5
Bayes' rule

Our conditional probabilities always depend on what we know. When our knowledge changes, these probabilities must also change. A 250-year-old mathematical principle called Bayes' Rule tells us how.

Bayesian search: finding the USS *Scorpion*

In February of 1968, the USS *Scorpion* set sail from the naval base in Norfolk, Virginia, under the command of Francis Slattery. The *Scorpion* was a Skipjack-class high-speed attack submarine, the fastest in the American fleet. Like other subs of her class, she played a major role in U.S. military strategy. Think *The Hunt for Red October* here: throughout the Cold War, both the Americans and the Soviets deployed large fleets of attack subs, whose mission was to locate, track, and—should the unthinkable happen—destroy the other side's ballistic-missile submarines.

On this deployment, the *Scorpion* sailed east, bound for the Mediterranean Sea, where for three months she participated in training exercises alongside the 6th Naval Fleet. Then in mid-May, the *Scorpion* was sent back west, past Gibraltar and out into the Atlantic. There she was ordered to observe Soviet naval vessels operating near the Azores—a remote island chain in the middle of the North Atlantic, about 850 miles off the coast of Portugal—and then to continue west, bound for home. The sub was due back in Norfolk at 1 PM on Monday, May 27th, 1968.

On the docks in Norfolk that day, the families of the *Scorpion*'s 99 crew members were gathered to welcome their loved ones back home. But as 1 PM came and went, the sub had not yet surfaced. Minutes stretched into hours; day gave way to night. Still the families waited. But there was no sign of the *Scorpion*.

With growing alarm, the Navy ordered a search. By 10 PM, the operation involved 18 ships; by the next morning, 37 ships and 17 long-range patrol aircraft. But the odds of a good outcome were slim. The *Scorpion* had last made contact off the Azores, 6 days ago, and 2,670 miles away from Norfolk. She could have been anywhere along that strip of ocean between the Azores and the eastern seaboard. As the hours ticked by, the chances that the sub could be located, and that rescue gear could be deployed in time, were rapidly diminishing. At a tense news conference on May 28th, President Lyndon Johnson summarized the mood of a nation: "Nothing encouraging to report. . . . We are all quite distressed."

Day after day went by, but the search for the *Scorpion* turned up no results. Finally, after eight days, the Navy was forced to concede the obvious: the *Scorpion*'s crew of 99 men were declared lost at sea, presumed dead.

The Navy now turned to the grim task of locating the *Scorpion*'s final resting place—a tiny needle in a vast haystack stretching three-fourths of the way across the North Atlantic. Although hopes for saving the crew had been dashed, the stakes were still high, and not only for the families of those lost: the *Scorpion* had carried two nuclear-tipped torpedoes, each capable of sinking an aircraft carrier with a single hit. These dangerous warheads were now somewhere on the bottom of the sea.

John Craven, Bayesian search guru

To lead the search for the *Scorpion*, the Pentagon turned to Dr. John Craven, chief scientist in the Navy's Special Projects Office, and the leading guru on finding lost objects in deep water.

Remarkably, Dr. Craven had done this kind of thing before. Two years earlier, in 1966, an American B-52 bomber had collided in mid-air with a refueling tanker over the Spanish coast, near the seaside village of Palomares. Both planes crashed, and the B-52's four hydrogen bombs, each of them 50 times more powerful than the Hiroshima explosion, were scattered for miles. Luckily none of the warheads had detonated, and three of the bombs were found more or less immediately.[1] But the fourth bomb was missing, and was presumed to have fallen into the sea. John Craven was called upon to help find it.

[1] Albeit after one of them had contaminated a roughly one-square-mile area of tomato farms and woodland with radioactive plutonium. The clean-up operation in the wake of this incident continues 50 years later, with the latest negotiations between Spain and the U.S. taking place in 2015.

Craven and his team had to ponder many unknown variables about the crash. Had the bomb remained in the plane, or had it fallen out? If the bomb had fallen out, had either or both of its parachutes deployed? If the parachutes had deployed, had the winds taken the bomb far out to sea? If so, in what direction, and exactly how far?

Bayesian search. To sort through this thicket of unknowns, Craven turned to his preferred strategy: *Bayesian search.* This search methodology had been pioneered during World War II, when the Allies used it to help locate German U-boats. But its origins stretched back much further, all the way to a mathematical principle called *Bayes' rule*, first worked out by an English reverend named Thomas Bayes, in the 1750s.

