Bayes’ Theorem: Part II

September 28, 2022

Note: This is the second part of a two-lecture series on Bayes’ Theorem–this half will finish up some of the things we did not get to last week, discuss exactly what conditional probability even means, and give a “proof with intuition” of Bayes’ Theorem.

Section -1: Bayes’ Theorem

Write Bayes’ Theorem below:

Now for some practice with the theorem that we didn’t get to last week. Try solving each of the following problems twice: once intuitively and once by explicitly using Bayes’ Theorem, and note that both methods generally end up doing the same thing:

Problem -1.1: You have two bags of marbles, one with 7 red and 3 blues marbles, and the other with 3 red and no blue marbles. You choose a bag randomly, then you draw a marble from the bag at random. The marble is blue. If you were to draw another marble from the same bag, what is the probability that it would be blue?

Problem -1.2: You have two bags of marbles, one with 7 red and 3 blues marbles, and the other with 3 red and 5 blue marbles. You choose a bag randomly, then you draw a marble from the bag at random. The marble is blue. If you were to draw another marble from the same bag, what is the probability that it would be blue?

Problem -1.3: A certain test for J-21315973 has a sensitivity of 90% and a specificity of 95%. Exactly 10% of people in your region of Mathland *test positive* for the disease. Given that you tested positive with this test, what is the probability that you actually have J-21315973?

Section 0: Remember Last Week?

Now, with our new Bayesian machinery, we can finally answer the question posed in Section 0 last week. First, let us set up our hypothesis and evidence.

Our hypothesis, H, should be that a given person in the Bay Area is the perpetrator of the crime, and our evidence, E, is that the given person (from the Bay Area) is 6 feet 6 inches tall, owns vintage edition Groucho Marx glasses, and owns a three-foot-tall tophat.

Now, let us compute. What is $P(H)$? Well, $P(H)$ should just be the probability a given person in the Bay Area is the perpetrator of the crime. Now there are approximately 8 million people in the Bay Area, and only one of them is the perpetrator. So $P(H)$ is $\frac{1}{8,000,000}$.

What about $P(E)$? Well, the Bureau of Strange Statistics tells us that approximately one in five million people are 6 feet 6 inches tall and own both vintage edition Groucho Marx glasses and a three-foot-tall tophat, so the odds that any person in the Bay Area satisfies these conditions, in the absence of other information, should be $\frac{1}{5,000,000}$.

And what about $P(E|H)$? Well, we know that the perpetrator is 6 feet 6 inches tall and own both vintage edition Groucho Marx glasses and a three-foot-tall tophat, so $P(E|H)$ should just be 1.

Putting it all together using Bayes' Theorem, we find that $P(H|E) = \frac{P(E|H)*P(H)}{P(E)} = \frac{\frac{1}{8.000.000}\cdot 1}{\frac{1}{5,000,000}} = \frac{5,000,000}{8,000,000} = \frac{5}{8}$. So since James Smith satisfies the evidence criteria and is in the Bay Area, the probability that he is the real perpetrator is $\frac{5}{8}$, which is a lot less than the $\frac{4,999,999}{5,000,000}$ that a certain Los Angeles jury might assume.

So Bayes' Theorem comes in to save the day, and we have a $\frac{5}{8}$ probability that James Smith is actually the handout thief… Just kidding, there's something else that we missed. We'll go over what's going on later, but first, onwards to conditional probability!

Section 1: The Intuition of "Given"

In our last class, we used the word "given" a lot. It's time to explain what that actually means.

In the following problems, we will have three bags: bag A, bag B, and bag C. Bag A contains 1 red and 2 blue marbles, bag B contains 2 red and 3 blue marbles, and bag C contains 3 red and no blue marbles.

In addition, R will be the event that a red marble is picked, L will be the event that a blue marble is picked, D will be the event that bag A is chosen, E will be the event that bag B is chosen, and F will be the event that bag C is chosen.

> Problem 1.1: You pick a random bag from bags A, B, and C, then pick a random marble from that bag. What is the probability that the marble you pick is red? In other words, compute P(R).

> Problem 1.2: You pick a random bag from bags A, B, and C, then pick a random marble from that bag. Given that you picked bag B, what is the probability that the marble you pick is red? In other words, compute P(R|E).

One way to solve a problem such as Problem 1.1 is to create a "space of possible outcomes," and then just count up all the probabilities where the event R happens. But what about a problem like Problem 1.2?

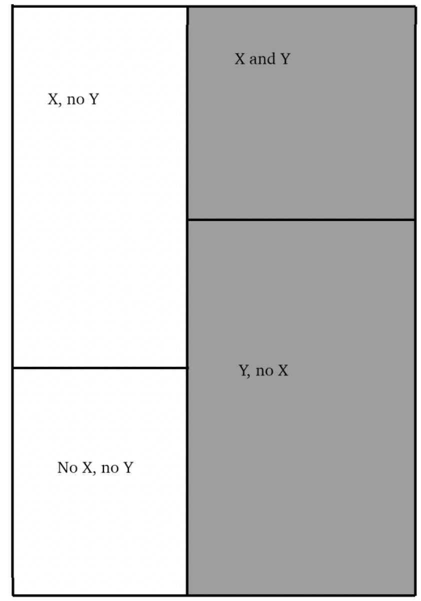
> Problem 1.3: Make a diagram of all possible outcomes of our bag-and-marble picking scenario. Use it to answer Problems 1.1 and 1.2.

