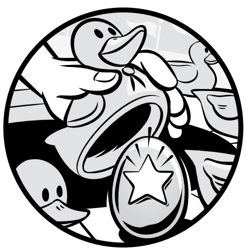
7

BAYES' THEOREM WITH LEGO



In the previous chapter, we covered conditional probability and arrived at a very important idea in probability, Bayes' theorem, which states:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Notice that here we've made a very small change from Chapter 6, writing $P(B|A)P(A)$ instead of $P(A)P(B|A)$; the meaning is identical, but sometimes changing the terms around can help clarify different approaches to problems.

With Bayes' theorem, we can reverse conditional probabilities—so when we know the probability $P(B|A)$, we can work out $P(A|B)$. Bayes' theorem is foundational to statistics because it allows us to go from having the probability of an observation given a belief to determining the strength of that belief given the observation. For example, if we know the probability of sneezing given that you have a cold, we can work backward to determine the probability that you have a cold given that you sneezed. In this way, we use evidence to update our beliefs about the world.

In this chapter, we'll use LEGO to visualize Bayes' theorem and help solidify the mathematics in your mind. To do this, let's pull out some LEGO bricks and put some concrete questions to our equation. Figure 7-1 shows a 6 × 10 area of LEGO bricks; that's a 60-stud area (*studs* are the cylindrical bumps on LEGO bricks that connect them to each other).

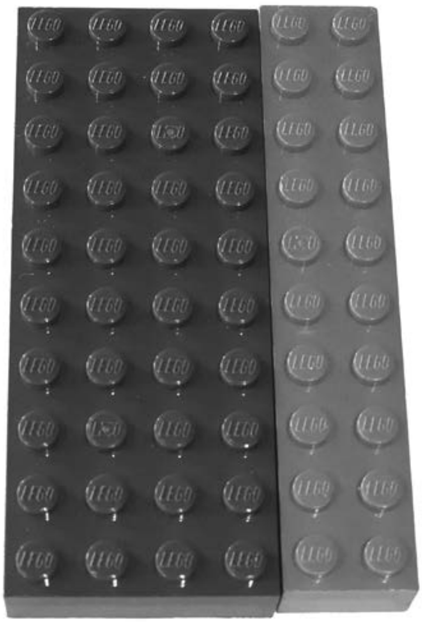


Figure 7-1: A 6 × 10-stud LEGO area to help us visualize the space of possible events

We can imagine this as the space of 60 possible, mutually exclusive events. For example, the blue studs could represent 40 students who passed an exam and the red studs 20 students who failed the exam in a class of 60. In the 60-stud area, there are 40 blue studs, so if we put our finger on a random spot, the probably of touching a blue brick is defined like this:

$$P(\text{blue}) = \frac{40}{60} = \frac{2}{3}$$

We would represent the probability of touching a red brick as follows:

$$P(\text{red}) = \frac{20}{60} = \frac{1}{3}$$

The probability of touching either a blue or a red brick, as you would expect, is 1:

$$P(\text{blue}) + P(\text{red}) = 1$$

This means that red and blue bricks alone can describe our entire set of possible events.

Now let's put a yellow brick on top of these two bricks to represent some other possibility—for example, the students that pulled an all-nighter studying and didn't sleep—so it looks like Figure 7-2.

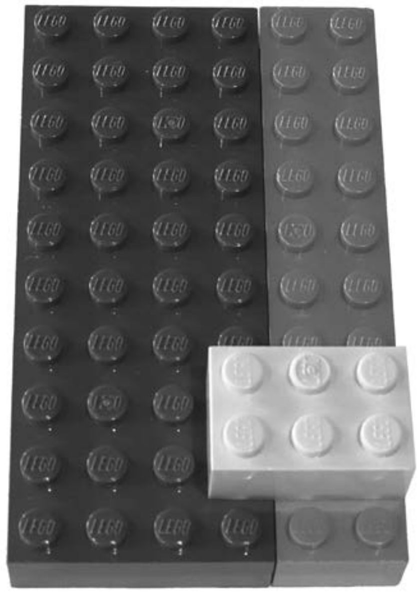


Figure 7-2: Placing a 2 × 3 LEGO brick on top of the 6 × 10-stud LEGO area

Now if we pick a stud at random, the probability of touching the yellow brick is:

$$P(\text{yellow}) = \frac{6}{60} = \frac{1}{10}$$

But if we add P(yellow) to P(red) + P(blue), we'd get a result greater than 1, and that's impossible!

The issue, of course, is that our yellow studs all sit on top of the space of red and blue studs, so the probability of getting a yellow brick is *conditional* on whether we're on a blue or red space. As we know from the previous chapter, we can express this conditional probability as $P(\text{yellow} \mid \text{red})$, or *the probability of yellow given red*. Given our example from earlier, this would be the probability that a student pulled an all-nighter, given that they had failed an exam.

Working Out Conditional Probabilities Visually

Let's go back to our LEGO bricks and work out $P(\text{yellow} \mid \text{red})$. Figure 7-3 gives us a bit of visual insight into the problem.