Bayesian search has three essential principles. First, you should combine the pre-search opinions of various experts about the plausibility of each possible scenario. In the case of the missing H-bomb, some of these experts would be familiar with mid-air crashes, some of them familiar with nuclear bombs, some with coastal winds and ocean currents, and so forth. These opinions should be synthesized to form a *prior probability* for each crash scenario—and, by extension, a prior probability that the bomb might be found in each possible search location. These probabilities are "prior," in the sense that they represent the best guess available, before anyone has any data.

Second, you must evaluate the capability of your search instruments to establish the likelihood that, if the object were in a given sector, you'd actually be able to find it there. This likelihood is combined with the prior probability to form a single search-effectiveness probability for each location. For example, let's say that the most plausible scenario puts the lost bomb at the bottom of a very deep ocean trench. Despite its high prior probability, this trench might still be a poor candidate location to begin your search, for the simple reason that the trench is so dark and remote that, even if the bomb were there, you'd be very unlikely to find it. To draw on a familiar metaphor, a Bayesian search has you start looking for your lost keys using a precise mathematical combination of two factors: where you think you lost them, and where the streetlight is shining brightest.

Third and finally, as new data comes in during the search process, you should use that new data to update your prior probability for each search location into a *posterior probability*. This Bayesian updating process is iterative, in the sense that today's posterior becomes tomorrow's prior. Suppose you search in today's region of high posterior probability, but find nothing. Then for tomorrow, you reduce the probability in the region you just searched, reassess your beliefs about each scenario, and bump up the probability in the other regions accordingly. You keep doing this day after day, always concentrating on that day's new region of highest probability, until you find what you're looking for.

Craven is stymied. Unfortunately, military politics, and a clash of personalities, prevented Dr. Craven and his team from actually applying these Bayesian principles to the 1966 search for the missing H-bomb off the coast of Palomares. In a classic military move, the Pentagon had asked the right hand to do one thing, and then asked the left hand to put some handcuffs on the right one, to make its job more difficult. The commanding officer on the scene, Rear Admiral William S. Guest—nickname: Bull Dog—had a notably different view of the way the search should be conducted. Bull Dog was a doer, not a thinker. He had little patience for probabilities, and even less patience for the team of twentysomething-year-old math Ph.D's he now found himself commanding. His initial orders to Craven's team, perhaps only half sarcastic, were for them to prove that the bomb had fallen on land rather than in the sea, so that it would be someone else's job to find it.[2]

[2] Sharon Bertsch McGrayne, *The Theory That Would Not Die*, Yale University Press, 2011 (pg. 190).

As a result, the search for the Palomares H-bomb was really two searches. There was Craven's Bayesian search, with its slide rules and probability maps, and with updated probabilities constantly chattering over the teletype machine as the mathematicians fed remote calculations to a mainframe computer back in Pennsylvania. But the insights arising from the Bayesian search were largely ignored in favor of Admiral Guest's "plan of squares," which guided the *real* search, and which was pretty much exactly what it sounds like. The frustrated Craven was like a high-school stock picker who records virtual trades in a ledger and watches his paper fortune grow, but never gets to buy and sell any shares.

Eventually, the bomb was found. It turns out that a fisherman named Francisco Orts had seen the bomb fall into the water under parachute, and he was able to guide the Navy to its exact point of entry. Thus while the search was a success, the Bayesian part of it had been a failure, for the simple reason that it had never been given a chance. Nonetheless, the Palomares incident taught John Craven some valuable lessons—both about the practicalities of conducting a Bayesian search, and about how to how to get the necessary support for that search from the military brass.

And two years later, when he was called upon to find the USS *Scorpion*, Craven was ready.

The search for the Scorpion continues

When the *Scorpion* disappeared in May of 1968, Craven and his Bayesian search team were quickly reconvened. At first, the task seemed vastly more daunting than the search for the Palomares H-bomb had been. Back then, they had known to confine the search to a relatively small area off the coast of southern Spain. But here, the team had to find a submarine under 2 miles of water, somewhere between Virginia and the Azores, without so much as a single clue.

Luckily, they caught a break. Starting in the early 1960s, the U.S. military had spent $17 billion installing an enormous, highly classified network of underwater microphones throughout the North Atlantic. Essentially, they had wired the entire ocean for sound, so that they could track the movements of the Soviet navy. Highly trained technicians at secret listening posts were monitoring these microphones around the clock. The technicians could look at the output from these devices and immediately distinguish the acoustic signature of a submarine from that of a whale, an oil tanker, or hot magma under the seabed.

After sniffing around, Craven discovered that one of these secret listening posts in the Canary Islands had, one day in late May, recorded a very unusual series of 18 underwater sounds. Then he learned that two other listening posts—both of them thousands of miles away, off the coast of Newfoundland—had recorded those very same sounds around the same time. Craven's team compared these three readings and, by triangulation, worked out that the sounds must have emanated from a very deep part of the Atlantic Ocean, about 400 miles southwest of the Azores.