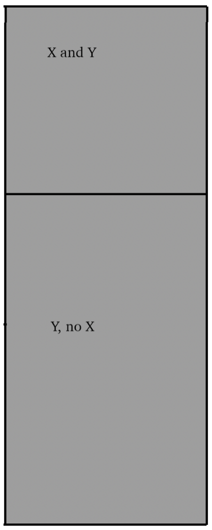
So one way to think about the answer to Problem 1.2 is that P(R|E) is 0.4, because when event E happens (you pick from bag B), there is a 0.4 chance that event R happens (a red marble is picked).

That is to say, R happens in 0.4 of the instances where E happens.

In other words, the probability of R and E both happening is 0.4 times the probability where only E happens.

On the other hand, there's nothing special about the events R and E. Indeed, this logic works for any events X and Y! This is because P(X|Y) asks for the probability that X happens only in the instances where Y occurs; in other words, the probability of X given Y is the probability that X happens when you restrict yourself to only the cases where Y occurs.





To see how this concludes, let the diagram on the right be a diagram representing the probabilities of X and Y, where the area of each region is proportional to the probability that the events (or lack of events) on the region occur.

It should be fairly clear that the probability that Y occurs is the ratio of the area of the shaded region to the area of the entire region.

Now, take a look at what happens when you restrict yourself to only the cases where Y happens, as shown on the left. Now the probability that X occurs, restricted to the cases where Y happens, should be the ratio of the area of the "X and Y" box to the area of the entire shaded region.

However, note that P(X and Y) is the ratio of the area of the "X and Y" box to the area of the entire region, and P(Y) is the ratio of the area of the shaded region to the area of the entire region. So P(X and Y) divided by P(Y) should be equal to the ratio of the area of the "X and Y" box to the area of the entire shaded region.

On the other hand, this ratio is just the probability that X occurs, restricted to the cases where Y happens, which is the same as P(X|Y).

This gives rise to the formula for conditional probability, which you should write below:

Another way to look at this formula is to rearrange it a bit. Consider the following few problems:

> Problem 1.3: You pick a random bag from bags A, B, and C, then pick a random marble from that bag. What is the probability that you picked bag A? In other words, compute P(D).
>
> Problem 1.4: Given that you picked bag A, what is the probability that the marble you pick is red? In other words, compute P(R|D).
>
> Problem 1.5: What is the probability that the marble you picked was red, and that you picked bag A? In other words, compute P(R and D).

Just from the way Problem 1.5 is solved, we find that P(X and D)=P(D)*P(X|D) must hold. Indeed, P(X and Y)=P(Y)*P(X|Y) for any events X and Y, and rearranging we get our conditional probability formula.

Section 2: Back to Bayes-ics

For this section, we have the same setup (bags A through C, event names) as Section 1.

I would recommend you solve each of these next three problems intuitively, and then think about why your answer lines up with the conditional probability formula.

> Problem 2.1: You pick a random bag from bags A, B, and C, then pick a random marble from that bag. Given that you picked bag A, what is the probability that the marble you pick is blue? In other words, compute P(L|D).
>
> Problem 2.2: You pick a random bag from bags A, B, and C, then pick a random marble from that bag. Given that you did *not* pick bag B, what is the probability that the marble you pick is red? In other words, compute P(R|not E).

Problem 2.3: You draw two marbles from bag B, randomly and without replacement. Given that the first marble was red, what is the probability that both marbles are the same color?

Problem 2.4: You draw two marbles from bag A, randomly and without replacement. Given that the first marble was blue, what is the probability that both marbles are different colors?

Alright, these problems so far have had the "given" condition occur *before* the other condition. What if we flipped that?

Problem 2.5: You pick a random bag from bags A, B, and C, then pick a random marble from that bag. Given that you drew a blue marble, what is the probability that you picked bag A? In other words, compute P(D|L).

And just like that, we're back to Bayes' Theorem. Why? Well, just take a look at how you most likely computed P(D and L) there.

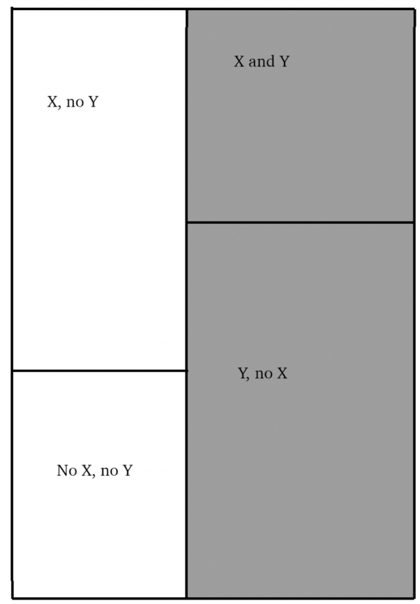
Problem 2.6: From just the conditional probability formula, prove Bayes' Theorem.

