Slide 1 – Title Slide

Hello and welcome to Week 7, Part 3 of EGM101: Even More Probability. In this lesson, we'll continue our introduction into the fundamental concepts of probability, covering independent and dependent events, and how that impacts probability.

Slide 2 – Independent Events

In last week's lecture on variables, we introduced the term "independent" variable, meaning a variable whose value does not depend on another variable. In the same way, a "dependent" variable is one whose value does depend on another variable.

If two events are independent, it means something similar – knowing that one event happens has no effect on the probability of the other event happening. If, on the other hand, knowing that an event happens does change the probability of another event, then the two events are dependent – that is, their probabilities depend on each other.

To show that two events are independent, we only have to show one of the following things. Either, the probability of event A given that event B happens is equal to the probability of event A; or, we need to show that the probability of event B given that event A happens is equal to the probability of event B; or, finally, we need to show that the probability of events A and B happening is equal to the probability of event A multiplied by the probability of event B.

If we can show one of those things, then we know that events A and B are independent; like with mutual exclusion, if we don't know that events are independent, or we haven't shown that they are independent, then we shouldn't assume that they are independent.

Slide 3 – Sampling With/Without Replacement

To help illustrate the impact that events being independent or dependent has on probability, let's consider a standard deck of playing cards, which has 13 cards of each suit of spades, hearts, clubs, and diamonds. In this example, we're interested in the probability of getting a heart when we draw a card from the deck – in other words, when we sample a card from the deck, what is the likelihood that the card is a heart?

There are two ways that we can sample – with replacement, and without replacement. Sampling "with replacement" means that every time we draw a card from the deck, we put it back.

Starting with the full 52-card deck, the probability that a random card drawn from the deck is a heart is 13 over 52, or 0.25. Let's say the card we draw is the ace of diamonds – after we've drawn the card, we put it back in the deck, and the probability of getting a heart on the next draw is still 0.25.

When we sample with replacement, the probability after each event stays the same, because the total number of possible events does not change. In other words, the outcomes are independent.

On the other hand, if we sample with replacement, each time we draw a card, we keep it.

Starting with our full 52-card deck, the probability of getting a heart is 0.25. When we draw the ace of diamonds, we keep it. This means that on the next draw, the probability of getting a heart is not 13 over 52, it's 13 over 51, or 0.2549 – because we now only have 51 cards in the deck.

Now, when we draw the seven of hearts and keep it, the probability of getting a heart on the next draw is no longer 13 over 51 – it's 12 over 50, or 0.24.

Because the probability changes after each event, we know that the outcomes are not independent.

Slide 4 – The Multiplication Rule

Now, we come the first important rule for combining probabilities – the multiplication rule.

The general multiplication rule says that if A and B are two events in a sample space, S, then we know that the probability of events A and B is equal to the probability of event B multiplied by the probability of event A given event B.

We know this because this is just the formula for the conditional probability of event A given event B, which told us that the probability of event A given event B is equal to the probability of events A and B, divided by the probability of event B. Here, all we've done is re-arranged it by multiplying across by the probability of event B.

The general rule holds for any two events that are part of the same sample space. But, if the two events are independent, we can simplify this. Remember that if the two events are independent, then the probability of event A given event B is just the probability of event A – event B has no impact on the probability of event A.

In this case, the multiplication rule becomes this: the probability of events A and B occurring is just the probability of event A, multiplied by the probability of event B.

Slide 5 – The Addition Rule

The other major rule for combining probabilities is the addition rule, which tells us how we can add probabilities together.

The general addition rule says that if A and B are two events in a sample space, S, then the probability of either event A or event B is equal to the probability of event A, plus the probability of event B, minus the probability of A and B.

If you look at the Venn Diagram for the union, this just says that the area of A union B is equal to the sum of their individual areas, minus the area of the shared part – the intersection.

But, if events A and B are mutually exclusive, we know that the intersection (and therefore the probability) of A and B is equal to zero – in which case, the general rule simplifies, and we see that the probability of event A or event B is equal to the sum of their individual probabilities.

Slide 6 – Tree Diagrams

One tool that we have for visualizing the different outcomes of a series of experiments is something called a tree diagram.

Let's say that we have 10 marbles in a bag: 7 blue marbles, and 3 red marbles, and we want to calculate the probability of different events where I draw multiple marbles from the bag. We'll start by considering the sampling with replacement example, where each time I draw a marble from the bag, I put it back.

So, when I draw a marble out of the bag on the first draw, I have a 7 in 10 chance of drawing a blue marble, and a 3 in 10 chance of drawing a red marble. In a tree diagram, each branch is labeled with frequencies or probabilities, which helps us to calculate the different probabilities of traveling along each branch.