This location fell along the *Scorpion*'s expected route home. Moreover, the sounds themselves were highly suggestive: a muffled underwater explosion; then 91 seconds of silence; and then 17 further sonic events in rapid succession that, to Craven, sounded like the implosion of various compartments of a submarine as it sank beneath its hull-crush depth.[3]

This acoustic revelation dramatically narrowed the size of the search area. Still, the team had about 140 square miles of ocean floor to cover, all of it 10,000 feet below the surface, and therefore inaccessible to all but the most advanced submersibles.

[3] PBS Nova documentary, "Submarines, Secrets, and Spies." Originally broadcast January 19, 1999. https://www.youtube.com/watch?v=NJWHiPSvzh8

The Bayesian search now kicked into high gear. Craven's first step was to take a map of the seabed and divide it up into a grid of little rectangles, each one a possible search location. Each rectangle got an alphanumeric code: B6, H3, and so on, just like in the board game *Battleship*. Craven and his team then interviewed expert submariners, and came up with nine possible scenarios—a fire on board, a torpedo exploding in its bay, a clandestine Russian attack, and so on—for how the submarine had sunk. They weighed the prior probability of each scenario, and ran computer simulations to understand how the *Scorpion*'s likely movements might have unfolded under each one. They assessed the capabilities of the search fleet: its cameras, its magnetic-sensing instruments, its sonars, its submersible robots. They even blew up depth charges at precise locations, in order to calibrate their original acoustical data from the listening posts in the Canary Islands and Newfoundland.

Finally, they put all this information together to form a single search-effectiveness probability for each cell on the grid. This map crystallized thousands upon thousands of hours of interviews, calculations, experiments, and careful thinking. It would have looked something like this:

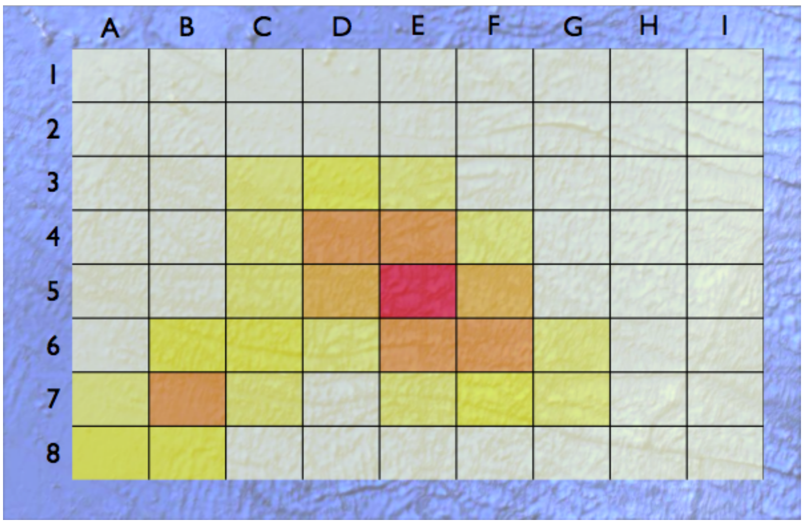


Figure 5.1: An attempted reconstruction of John Craven's probability map for the *Scorpion* search. Darker red rectangles indicate regions of relatively higher prior probability.

Mathematically speaking, this map represented the best chance for finding the *Scorpion*.

Predictably, Craven encountered both logistical and bureacratic difficulties in getting the Pentagon to pay attention to his map of probabilities. Summer came and went. By this point, the search for the *Scorpion* had been going on since early June, to no avail.

But eventually his cajoling paid off, and the military brass ordered that Craven's map be used to guide the now months-old search. So starting in October, when commanders leading the search aboard the USS *Mizar* finally got ahold the map, the operation became truly Bayesian. Day by day, the team rigorously searched the region of highest probability, and crunched the numbers to update the map for tomorrow. And day by day, those numbers were slowly homing in on rectangle F6.

Found

On October 28th, Bayes finally paid off.

The *Mizar* was in the midst of its 5th cruise, and its 74th individual run over the ocean floor. All of a sudden the ship's magnetometer spiked, suggesting an anomaly on the sea floor. Cameras were hurriedly deployed to investigate—and sure enough, there she was:

Figure 5.2: A photo of the bow section of the USS *Scorpion*, taken in 1968 by the crew of the bathyscape *Trieste II*. USN photo #1136658.

