Contents

Chapter 1

Conditional Probability

Let's say there is a virus going around, and there is a vaccine for it that requires 2 shots. You are working at a school, and you are wondering how effective the vaccines are. Some students are unvaccinated, some have had one shot, and some have had two shots. One day you test all 644 students to see who has the virus. You end up with the following table:

	V_0	V_1	V_2
T_+	88 students	36 students	96 students
T_-	92 students	76 students	256 students

Here are what the symbols mean:

- V_0: student has had zero vaccination shots
- V_1: student has had one vaccination shot
- V_2: student has had both vaccination shots
- T_+: student tested positive for the virus
- T_-: student tested negative for the virus

So, for example, your data indicates that 76 students who had only one of the two shots and tested negative for the virus.

Your principal has a few questions. The first is "If I put five randomly chosen students in a study group together, what is the probabiltiy that one of them has the virus?"

The first thing you might do is make a new table that shows what is the probability of a randomly chosen student being in any particular group. You just divide each entry by 644 (the total number of students).

	V_0	V_1	V_2
T_+	$p(V_0 \text{ AND } T_+) = 13.7\%$	$p(V_1 \text{ AND } T_+) = 5.6\%$	$p(V_2 \text{ AND } T_+) = 14.9\%$
T_-	$p(V_0 \text{ AND } T_-) = 14.3\%$	$p(V_1 \text{ AND } T_-) = 11.8\%$	$p(V_2 \text{ AND } T_-) = 39.8\%$

(In this table, I expressed the number as a percentage with a decimal point – you had to round off the numbers. If you wanted exact answers, you would have to keep each as a fraction: 36 students represents $\frac{9}{161}$ of the student body.)

1.1 Marginalization

Now we can sum across the columns and rows.

	V_0	V_1	V_2	sum
T_+	0.137	0.056	0.149	$p(T_+) = 0.342$
T_-	0.143	0.118	0.398	$p(T_+) = 0.547$
sum	$p(V_0) = 0.280$	$p(V_1) = 0.174$	$p(V_2) = 0.547$	

If a child is chosen randomly from the entire student body, there is a 34.2% that the student has tested positive for the virus. And there is 17.4% chance that the student has one shot of the vaccine.

This summing of the probabilities across one dimension is known as *marginalizing*. Marginalization is just summing across all the variables that you don't care about. You don't care who has the virus, just the probability that a student has not received even one shot of the vaccine? You marginalize all the vaccine statuses.

To answer the principal's question, the easy thing to do is find the answer of the opposite "if I put five randomly chosen students in a study group together, what is the probabiltiy that *none* of them has tested positive for the virus?"

The chance that a randomly chosen student doesn't have the virus $(p(T_-)$ is 54.7%. Thus the chance that 5 randomly chosen students don't have the virus is $0.547 \times 0.547 \times 0.547 \times 0.547 \times 0.547 = 0.0489$ Thus the probability of the opposite is $1.0 - 0.0489 = 0.951$

The answer, then, is "If you put 5 kids in a study group together, there is a 95.1 % probability that at least one of them has the virus."

1.2 Conditional Probability

Now the principal asks you, "What if I make a group of 5 kids who have had both shots of the vaccine? What are the odds that one of them has tested positive for the virus?"

This involves the idea of *Conditional probability*. You want to know the odds that a student doesn't have the virus given that the student has had both shots of the vaccine.

There is a mathematical notation for this:

$$p(T_-|V_2)$$

That is the probability that a student who has had both vaccination shots will test negative for the virus.

How would you calculate this? You would count all the students who had a positive test *and* both vaccination shots, which you would divide by the total number of students who had both vaccination shots.

$$p(T_-|V_2) = \frac{256}{96 + 256} = \frac{8}{11} \approx 72.7\%$$

If we are working from the probabilities, you can get the same result this way: Divide the probability that a randomly chosen student had a positive test *and* both vaccination shots by the probabiltiy that a student had both vaccination shots:

$$p(T_-|V_2) = \frac{p(T_- \text{ AND } V_2)}{p(T_-)} = \frac{0.398}{0.547} \approx 72.7\%$$

Notice that this is different from $p(V_2|T_-)$, which is the probability that a student has had both vaccinations, given they tested negative for the virus.