On the second draw, I again have a 7 in 10 chance of drawing a blue marble, and a 3 in 10 chance of drawing a red marble, no matter what I got on the first draw. The events are independent, because what I got on the first draw has no impact on the second draw, and so the probabilities are all the same.

In the sampling without replacement example, each time I draw a marble out of the bag, I keep it. For the first draw, this looks the same as the with replacement example – just like we saw with the deck of cards.

On the second draw, though, we see how useful the tree diagram can be in keeping track of different probabilities. Without replacement, if I get a blue marble on the first draw, the likelihood of getting a blue marble on the second draw is 6 in 9, and 3 in 9 that I get a red marble. If, on the other hand, I got a red marble on the first draw, the likelihood of getting a blue marble on the second draw is 7 in 9, and it is 2 in 9 that I get a red marble.

We can then calculate the probability of each individual branch by multiplying the probability of each individual outcome that we see on the branch together, and write this at the bottom of the tree.

If we then assume that the order doesn't matter, we can calculate the probability of different events – the probability that we end up with two blue marbles if we sample with replacement is 7 over 10 times 7 over 10, or 49 over 100. Without replacement, it's 7 over 10 times 6 over 9, or 42 over 90.

We see that there are two ways that we can end up with one marble of each color – we either get a blue, followed by a red, or we get a red followed by a blue. If we sample with replacement, the likelihood of either of those two events happening is 21 over 100, so together, the likelihood is 42 over 100. Without replacement, the likelihood is 21 over 90 for each event, or 42 over 90 for either one.

And finally, we calculate the probability of getting two reds in the same way that we calculated the probability of getting two blues – it is 9 over 100 if we sample with replacement, and 6 over 90 if we sample without replacement.

Slide 7 – Contingency Tables

One other useful tool we have for working out combined probabilities is something called a contingency table, where we display different numbers of outcomes based on different values for two variables which may or may not be dependent.

Contingency tables help us calculate conditional probabilities directly. To help illustrate this, we'll look at a sample of 100 pets and their preference of fish: salmon or tuna. In the table here, we can see that we had 49 dogs and 51 cats, based on the totals in the last column; and using the row totals, we see that 60 pets preferred salmon, while 40 preferred tuna fish.

If we were to select a pet at random from this sample, what is the probability that it's a dog? This is just the total number of dogs, divided by the number of animals – from the table, we have 49 dogs and 100 animals, so the probability of picking a dog at random is 49 over 100, or 0.49.

Similarly, we can calculate the likelihood of picking a pet that prefers salmon, which is the proportion of animals that prefer salmon. We have 60 pets that prefer salmon and 100 animals, so this probability is 60 over 100, or 0.60.

What if, instead, we want to know the likelihood of picking an animal that prefers salmon, if we only select a dog? In this case, we know that the conditional probability is the probability that a pet prefers salmon and is a dog, divided by the probability that the pet is a dog. But, we can also calculate this probability directly from the table – it's the number of dogs that prefer salmon, 39, divided by the number of dogs, 49. And when we plug this into our calculator, we get approximately 0.796 – roughly 80% of dogs surveyed preferred salmon.

Now, what if we want to know the opposite? If we randomly select a pet that prefers salmon, what are the chances that it's a dog? We can answer this in the same way from the table – this is the number of dogs that prefer salmon, divided by the number of pets that prefer salmon, or 39 over 60, or 0.65.

We'll work with contingency problems again later this week – they're an incredibly useful tool for calculating conditional probabilities, since they make it easy for us to have all of the relevant numbers in one place.

Slide 8 – Summary

In this lesson, we've discussed one of the key concepts of probability: whether or not events are independent. If they are independent, it changes the way that we analyze the probability, in effect by making the calculations a bit easier.

We've also seen how our sampling method affects our calculated probabilities – if we sample with replacement, it means that our samples are wholly independent, though it also means that we have a chance of selecting the same member of the population again.

We also looked at two useful tools for helping us to visualize and calculate probabilities – tree diagrams and contingency tables. We saw how tree diagrams can especially help in cases where events are not independent, and we saw how contingency tables are useful for calculating conditional probabilities.

Slide 9 – Additional resources

You can read more about the topics we've discussed here in the textbooks – Illowsky and Dean, Chapters 3.2 to 3.4; Caswell, Chapter 8; or Weiss, Chapters 5.1 to 5.3.

I've also included a link to a video from Khan Academy that provides more information about independent and dependent events.

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!