So Bayes' Theorem is bayes-ically just telling you the conditional probability formula, and then giving a method to compute the numerator.

The reason Bayes' Theorem might feel less intuitive even than the conditional probability formula is that it applies conditional probability in a direction that might feel less familiar. Normally, with conditional probability, you have some events X and Y, and you compute P(X|Y), but Y occurs *after* X or is in some way *affected by* X, which makes the "given" feel intuitively sensible. Whereas in Bayes' Theorem, you're typically calculating the probability of some hypothesis based on some evidence, where the evidence, the *given* condition, is the thing affected by the hypothesis. Here is an example, and not how different it feels from Problem 2.3:

Problem 2.7: You draw two marbles from bag B, randomly and without replacement. Given that the *second* marble was red, what is the probability that both marbles are the same color?

The math doesn't care that Problem 2.7 feels more like inference and while Problem 2.3 feels more direct, but it is worth exploring why the conditional probability theorem applies even when it feels the given condition is the dependent one.



Maybe the best explanation for this phenomenon is the diagram used in our initial justification of conditional probability. Note that this diagram absolutely does not care whether X or Y came "first" or which is dependent on the other, so the logic we used earlier to justify the formula should still apply.

It is also easier to justify with concrete numbers, such as with the false positives cases or the following problems:

Problem 2.8: In Mathland (population 10,000), there are nine times as many farmers as librarians. 25% of the farmers wear glasses, and 50% of the librarians wear glasses. Given that a random mathlandian citizen wears glasses, what is the probability that they are a librarian?

Problem 2.9: In Mathland (population 10,000), there are 19 times as many farmers as librarians. 25% of the farmers wear glasses, and 27% of citizens in total wear glasses. Given that a random mathlandian citizen wears glasses, what is the probability that they are a librarian?

The principles behind these problems carry over to cases where things are less discrete, but it's definitely more apparent why Bayes' Theorem works when we're dealing with actual numbers of people; choosing a random outcome in a more continuous problem is slightly less intuitive than choosing a random person in a problem like 2.8 or 2.9.

Section 3 (if time): Monty Hall with Bayes

Let's go back to last week's Monty Hall problems, but now use Bayes to figure them out! For a reminder, here are the problems:

> Problem 3.1 (Monty Hall): You are in a game show, trying to win a car. In this game show, there are three doors, one on the left, one in the middle, and one on the right. Behind one door is a sports car, and behind the other two doors is nothing. You are allowed to choose a door, and you choose the door on the right. The game show host then says "now, I will open one of the other two doors with nothing behind it," and opens the door on the left, revealing nothing. You are allowed to switch your guess to the unopened middle door. What is the probability that the middle door has the sports car behind it?
>
> Problem 3.2 (Monty Hall Variant): In the same scenario as before, you choose the door on the right, but the game show host then says "now, I will open the leftmost of the remaining two doors with nothing behind it," and opens the door on the left, revealing nothing. You are allowed to switch your guess to the unopened middle door. What is the probability that the middle door has the sports car behind it?
>
> Problem 3.3 (Another Monty Hall Variant): In the same scenario as before, you choose the door on the right, but the game show host then says "now, I will choose one of the two other doors at random and open it," and opens the door on the left, revealing nothing. You are allowed to switch your guess to the unopened middle door. What is the probability that the middle door has the sports car behind it?

Bayes' Theorem can be pretty helpful with answering questions that have difficult-to-intuit answers, like the Monty Hall problems. Indeed, the next question might seem daunting–almost impossible, even–but Bayes' Theorem answers it quite nicely.

Problem 3.4 (Monty Chaos): You are in a game show, trying to win a car. In this game show, there are four doors, labeled “1,” “2,” “3,” and “4”. Behind one door is a sports car, and behind the other three doors is nothing. You are allowed to choose a door, and you choose door “4.” The game show host then says “now, I will open one of the other three doors with nothing behind it. Here is how I will choose how to pick:

-If door 1 has the sports car, I will open door 2 or door 3 each with probability 1/2.

-If door 2 has the sports car, I will open door 1 with probability 2/3, or door 3 with probability 1/3.

-If door 3 has the sports car, I will open door 1 with probability 3/4, or door 2 with probability 1/4.

-If door 4 has the sports car, I will open door 1 with probability 7/12, door 2 with probability 1/4, and door 3 with probability 1/6.”

Then, she opens door 1, revealing nothing. You are allowed to switch your guess to the unopened middle door. What is the probability that each door has the sports car behind it?

Section 4: Back to the Original Problem

Alright, now with our new experience with Bayes’ Theorem, let’s figure out what the exact odds that James Smith was the killer are.

Problem 4.1: Figure out the error in the logic of the “solution” in section 0. (Hint: there’s a very subtle piece of information that was left out).