Back to the principal's question: "If you have 5 students who have had both vaccinations, what is the probability that all of them tested negative for the virus?" The probability that one student is virus-free is $\frac{8}{11}$, so the probability that 5 students are virus-free is $\frac{8}{11}^5 \approx 0.203$. So, there is a 79.6% chance that at least one of the five has the virus.

1.3 Chain Rule for Probability

You just used this equality: For any events A and B

$$p(A|B) = \frac{p(A \text{ AND } B)}{p(B)}$$

This is more commonly written like this:

$$p(A \text{ AND } B) = \frac{p(A|B)}{p(B)}$$

This is an abstract way of writing the idea, but the idea itself is pretty intuitive: The probability that I'm going to buy and ticket and win the lottery is equal to the probability that I buy a ticket times the probability that I win, given that I have bought a ticket. (Here A is "win the lottery" and B is "buy a ticket".)

This is known as *The Chain Rule of Probability*. And we can chain together as many events as we want: The probability that you are going to die in the car that you bought with your winnings from the lottery ticket you bought is:

$$p(W \text{ AND } X \text{ AND } Y \text{ AND } Z) = p(W|X \text{ AND } Y \text{ AND } Z)p(X|Y \text{ AND } Z)p(Y|Z)p(Z)$$

where

- W = Dying in car accident
- X = Buying a car with lottery winnings
- Y = Winning the lottery
- Z = Buying a lottery ticket

In English, then, the equation says:

"The probability that you will die in a car accident, buy a car with lottery winnings, win the lottery, and buy a lottery ticket is equal to the probability that you buy a lottery ticket times the probability that you win the lottery (given that you have bought a ticket) times the probability that buy a car with those lottery winnings (given that bought a ticket and won) times the probability that you crash that car (given that you have bought the car, won the lottery, and bought a ticket)."

Chapter 2

Bayes' Theorem

Let's say that you are holding two bags of marbles. You know that one bag contains 60 white marbles and 40 red marbles. And you know that the other holds 10 white marbles and 90 red marbles. You don't know which is which – and you can't see the marbles.

I say "Guess which bag is mostly red marbles." You pick one.

"What is the probability that this is the bag that is mostly red marbles?" You think "50 percent and there is also a 50 percent probabilitiy that it is the mostly-white-marbles bag."

Then you pick one marble from the bag. It is red. Now you must update your beliefs. It is more likely that this is the mostly-red-marbles bag. What is the probability now?

Bayes Theorem gives you the rule for updating your beliefs based on new data.

2.1 Bayes Theorem

Let's say you have two events or conditions C and D. C is "The person has a cough" and D is "The person is waiting to see a doctor."

Using the chain rule of probability, we now have two ways to calculate $p(C \text{ AND } D)$:

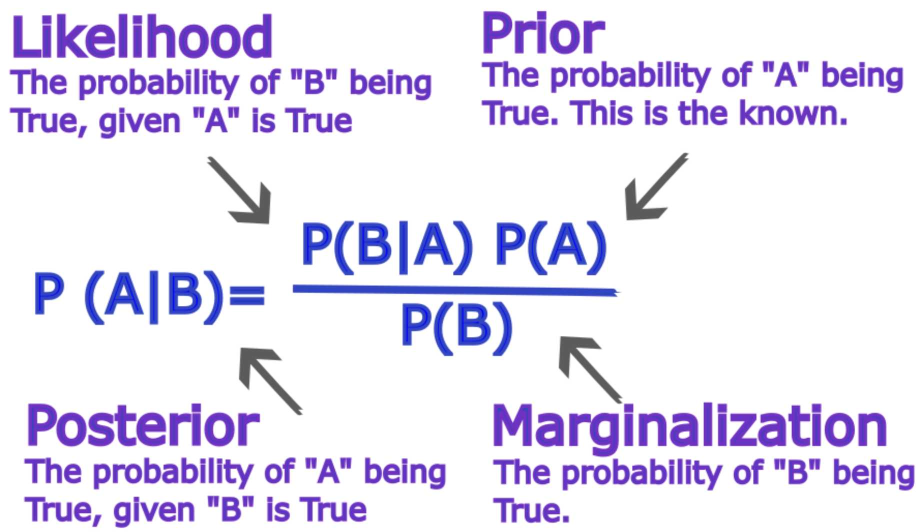
$$p(C \text{ AND } D) = p(C|D)p(D)$$

(The probability the person is at the doctor multiplied by the probability they have a cough if they are at the doctor.)

