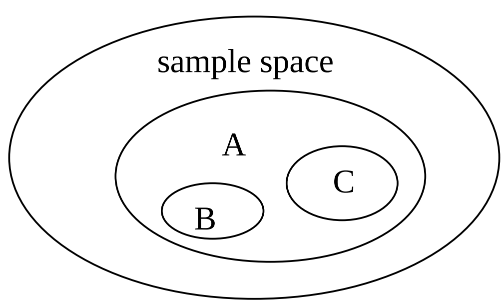
Many general probability identities still hold when all probabilities are conditioned on the same event. For example, the following identity is analogous to the Inclusion-Exclusion formula for two sets, except that all probabilities are conditioned on an event C.

$$\Pr(A\cup B \mid C) = \Pr(A \mid C) + \Pr(B \mid C) - \Pr(A\cap B \mid C)$$

Be careful not to mix up events before and after the conditioning bar! For example, the following is *not* a valid identity:

$$\Pr(A \mid B\cup C) = \Pr(A \mid B) + \Pr(A \mid C) \qquad (B\cap C=\phi)$$

A counterexample is shown below. In this case, $\Pr(A \mid B) = 1$, $\Pr(A \mid C) = 1$, and $\Pr(A \mid B\cup C) = 1$. However, since $1 \neq 1+1$, the equation above does not hold.



So you're convinced that this equation is false in general, right? Let's see if you *really* believe that.

4.3 Discrimination Lawsuit

Several years ago there was a sex discrimination lawsuit against Berkeley. A female professor was denied tenure, allegedly because she was a woman. She argued that in every one of Berkeley's 22 departments, the percentage of male applicants accepted was greater than the percentage of female applicants accepted. This sounds very suspicious!

However, Berkeley's lawyers argued that across the whole university the percentage of male tenure applicants accepted was actually *lower* than the percentage of female applicants accepted. This suggests that if there was any sex discrimination, then it was against men! Surely, at least one party in the dispute must be lying.

Let's simplify the problem and express both arguments in terms of conditional probabilities. Suppose that there are only two departments, EE and CS, and consider the experiment where we pick a random applicant. Define the following events:

- Let A be the event that the applicant is accepted.
- Let F_{EE} the event that the applicant is a female applying to EE.
- Let F_{CS} the event that the applicant is a female applying to CS.
- Let M_{EE} the event that the applicant is a male applying to EE.
- Let M_{CS} the event that the applicant is a male applying to CS.

Assume that all applicants are either male or female, and that no applicant applied to both departments. That is, the events F_{EE}, F_{CS}, M_{EE}, and M_{CS} are all disjoint.

In these terms, the plaintiff is make the following argument:

$$\Pr(A \mid F_{EE}) < \Pr(A \mid M_{EE})$$
$$\Pr(A \mid F_{CS}) < \Pr(A \mid M_{CS})$$

That is, in both departments, the probability that a woman is accepted for tenure is less than the probability that a man is accepted. The university retorts that overall a woman applicant is *more* likely to be accepted than a man:

$$\Pr(A \mid F_{EE} \cup F_{CS}) > \Pr(A \mid M_{EE} \cup M_{CS})$$

It is easy to believe that these two positions are contradictory. In fact, we might even try to prove this by adding the plaintiff's two inequalities and then arguing as follows:

$$\begin{aligned} & \Pr(A \mid F_{EE}) + \Pr(A \mid F_{CS}) < \Pr(A \mid M_{EE}) + \Pr(A \mid M_{CS}) \\ \Rightarrow \quad & \Pr(A \mid F_{EE} \cup F_{CS}) < \Pr(A \mid M_{EE} \cup M_{CS}) \end{aligned}$$

The second line exactly contradicts the university's position! But there is a big problem with this argument; the second inequality follows from the first only if we accept the "false identity" from the preceding section. This argument is bogus! Maybe the two parties do not hold contradictory positions after all!

In fact, the table below shows a set of application statistics for which the assertions of both the plaintiff and the university hold:

CS	0 females accepted, 1 applied	0%
	50 males accepted, 100 applied	50%
EE	70 females accepted, 100 applied	70%
	1 male accepted, 1 applied	100%
Overall	70 females accepted, 101 applied	≈ 70%
	51 males accepted, 101 applied	≈ 51%

In this case, a higher percentage of males were accepted in both departments, but overall a higher percentage of females were accepted! Bizarre!

4.4 On-Time Airlines

Newspapers publish on-time statistics for airlines to help travelers choose the best carrier. The on-time rate for an airline is defined as follows:

$$\text{on-time rate} = \frac{\text{\# flights less than 15 minutes late}}{\text{\# flights total}}$$

This seems reasonable, but actually can be badly misleading! Here is some on-time data for two airlines in the late 80's.

	Alaska Air			America West		
Airport	#on-time	#flights	%	#on-time	#flights	%
Los Angeles	500	560	89	700	800	87
Phoenix	220	230	95	4900	5300	92
San Diego	210	230	92	400	450	89
San Francisco	500	600	83	320	450	71
Seattle	1900	2200	86	200	260	77
OVERALL	3330	3820	87	6520	7260	90

America West had a better overall on-time percentage, but Alaska Airlines did better at *every single airport!* This is the same paradox as in the Berkeley tenure lawsuit. The problem is that Alaska Airlines flew proportionally more of its flights to bad weather airports like Seattle, whereas America West was based in fair-weather, low-traffic Phoenix!