Slide 1 – Title Slide

Hello and welcome to Week 7, Part 4 of EGM101: Discrete Probability Distributions. In this lesson, we'll start to learn about probability distributions, and how they can tell us about the probability of different events.

Slide 2 – Roll 2 D6

To begin, we'll consider two six-sided dice. We'll assume that these dice are fair, meaning that there's an equal chance for each number from 1 to 6 to appear on either D6 when we roll them.

The question we want to answer is, what is the probability of getting a particular number? When we roll two six-sided dice, we can get any number from 2 (if we roll two 1s) to 12 (if we roll two 6s).

Another way of phrasing this question is, "what is the probability that the outcome, X, is equal to some discrete value?" Hopefully the "discrete" part of this makes sense – the dice are limited to integer values, which means that we are dealing with discrete numbers.

To answer this question, we can make a table of all of the different ways that the sum of the two values equals a particular value between 2 and 12. With two dice, there is only one way for us to roll a 2 – both dice have to come up with a 1 for that to happen. What about 3, though? Well, there are two ways for the sum of the two dice to equal a 3: either the first D6 is a 1 and the other is a 2, or the first D6 is a 2 and the other is a 1.

We can do this for all of the other possible values, and the end result should look something like this. The number of possible combinations increases until we get to 7, at which point it decreases again until we get to 12 – just like with 2, there's only 1 way to get a 12 if we roll two six-sided dice.

We can then count up the number of ways to roll a particular value, and add that column to this table. What we have done is calculate the frequency distribution for this set of events – each total is an event made up of 2 outcomes, and the number of different ways that those events can happen is the frequency. We can take this a step further and calculate the probability distribution, by dividing by the total number of events – we have 36 possible combinations, and so the probability of each event is equal to the number of ways to get that event divided by 36.

Not only that, but we can plot the distribution as a histogram, like this – we see that it has effectively the same shape as the diagram in the table, with a peak at 7 and steps off to either side.

Probability distributions are an extremely powerful tool that we can use to estimate probability – not only can we calculate the probability of a particular outcome or event, like rolling a total of 6: 5 over 36, or about 14%. We can also calculate the probability of a range of outcomes – for example, the probability of rolling a 9 or greater. From the table, we see that there are 4 + 3 + 2 + 1 = 10 ways to get a 9 or higher, so the probability of rolling a 9 or higher is equal to 10 over 36, or about 28%.

Slide 3 – Successive coin flips

Now, let's move away from dice games, and focus on a different game of chance: flipping a coin. Let's assume that we have a fair coin, and that we flip it some number of times – after one flip, we could end up with either a head or a tail. On the second flip, we could also end up with a head or a tail, but the different outcomes form a tree diagram like we saw in the previous lesson. After 3 flips, the possible outcomes would look like this, and we could keep on going in this way.

So, let's say we flip this coin n times, and each time, we write down whether the result was a head or a tail. Rather than flipping a single coin n times, we could flip n coins a single time and write down the results – though this approach only really works if you're Scrooge McDuck rich.

So, the question is: after we flip the coin n times, how many different ways can we end up with the following results: n heads? (in other words, all of the results are heads); n – 1 heads? (in other words, we have 1 tail and the rest are heads); and so on, until we get to 0 heads, or n tails.

To help figure this out, we can write this up in a table like this, with the number of ways that we can get a given number of heads written in each column. After 1 flip, we see that we have one way to end up with 0 heads, and 1 way to end up with 1 head – we either get a head or a tail on the first flip.

After two flips, we see that we have 1 way to get 0 heads, 2 ways to get 1 head – first a head, then a tail, or first a tail, then a head – and one way to get 2 heads, for a total of four possible outcomes.

And, we can keep going like this, writing down the number of different ways there are to get a given result. And then, using the total number of possibilities in each row, we can calculate the probability of any given outcome, like this.

It turns out there's an easier way to figure this out, though, because this is an example of something called a "binomial experiment". A binomial experiment has a fixed number of n trials, each of which has two possible outcomes – either we succeed, or we fail. There is no try. Incidentally, this kind of trial, with only two possible outcomes, also has a special name – it's a Bernoulli trial, and if you continue learning statistics or probability you will probably see this name come up again.

Anyway, a binomial experiment requires that the trials are independent – in other words, the outcome of one trial cannot have an influence on the outcome of any other trial. It also requires that the trials are identical – that is, each trial has to have the same probability of success or failure.

Slide 4 – The Binomial Distribution

The outcome of a binomial experiment, which we will denote as an uppercase X, follows something called a binomial probability distribution.