Partially buried in the sand, 400 miles from the nearest landfall and two miles below the surface of the sea, the USS *Scorpion* had been found at last.

To this day, nobody knows for sure what actually happened to the *Scorpion*—or if they do, they're not talking. The Navy's official version of events, though inconclusive, cites the accidental explosion of a torpedo or the malfunctioning of a garbage-disposal unit as two of the most likely possible causes of the tragedy. Many other explanations have been proposed over the years. And as with any famous mystery, conspiracy theories abound.

But there was at least one definitive conclusion to come out of the *Scorpion* incident: Bayesian search was a truly winning idea. As it turned out, the sub's final resting place lay a mere 260 yards away from rectangle E5, the initial region of highest promise on Craven's map of prior probabilities. The search team had actually passed over that location on a previous cruise, but had missed the telltale signs due to a broken sonar.[4]

[4] McGrayne, ibid.

Ponder that for a moment more. A lone submarine had been lost somewhere in a 2600-mile stretch of open ocean, and the Bayesian search had pinpointed her location to within 260 yards—only three lengths of the submarine itself. It was a remarkable triumph for Craven's team, and for Bayes' rule, the 250-year-old mathematical formula that had served as the search's guiding principle.

Today, Bayesian search is a small industry, with at least one college textbook[5] explaining the details, and with entire companies whose mission is to apply Bayesian principles to find what has been lost. To cite a recent example, many readers will remember the tragedy of Air France Flight 447, which crashed in the Atlantic Ocean on its way from Rio de Janeiro to Paris, in June of 2009. The search for the wreckage had been going on for two fruitless years; then in 2011, a Bayesian search firm was hired, a map of probabilities was drawn up—and the plane was found within one week of undersea search.[6]

[5] The Theory of Optimal Search (Operations Research Society of America, 1975), by Lawrence D. Stone.

[6] Stone et. al. "Search for the wreckage of Air France Flight AF 447." *Statistical Science* 2014, Vol. 29(1), pp. 69-80.

Moreover, the broader principle behind Bayesian search, Bayes' rule, is used almost everywhere: from courtrooms to doctor's offices, and from spam filters to self-driving cars. So if you want to learn more about the key equation that found the *Scorpion* and that helps power the modern world, then this chapter is for you.

Updating conditional probabilities

When our knowledge changes, our probabilities must change, too. Bayes' rule tells us how to change them.

Imagine the person in charge of a Toyota factory who starts with a subjective probability assessment for some proposition A, like "our engine assembly robots are functioning properly." Just to put a number on it, let's say $P(A) = 0.95$; we might have arrived at this judgment, for example, based on the fact that the robots have been down for 5% of the time over the previous month. In the absence of any other information, this is as good a guess as any.

Now we learn something new, like information B: the last 5 engines off the assembly line all failed inspection. Before we believed there was a 95% chance that the assembly line was working fine. What about now?

Figure 5.3: Bayes' rule is named after Thomas Bayes (above), an English reverend of the 18th century who first derived the result. It was published posthumously in 1763 in "An Essay towards solving a Problem in the Doctrine of Chances."

Bayes's rule is an explicit equation that tells us how to incorporate this new information, turning our initial probability $P(A)$ into a new, updated probability:

$$P(A \mid B) = \frac{P(A) \cdot P(B \mid A)}{P(B)} . \qquad (5.1)$$

Each piece of this equation has a name:

- $P(A)$ is the prior probability: how probable is A, before ever having seen data B?
- $P(A \mid B)$ is the posterior probability: how probable is A, now that we've seen data B?
- $P(B \mid A)$ is the likelihood: if A were true, how likely is it that we'd see data B?
- $P(B)$ is the total (or marginal) probability of B: how likely is it that we'd see data B anyway, regardless of whether A is true or not? This calculation is usually the tedious part of applying Bayes' rule. Usually, as we'll see in the examples, we use the rule of total probability, which we learned in a previous chapter.

Have you found the two-headed coin?

To get a feel for what's going on here, let's see an example of Bayes' rule in action.

Imagine a jar with 1024 normal quarters. Into this jar, a friend places a single two-headed quarter (i.e. with heads on both sides). Your friend then gives the jar a good shake to mix up the coins. You draw a single coin at random from the jar, and without examining it closely, flip the coin ten times. The coin comes up heads all ten times. Are you holding the two-headed quarter, or an ordinary quarter?

Now, you might be thinking that this example sounds pretty artificial. But it's not at all. In fact, in the real world, an awful lot of time and energy is spent looking for metaphorical two-headed coins—specifically, in any industry where companies compete strenuously for talented employees. To see why, let's change the story just a little bit.