or

$$p(C \text{ AND } D) = p(C|D)p(D)$$

(The probabilitiy the person has a cough multiplied by the probabilitiy they are at the doc-



tor if they have a cough.)

Thus:

$$p(D|C) = \frac{p(C|D)p(D)}{P(C)}$$

Now you can calculate $p(D|C)$ (in this case, the probability that you are waiting to see a doctor given that you have a cough.) if you know:

- $p(C|D)$ (The probability that you have a cough given that you are waiting to see a doctor)
- $p(D)$ (The probability that you are waiting for a doctor for any reason.)
- $p(C)$ (The probability that you have a cough anywhere)

Pretty much all modern statistical methods (including most artificial intelligence) are based on this formula, which is known as Bayes' Theorem. It was written down by Thomas Bayes before he died in 1761. It was then found and published after his death.

2.2 Using Bayes' Theorem

Back to the example at the beginning. To review:

- There are two bags that look exactly the same.
- Bag W has 60 white marbles and 40 red marbles.

- Bag R has 10 white marbles and 90 red marbles.
- You pull one marble from the selected bag – it is red.

What is the probability that the selected bag is Bag R? Intuitively, you know that the probability is now more than 0.5. What is the exact number?

In terms of conditional probability, we say we are looking for "the probability that the selected bag is Bag R, given that you drew a red marble?" or $p(B_R|D_R)$, where B_R is "The selected bag is Bag R" and D_R is "You drew a red marble from the selected bag".

From Bayes' Theorem, we can write:

$$p(B_R|D_R) = \frac{P(D_R|B_R)P(B_R)}{P(D_R)}$$

$P(D_R|B_R)$ is just the probability of drawing a red marble given that the selected bag is Bag R. That is easy to calculate: There are 100 marbles in the bag, and 90 are red. Thus $P(D_R|B_R) = 0.9$.

$P(B_R)$ is just the probability that you chose Bag R before you drew out a marble. Both bags look the same, so $P(B_R) = 0.5$. This is called *the prior* because it represents what you thought the probability was before you got more information.

$P(D_R)$ is the probability of drawing a red marble. There was 0.5 probability that you put your hand into Bag W (in which 40 of the 100 marbles are red) and a 0.5 probability that you put your hand into Bag R (in which 90 of the 100 marbles are red). So

$$P(D_R) = 0.5\frac{40}{100} + 0.5\frac{90}{100} = 0.65$$

Putting it together:

$$p(B_R|D_R) = \frac{P(D_R|B_R)P(B_R)}{P(D_R)} = \frac{(0.9)(0.5)}{0.65} = \frac{9}{13} \approx 0.69$$

Thus, given that you have pulled a red marble, there is about a 69% chance that you have selected the bag with 90 red marbles.

2.3 Confidence

Bayes' Theorem, then, is about updating your beliefs based on evidence. Before you drew out the red marble, you selected one bag thinking it might contain 90 red marbles. How certain were you? 0.0 being complete disbelief and 1.0 entirely confidence, you were 0.5. After pulling out the red marble, you were about 0.69 confident that you had chosen the bag with 90 red marbles.

The question "How confident are you in your guess?" is very important in some situations. For example in medicine, diagnoses often lead to risky interventions. Few diagnoses come with 100% confidence. All doctors should know how to use Bayes' Theorem.

And in a trial, a jury is asked to determine if the accused person is guilty of a crime. Few jurors are ever 100% certain. In some trials, Bayes' Theorem is a really important tool.

Appendix A

Answers to Exercises