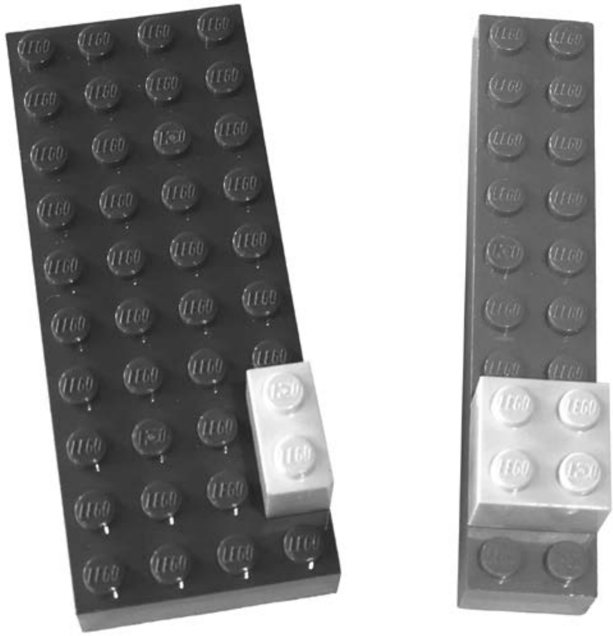


Figure 7-3: Visualizing P*(yellow | red)*

Let's walk through the process for determining $P(\text{yellow} \mid \text{red})$ by working with our physical representation:

1. Split the red section off from the blue.
2. Get the area of the entire red space; it's a 2 × 10-stud area, so that's 20 studs.
3. Get the area of the yellow block on the red space, which is 4 studs.
4. Divide the area of the yellow block by the area of the red block.

This gives us $P(\text{yellow} \mid \text{red}) = 4/20 = 1/5$.

Great—we have arrived at the conditional probability of yellow given red! So far, so good. So what if we now reverse that conditional probability and ask what is $P(\text{red} \mid \text{yellow})$? In plain English, if we know we are on a yellow space, what is the probability that it's red underneath? Or, in our test example, what is the probability that a student failed the exam, given that they pulled an all-nighter?

Looking at Figure 7-3, you may have intuitively figured out $P(\text{red} \mid \text{yellow})$ by reasoning, "There are 6 yellow studs, 4 of which are over red, so the probability of choosing a yellow that's over a red block is 4/6." If you did follow this line of thinking, then congratulations! You just independently discovered Bayes' theorem. But let's quantify that with math to make sure it's right.

Working Through the Math

Getting from our intuition to Bayes' theorem will require a bit of work. Let's begin formalizing our intuition by coming up with a way to *calculate* that there are 6 yellow studs. Our minds arrive at this conclusion through spatial reasoning, but we need to use a mathematical approach. To solve this, we just take the probability of being on a yellow stud multiplied by the total number of studs:

$$\text{numberOfYellowStuds} = P(\text{yellow}) \times \text{totalStuds} = \frac{1}{10} \times 60 = 6$$

The next part of our intuitive reasoning is that 4 of the yellow studs are over red, and this requires a bit more work to prove mathematically. First, we have to establish how many red studs there are; luckily, this is the same process as calculating yellow studs:

$$\text{numberOfRedStuds} = P(\text{red}) \times \text{totalStuds} = \frac{1}{3} \times 60 = 20$$

We've also already figured out the ratio of red studs covered by yellow as $P(\text{yellow} \mid \text{red})$. To make this a count—rather than a probability—we multiply it by the number of red studs that we just calculated:

$$\text{numberOfRedStuds} = P(\text{yellow} \mid \text{red}) \times \text{numberOfRedStuds} = \frac{1}{5} \times 20 = 4$$

Finally, we get the ratio of the red studs covered by yellow to the total number of yellow:

$$P(\text{red} \mid \text{yellow}) = \frac{\text{numberOfRedUnderYellow}}{\text{numberOfYellowStuds}} = \frac{4}{6} = \frac{2}{3}$$

This lines up with our intuitive analysis. However, it doesn't quite look like a Bayes' theorem equation, which should have the following structure:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

To get there we'll have to go back and expand the terms in this equation, like so:

$$P(\text{red} \mid \text{yellow}) = \frac{P(\text{yellow} \mid \text{red}) \times \text{numberOfRedStuds}}{P(\text{yellow}) \times \text{totalStuds}}$$

We know that we calculate this as follows:

$$P(\text{red} \mid \text{yellow}) = \frac{P(\text{yellow} \mid \text{red})P(\text{red}) \times \text{totalStuds}}{P(\text{yellow}) \times \text{totalStuds}}$$

Finally, we just need to cancel out totalStuds from the equation, which gives us:

$$P(\text{red} \mid \text{yellow}) = \frac{P(\text{yellow} \mid \text{red})P(\text{red})}{P(\text{yellow})}$$

From intuition, we have arrived back at Bayes' theorem!

Wrapping Up

Conceptually, Bayes' theorem follows from intuition, but that doesn't mean that the formalization of Bayes' theorem is obvious. The benefit of our mathematical work is that it extracts reason out of intuition. We've confirmed that our original, intuitive beliefs are consistent, and now we have a powerful new tool to deal with problems in probability that are more complicated than LEGO bricks.

In the next chapter, we'll take a look at how to use Bayes' theorem to reason about and update our beliefs using data.

Exercises

Try answering the following questions to see if you have a solid understanding of how we can use Bayes' Theorem to reason about conditional probabilities. The solutions can be found at *https://nostarch.com/learnbayes/*.

1. Kansas City, despite its name, sits on the border of two US states: Missouri and Kansas. The Kansas City metropolitan area consists of 15 counties, 9 in Missouri and 6 in Kansas. The entire state of Kansas has 105 counties and Missouri has 114. Use Bayes' theorem to calculate the probability that a relative who just moved to a county in the Kansas City metropolitan area also lives in a county in Kansas. Make sure to show $P(\text{Kansas})$ (assuming your relative either lives in Kansas or Missouri), $P(\text{Kansas City metropolitan area})$, and $P(\text{Kansas City metropolitan area} \mid \text{Kansas})$.
2. A deck of cards has 52 cards with suits that are either red or black. There are four aces in a deck of cards: two red and two black. You remove a red ace from the deck and shuffle the cards. Your friend pulls a black card. What is the probability that it is an ace?