Suppose you're in charge of a large trading desk at a major Wall Street bank. You have 1025 employees under you, and each one is responsible for managing a portfolio of stocks to make money for your firm and its clients.

One day, a young trader knocks on your door and confidently asks for a big raise. You ask her to make a case for why she deserves one. She replies:

> Look at my trading record. I've been with the company for ten months, and in each of those ten months, my portfolio returns have been in the top half of all the portfolios managed by my peers on the trading floor. If I were just an average trader, this would be very unlikely. In fact, the probability that an average trader would see above-average results for ten months in a row is only $(1/2)^{10}$, which is less than one chance in a thousand. Since it's unlikely I would be that lucky, the implication is that I am a talented trader, and I should therefore get a raise.

The math of this scenario is exactly the same as the one involving the big jar of quarters. Metaphorically, the trader is claiming to be a two-headed coin (T), on the basis of some data D: that she performs above average, every single month without fail.

But from your perspective, things are not so clear. Is the trader lucky, or good? There are 1025 people in your office (i.e. 1025 coins). Now you're confronted with the data that one of them has had an above-average monthly return for ten months in a row (i.e. D = "flipped heads ten times in a row"). This is admittedly unlikely, and this person might therefore be an excellent performer, worth paying a great deal to retain. But excellent performers are probably also rare, so that the prior probability $P(T)$ is pretty small to begin with. To make an informed decision, you need to know $P(T \mid D)$: the posterior probability that the trader is an above-average performer, given the data.

Applying Bayes' rule. So our two-headed coin example definitely has real-world applications. Let's use it to see how a posterior probability is calculated using Bayes' rule:

$$P(T \mid D) = \frac{P(T) \cdot P(D \mid T)}{P(D)} .$$

We'll take this equation one piece at a time. First, what is $P(T)$, the prior probability that you are holding the two-headed quarter? Well, there are 1025 quarters in the jar: 1024 ordinary ones, and one two-headed quarter. Assuming that your friend mixed the coins in the jar well enough, then you are just as likely to draw one coin as another, and so $P(T)$ must be 1/1025.

Next, what about $P(D \mid T)$, the likelihood of flipping ten heads in a row, given that you chose the two-headed quarter? Clearly this is 1—if the quarter has two heads, there is no possibility of seeing anything else.

Finally, what about $P(D)$, the marginal probability of flipping ten heads in a row? As is almost always the case when using Bayes' rule, $P(D)$ is the hard part to calculate. We will use the law of total probability to do so:

$$P(D) = P(T) \cdot P(D \mid T) + P(\text{not } T) \cdot P(D \mid \text{not } T) .$$

Taking the pieces on the right-hand one by one:

- As we saw above, the prior probability of the two-headed coin, $P(T)$, is 1/1025.

- This means that the prior probability of an ordinary coin, $P(\text{not } T)$, must be 1024/1025.

- Also from above, we know that $P(D \mid T) = 1$.

- Finally, we can calculate $P(D \mid \text{not } T)$ quite easily. If the coin is an ordinary quarter, then there is a 50% chance of getting heads on any one coin flip. Each flip is independent. Therefore, the probability of a 10-head winning streak is

$$\begin{aligned} P(D \mid \text{not } T) &= \frac{1}{2} \times \frac{1}{2} \times \cdots \times \frac{1}{2} \quad \text{(10 times)} \\ &= \left(\frac{1}{2}\right)^{10} = \frac{1}{1024} . \end{aligned}$$

We can now put all these pieces together:

$$\begin{aligned} P(T \mid D) &= \frac{P(T) \cdot P(D \mid T)}{P(T) \cdot P(D \mid T) + P(\text{not } T) \cdot P(D \mid \text{not } T)} \\ &= \frac{\frac{1}{1025} \cdot 1}{\frac{1}{1025} \cdot 1 + \frac{1024}{1025} \cdot \frac{1}{1024}} = \frac{1/1025}{2/1025} \\ &= \frac{1}{2} \,. \end{aligned}$$

Perhaps surprisingly, there is only a 50% chance that you are holding the two-headed coin. Yes, flipping ten heads in a row with a normal coin is very unlikely. But so is drawing the one two-headed coin from a jar of 1024 normal coins! In fact, as the math shows, both explanations for the data are equally unlikely, which is why we're left with a posterior probability of 0.5.

Two-headed coins in the wild. Let's return to the scenario of the trader knocking at your door, asking for a rise on the basis of a 10-month winning streak. In light of what you know about Bayes' rule, which of the following replies is the most sensible?

(A) "You're right. Here's a giant raise."

