Slide 1 – Title Slide

Hello and welcome to Week 8, Part 1 of EGM101: Bayes' Theorem. In this lesson, we'll learn about one of the other most important theorems in probability, Bayes' Theorem, which tells us how to relate different conditional probabilities.

Slide 2 – Week 8 Outline

Over the remaining lessons this week, we'll look at various forms of hypothesis testing – how we can use inferential statistics to make conclusions about our data.

Slide 3 – The ~~Cookie~~ Biscuit Problem

Let's say that I have two bowls of... biscuits, and in each bowl, I have a miof two different biscuits: chocolate chip and custard crème. The breakdown for the number of each type of biscuit in each bowl is shown here in the table – we have 40 biscuits in each bowl. In the first bowl, we have 30 chocolate chip and 10 custard crème, and in the second bowl, we have an even split.

This is a contingency table, and we can use this to calculate different probabilities. If we sample a biscuit at random, without looking at which bowl we reach into, what is the probability that we get a chocolate chip biscuit? This is the number of chocolate chip biscuits, divided by the total number of biscuits; plugging in the numbers, we have 50 chocolate chip cookies and 80 biscuits total, so the probability is 0.625.

We can also calculate the probability of getting a custard crème – this is the number of custard crème biscuits, divided by the total number of biscuits, or 30 divided by 80; or 0.375.

And, we can calculate conditional probabilities, as well. If we reach into bowl one and pull out a biscuit, what are the chances that it is a chocolate chip biscuit? As we saw previously, this is the number of chocolate chip biscuits in bowl 1, divided by the number of biscuits in bowl 1 – plugging in the numbers, we have 30 over 40, or 0.75.

Now, let's ask a slightly different question. Let's assume that we randomly sample from a bowl without looking, and we get a custard crème. What is the probability it came from either bowl?

In other words, now we want to know – what is the probability that we chose Bowl 1, given that we selected a custard crème? And what is the probability that we chose Bowl 2, given that we selected a custard crème?

Slide 4 – The ~~Cookie~~ Biscuit Problem (cont.)

Now, obviously, we can easily calculate this from the contingency table. The probability that we chose from Bowl 1, given that we selected a custard crème, is equal to the number of custard crème biscuits in Bowl 1, divided by the total number of custard crème biscuits. Plugging in the numbers, we have 10 over 30, or 0.333.

Likewise, the probability that we chose from Bowl 2, given that we selected a custard crème, is equal to the number of custard crème biscuits in Bowl 2, divided by the total number of custard crème biscuits. Plugging in the numbers, we have 20 over 30, or approximately 0.667.

But, we might not always have a contingency table – we may only have some of the information that we need to answer the question. In that case, what we want to know is, how can we relate these different conditional probabilities?

Slide 5 – Conditional Probabilities, Revisited

Let's start with two different events, A and B. Then, let's assume that event B has happened. What is the probability of event A happening? In other words, what is the probability of event A, given event B? Earlier this week, we saw that we can calculate this conditional probability as the probability of an outcome being in both A and B, that is, the probability of the intersection of A and B, divided by the probability of event B.

From the Venn diagram, we can see that this is the same as the proportion of the intersection of events A and B within event B.

Similarly, we know that the probability of event B, given event A, is equal to the probability of the intersection of events B and A, divided by the probability of event A – in other words, the probability of B given A is just the proportion of the intersection of events B and A within event A.

We also remember, and we can see from the Venn diagram, that the intersection of A and B is equal to the intersection of B and A. This is because intersection is commutative – the order doesn't matter. This means that the probability of the intersection of A and B is equal to the probability of the intersection of B and A.

And so, we can re-arrange these two equations and put them together. If we multiply this equation by the probability of event A, and this equation by the probability of event B, and we use the fact that the probability of the intersection has to be the same, we end up with this equation. That is, the probability of event A given event B, multiplied by the probability of event B, is equal to the probability of event B given event A, multiplied by the probability of event A.

Slide 6 – Bayes' Theorem

This is very nearly the most common form of Bayes' Theorem that you will see – we just have to divide by the probability of event B. In this form, Bayes' Theorem tells us how conditional probabilities of two events are related.

You may also see this written or talked about in terms of a "hypothesis", H, and "data", or "evidence", D – in which case, Bayes' Theorem helps us answer the question, "what is the probability of the hypothesis, given the evidence?" This is the most common interpretation in something called "Bayesian inference" – we don't have the time to go into any details about this, but there are some links at the end of the presentation that you can follow to learn a bit more about this.

As you read more about Bayesian statistics, you will very often each of the components of this formula referred to by other names, beginning with the posterior probability, the probability of A given B. This is called the "posterior" probability, because it tells us the likelihood of event A after event B has happened. The contrasting probability is the likelihood – the probability of event B, given that event A occurs. Or, the probability of observing B, given A.

We also have the "prior" probability – this is the probability of event A, regardless of event B. This is called the "prior" probability because it is the probability of event A before event B has been observed. The final part of the formula here, the probability of event B, is known as the "marginal likelihood" – this is the probability that event B occurs, regardless of whether event A happens.

Slide 7 – Testing and Type I, Type II Errors

Later this week, we're going to talk about hypothesis testing – that is, the method by which we use inferential statistics to make conclusions about our observations. Before we get there, though, we should talk about what a "test" is – most often, we're testing whether or not a condition is positive. For example, if you take a test for an illness, the test is checking for indications that you do indeed have the illness. If the test is positive, we interpret this as the condition being present, and if it is negative, we interpret this as the condition not being present.

Tests can be incorrect, though – this contingency table helps illustrate how. The columns of this test tell us the "true" situation – that is, whether or not the condition is actually present, while the rows tell us the results of the test – whether the test says that the condition is present or not.