(B) "Thank you for letting me know. While I need more data to give you a raise, you've had a good ten months. I'll review your case again in 6 months and will look closely at the facts you've showed me."

The best answer depends very strongly on your beliefs about whether excellent stock traders are common or rare. For example, suppose you believe that 10% of all stock traders are truly excellent, in the sense that they can reliably finish with above-average returns, month after month; and that the other 90% just muddle through and collect their thoroughly average bonus checks. Then $P(T) = 0.1$, and

$$P(T \mid D) = \frac{0.1 \cdot 1}{0.1 \cdot 1 + 0.9 \cdot \frac{1}{1024}} \approx 0.991 \,,$$

so that there is better than a 99% chance that your employee is among those 10% of excellent performers. You should give her a raise, or risk letting some other bank save you the trouble.

What if, however, you believed that excellence were much rarer, say $P(T) = 1/10000$? In that case,

$$P(T \mid D) = \frac{0.0001 \cdot 1}{0.0001 \cdot 1 + 0.9999 \cdot \frac{1}{1024}} \approx 0.093 \,.$$

In this case, even though the ten-month hot streak was unusual—$P(D \mid \text{not } T)$ is small, at $1/1024$—there is still more than a 90% chance that your employee got lucky.

The moral of the story is that the prior probability in Bayes' rule—in this case, the baseline rate of excellent stock traders, or two-headed coins—plays a very important role in correctly estimating conditional probabilities. Ignoring this prior probability is a big mistake, and such a common one that it gets its own name: the base-rate fallacy.[7]

[7] en.wikipedia.org/wiki/Base_rate_fallacy

So just how rare are two-headed coins? While it's very difficult to know the answer to this question in something like stock-trading, it is worth pointing out one fact: in the above example, a prior probability of 10% is almost surely too large. Remember the NP rule: if this prior probability were right, then out of your office of 1025 traders, you would expect there to be $0.1 \times 1025 \approx 100$ of them with 10-month winning streaks, all at your door at once clamoring for a raise. (Traders are not known for being shy about their winning streaks, asking for raises, or anything else.) Since this hasn't happened, the prior probability $P(T) = 0.1$ is too high to be consistent with all the data available, and should be revised downward.

On the flip side, we also know that two-headed coins in stock-picking do exist, or else there would be no explanation for Warren Buffett, known as the "Oracle of Omaha." Over the last 50 years, Warren Buffett has beaten the market so consistently that it almost defies belief: between 1964 and 2013, the share price of his holding company, Berkshire Hathaway, rose by about 1 million percent, versus only 2300% for the S&P 500 stock index.

This line of reasoning demonstrates that, while the prior probability often reflects your own knowledge about the world, it can also be informed by data. Either way, the prior is very influential in real-word probability calculations, and should not be ignored.

Understanding Bayes' rule using trees

Let's try a second example of Bayes' rule in action.

> You are driving through unfamiliar territory in East Texas in your burnt-orange car sporting a bumper sticker from the University of Texas. You reach a fork in the road. In one direction lies College Station; in another direction, Austin. The road sign pointing to Austin has been stolen, but you see a man by the side of the road. You pull over and ask him for

directions.

You know that there are two kinds of people in this part of Texas: Longhorns and Aggies, with Aggies outnumbering Longhorns by a 60/40 margin. But you don't know which one this man is. If he's a Longhorn, he is sure to help you out to the best of his ability, and you judge that there is only a 5% chance that he will get confused and point you in the wrong direction. But you believe that, if he is an Aggie and you ask him for directions to some specific place, there is a 70% chance that he will see the bumper sticker on your car and send you the opposite way to wherever you ask.

You're clever, though, and so you decide to ask him "Which way is your university?" (Think for a minute about why that's the smart question to ask, rather than "Which way is UT?") He stares for a moment at your bumper sticker, then smiles and points to the left. You go in the direction indicated, and two hours later you arrive in Austin.

Given that the man has pointed you to Austin, what is the posterior probability that he was a Longhorn: $P(\text{Longhorn} \mid \text{pointed to Austin})$?

One way of solving this is to use Bayes' rule directly:

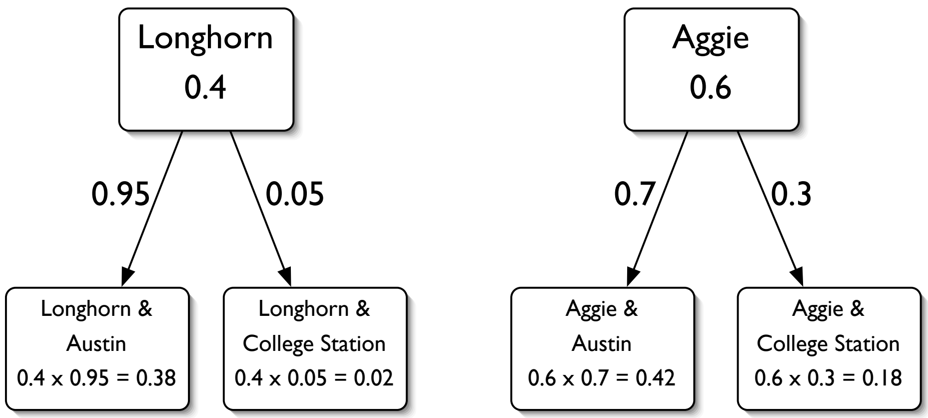
$$
\begin{aligned}
P(\text{Longhorn} \mid \text{pointed to Austin}) &= \frac{P(\text{Longhorn}) \cdot P(\text{pointed to Austin} \mid \text{Longhorn})}{P(\text{pointed to Austin})} \\
&= \frac{0.4 \cdot 0.95}{0.6 \cdot 0.7 + 0.4 \cdot 0.95} \\
&= 0.475\,.
\end{aligned}
$$

We used three facts here in the calculation. First, the prior probalities are $P(\text{Longhorn}) = 0.4)$, and therefore $P(\text{Aggie}) = 0.6$. Second, we know that $P(\text{pointed to Austin} \mid \text{Longhorn}) = 0.95$: he saw your bumper sticker, knew you were a Longhorn, and therefore had a 95% chance of pointing you towards his university (that is, UT). Finally, we know that $P(\text{pointed to Austin} \mid \text{Aggie}) = 0.7$: he saw your bumper sticker, knew you were a Longhorn, and therefore had a 70% chance of pointing you towards *away from* his university (that is, away from A&M and towards UT). Then we use the rule of total probability to calculate $P(\text{pointed to Austin})$.

So overall, there is slightly better than an even chance that you were talking to an Aggie. Bayes' rule gets us the answer with little fuss, by simplying plugging in the appropriate terms to the formula. But an alternative, very intuitive way of solving this problem—and of understanding Bayes' theorem more generally—is to use a tree.

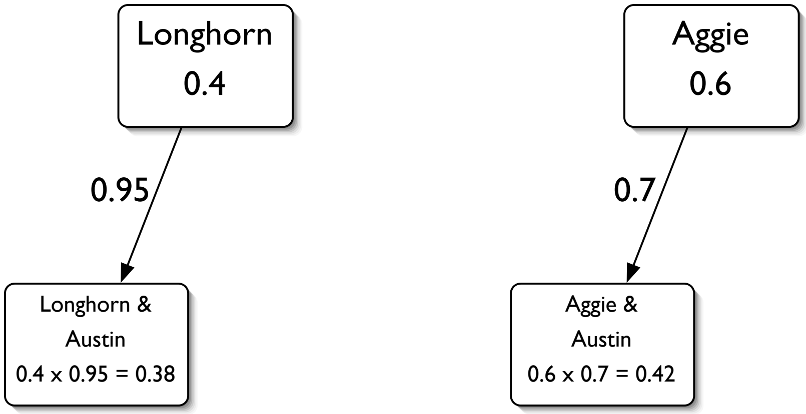
Let's see how this works. First, start by listing the possible states of the world, along with their probabilities. I like to put these in boxes:

Next, draw arrows from each state of the world to the possible observational consequences. Along the arrows, put the conditional probabilities that you will observe each data point, given the corresponding state of the world:

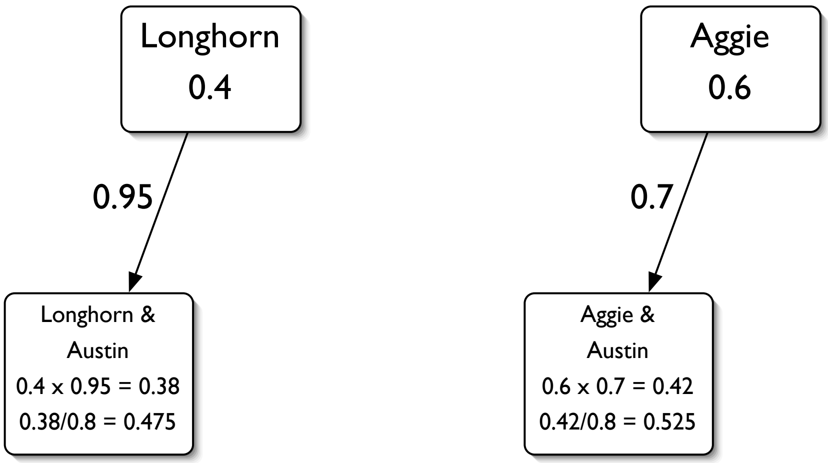


At the terminal leaves of the tree, multiply out the probabilities according to the multiplication rule: $P(A, B) = P(A) \cdot P(B \mid A)$. So, for example, the probability that the man is an Aggie and that he points you to Austin is $0.7 \times 0.6 = 0.42$. The sum of all the probabilities in the leaves of the tree must be 1, since they exhaust all possibilities.