The sensitivity of the test is the probability of getting a true positive – that is, the probability that the test is positive, given that the condition is actually present. A type I error happens when the test says incorrectly says that the condition is present – this is also known as a false positive.

The specificity of the test is the probability of getting a true negative – that is, the probability that the test is negative, given that the condition is not actually present. A type II error happens when the test incorrectly says that the condition is not present – also known as a false negative.

As we will see on the next slide, the other important thing to consider with testing is the prevalence – that is, the likelihood of someone having the condition. We will cover this in even more detail this week in the context of hypothesis testing, but this is enough to get started for now.

Slide 8 – The Prosecutor's Fallacy

We can combine what we know about Bayes' Theorem, and what we just learned about Type I and Type II errors, to see how we can evaluate different claims of evidence – in this example, in the context of a crime being committed.

Let's assume that you have a tin full of your favorite biscuits, and you walk into the room where you keep it and find it empty – only crumbs left behind. And, let's say for the sake of argument, that you dust for fingerprints, find one, and you send it off to the police to have it matched. And, predictably, the fingerprint that you found matches my fingerprint.

Before we determine that I'm clearly guilty, let's look at the details of the fingerprint match – in effect, this is a test. The sensitivity of the fingerprint test – the probability that the test correctly identifies the fingerprint of the guilty party, is 1 in 1,000,000. That is, if we get 1,000,000 matches, we would expect that 999,999 of these are true positives.

The specificity of the test is also high. Remember that the specificity is the probability of getting a negative match, given that the person matching is not guilty. So, if we test 1,000,000 fingerprints, we would expect one false match. And finally, we need to know the number of people who are in the database that we're matching. In this case, we have 10,000,000 people whose fingerprints we're matching.

Now, the question we need to answer is: given that we have a positive match to my fingerprint, what are the chances that I am truly guilty? That is, what is the probability of guilt, given a positive result?

I will spoil the answer by showing the contingency table here – for simplicity, we're going to assume that the culprit acted alone, meaning we have only one guilty party. The second row of the table here is the one that matters, though – if we matched against 10,000,000 fingerprints, with a specificity of 1 in 1,000,000, we would expect to see 10 false positives. That is, we would expect to see 10 incorrect matches. But now look – the probability that I am guilty, given a positive result, is actually very low – it's 1 in 11.

Even though the ending has been spoiled, let's look at how we work this out with Bayes' Theorem. From Bayes' Theorem, we know that the posterior probability, the probability that I am guilty, given a positive match, is equal to the likelihood, the probability of a positive match given that I am guilty, multiplied by the prior probability, the probability of being guilty, divided by the marginal likelihood – the probability of a positive match.

The important thing to note here is that the marginal likelihood, the probability of a positive match, includes both the true positives, <u>and</u> the false positives! That is, we can re-write this formula like this, where we have the probability of a true positive given by the terms in the blue box, and the probability of a false positive given by the terms in the red box.

If we plug in all of the numbers, it looks something like this. The probability of a true positive is here – it's 999,999 divided by 1,000,000, and we multiply this by the probability that any person in the database is guilty – 1 in 1,000,000. And, if we do the math here, we end up with the same answer that we got from the contingency table – 1 in 11.

This is an example of the "prosecutor's fallacy" – the incorrect assumption that the probability that the defendant is guilty, given the evidence, is approximately equal to the probability of a false positive. And, as we can see, this is not the case. In the equation here, we can see that the terms in these two boxes are very small – this simplifies to approximately 1 over 1,000,000. But, the terms in the red boare also very small – but most importantly, they simplify to approximately 10 in 1,000,000 – so, when we combine all this together, we have 1 over 11.

As always, this is something of a silly example, but it has wide-reaching applications, from actual criminal trials to drugs testing in sports, to testing for specific diseases or medical conditions, and beyond.

Slide 9 – Probability Distributions

We can also look at this in terms of probability distributions. If the likelihood or the prior distribution are very wide, then so is the posterior. In other words, we have a larger range of possible outcomes.

Look at the first example on the left - we see that the prior distribution, in red, has a large dispersion, as does the likelihood in blue. The posterior distribution, in black, is also comparatively wide – it has a similar width as the posterior distribution. On the right, we see that a narrower likelihood also decreases the dispersion of the posterior distribution – the peak gets both thinner and taller.

This gives us an alternative way of interpreting Bayes' Theorem – the more alternative explanations there are, the less likely it is that your chosen explanation is correct. On the previous slide, we saw how even with a high-specificty test, the fact that we would expect a number of false positives – alternative explanations – meant that the likelihood of our chosen explanation was also comparatively low. Another way of saying this is, the less sure that you are of the evidence, the less likely your hypothesis is.

Slide 10 – Summary

In this lesson, we've learned about how Bayes' Theorem gives us a way to relate conditional probabilities, and allows us to evaluate the likelihood of an explanation, given the available evidence.

We've also seen how this means that if there are many different explanations for the available evidence, the probability that our chosen explanation is correct decreases.

More than anything else, this is due to the existence of false positives – we need to consider the possibility of false positives in order to correctly evaluate the outcome of a particular test.

Slide 11 – Additional resources

In the Bergstrom and West book that is part of the recommended reading for this module, you can read more about the Prosecutor's Fallacy in Chapter 9, titled "The Susceptibility of Science." I've also included a link to an article from Scientific American about Bayes' Theorem and why it's so important, as well as a link to a good explainer on Bayesian Inference. I've also linked to two further videos – the first is a good introduction to Bayes' Theorem and Bayesian Inference. The second video linked here is from the class that the Bergstrom and West book came out of – it covers similar ground as Chapter 9 in the book.

That's all for this lesson – I hope you found it interesting, and if you have any questions, please don't hesitate to e-mail me or post in the discussion forum on blackboard. Bye!