But now that you've arrived back home, you know that the man was pointing to Austin. To use the tree to compute the probability that he was a Longhorn, simply cut off all the branches corresponding to data that wasn't actually observed, leaving only the actual data:



The remaining leaf probabilities are proportional to the posterior probabilities of the corresponding states of the world. These probabilities, of course, do not sum to 1 anymore. But this is easily fixed: simply divide each probability by the sum of all remaining probabilities (in this case, $0.38 + 0.42 = 0.8$):



As before, we find that there is a probability of 0.475 that the man at the watermelon stand was a Longhorn. Just so you can keep things straight, the sum of 0.8, by which you divided the terminal-leaf probabilities in the final step to ensure that they summed to 1, corresponds exactly to the denominator, P(points to Austin), in Bayes' rule.

Bayes' rule and the law

SUPPOSE you're serving on a jury in the city of New York, with a population of roughly 10 million people. A man stands before you accused of murder, and you are asked to judge whether he is guilty (G) or not guilty ($\sim G$). In his opening remarks, the prosecutor tells you that the defendant has been arrested on the strength of a single, overwhelming piece of evidence: that his DNA matched a sample of DNA taken from the scene of the crime. Let's call denote this evidence by the letter D. To convince you of the strength of this evidence, the prosecutor calls a forensic scientist to the stand, who testifies that the probability that an innocent person's DNA would match the sample found at the crime scene is only one in a million. The prosecution then rests its case.

Would you vote to convict this man?

If you answered "yes," you might want to reconsider! You are charged with assessing $P(G \mid D)$—that is, the probability that the defendant is guilty, given the information that his DNA matched the sample taken from the scene. Bayes' rule tells us that

$$P(G \mid D) = \frac{P(G) \cdot P(D \mid G)}{P(D)} = \frac{P(G) \cdot P(D \mid G)}{P(D \mid G) \cdot P(G) + P(D \mid \sim G) P(\sim G)} .$$

We know the following quantities:

- The prior probability of guilt, $P(G)$, is about one in 10 million. New York City has 10 million people, and one of them committed the crime.
- The probability of a false match, $P(D \mid \sim G)$, is one in a million, because the forensic scientist testied to this fact.

To use Bayes' rule, let's make one additional assumption: that the likelihood, $P(D \mid G)$, is equal to 1. This means we're assuming that, if the accused were guilty, there is a 100% chance of seeing a positive result from the DNA test.

Let's plug these numbers into Bayes' rule and see what we get:

$$\begin{aligned} P(G \mid D) &= \frac{\frac{1}{10{,}000{,}000} \cdot 1}{1 \cdot \frac{1}{10{,}000{,}000} + \frac{1}{1{,}000{,}000} \cdot \frac{9{,}999{,}999}{10{,}000{,}000}} \\ &\approx 0.09 . \end{aligned}$$

The probability of guilt looks to be only 9%! This result seems shocking in light of the scientist's claim that $P(D \mid \sim G)$ is so small:

a "one in a million chance" of a positive match for an innocent person. Yet the prior probability of guilt is very low—$P(G)$ is a mere one in 10 million—and so even very strong evidence still only gets us up to $P(G \mid D) = 0.09$.

Conflating $P(\sim G \mid D)$ with $P(D \mid \sim G)$ is a serious error in probabilistic reasoning. These two numbers are typically very different from one another, because conditional probabilities aren't symmetric, as we've said more than once. Getting this wrong in the context of a courtroom—that is, conflating $P(A \mid B)$ with $P(B \mid A)$—is so common that it has its own name: the prosecutor's fallacy.[8]

[8] en.wikipedia.org/wiki/Prosecutor's_fallacy

An alternate way of thinking about this result is the following. Of the 10 million innocent people in New York, ten would have DNA matches merely by chance. The one guilty person would also have a DNA match. Hence there are 11 people with a DNA match, only one of whom is guilty, and so $P(G \mid D) \approx 1/11$. Your intuition may mislead, but Bayes' rule never does!