To work out the probability of any number of successes, we can use a formula. I'm not going to show this formula on the slide, because it goes beyond what we need to know in this module – but, if you're interested, you should be able to find it. But, it does depend on the following things.

First, we need to know the probability of success, which we denote using a lowercase p. For the coin flip example, this is of course 1 in 2, or 0.5.

We also need to know the total number of trials, n. And that's it – those are the only things we need to know, to be able to work out the probability.

If you look at the figure here, you should see that as the number of trials increases, the "peak" probability decreases – the reason for this is that the sum of the individual probabilities has to equal 1, and as we increase the number of trials, we increase the number of possible outcomes. This means that the likelihood of any single outcome gets smaller and smaller.

In effect, what's happening is that we're spreading the total probability out over a much larger range of values – this is an idea that we'll come back to in the next lesson.

Slide 5 – The Geometric Distribution

Instead of asking the probability of getting a certain number of heads or tails, we might instead ask what is the probability that we fail a certain number of times before we succeed? Instead of using the binomial distribution, this kind of question uses something called the geometric probability distribution.

Here, we have a similar setup to the binomial distribution – we have a fixed number of Bernoulli trials, and these trials must be independent – the probability of success or failure of any individual trial does not have any effect on the probability of success or failure of any other trial.

Just like with the binomial distribution, I'm not going to write the formula for calculating the probability distribution. Also like the binomial probability distribution, if we want to calculate the probability that we don't get a success until the k-th trial, we need to know the probability of success, denoted p, and that's it.

This figure shows how the probability distribution changes for different probabilities of success or failure – for probabilities of success very close to 1, the likelihood that we go more than 1 or 2 trials without success is very small. As the probability decreases, though, we see that the probability of going a long time without success increases. It's important to note here even though it doesn't show up on the graph, we do have a nonzero probability of failing a large number of times. For p equals 0.8, the chances of failing 10 times in a row before succeeding is about 1 in 2.4 million – it's a very small probability, but it still exists.

Slide 6 – The Hypergeometric Distribution

Despite the name, the hypergeometric distribution helps us answer a slightly different kind of question. Let's say that we have a mixture of 15 marbles in a bag: 9 red, and 6 blue. If we randomly select 7 marbles from this bag, what is the probability that we end up with x blue marbles?

This kind of experiment is known as a hypergeometric experiment, and most importantly, we're sampling without replacement – that is, once we select a marble, we don't put it back in the bag. This means that the trials are not independent, and it also means that they're not Bernoulli trials.

To calculate the probability of ending up with a certain number of blue marbles, we need to know: the size of the whole group, denoted by an uppercase M – for our marble example, this is 15. We also need to know the size of the group of interest, denoted by a lowercase n. By “group of interest,” we mean the group that we’re working out the probability for. In our example, this is the number of blue marbles in the bag, which is 6.

Finally, we need to know how many samples we’re taking – that is, the number of marbles that we select from the bag. We denote this using an uppercase N, and for our example it’s 7.

Based on these values, the probability distribution for our example problem is shown here. We can see that the most likely number of blue marbles is 3, but any number of blue marbles between 0 and 6 has a non-zero probability. Obviously, we can’t select more blue marbles than there are in the bag, nor can we select a negative number of marbles, so values of x that are greater than small n, or less than 0, are impossible – the probability for these values is equal to 0.

Slide 7 – The Poisson Distribution

The last discrete probability distribution that we’ll look at is the Poisson distribution. To illustrate the kind of problem that the Poisson distribution helps us answer, consider the following question. If, on average, we catch 10 salmon per hour, how many salmon can we expect to catch in the next 15 minutes?

The Poisson distribution tells us the probability of x events happening in a fixed interval – for example, the number of fish that we might catch in the next 15 minutes. In using the Poisson distribution, we assume that all of the events are independent. The only thing that we need to know to calculate the probability is the average rate of occurrence, denoted using the lowercase Greek letter lambda.

The peak of the Poisson distribution occurs close to lambda – that is, the answer to the question “how many fish can we expect to catch in the next 15 minutes?” is “whatever the average number of fish that we catch in 15 minutes is.” From the figure here, you can also see how the distribution changes as we increase the value of lambda – not only does the peak value decrease, but we also see how the distribution becomes broader. Tehnically, the distribution is unbounded – just like we saw for the geometric distribution, any value greater than zero has a non-zero probability, even if it’s very very small.

Slide 8 – More on Probability Distributions

As we saw earlier in the lesson, we can use distributions to calculate the probability of a range of events, and not just the probability of a single event.

As an example, let’s say that we flip a fair coin 10 times – as you should hopefully remember, this is a binomial experiment with 10 trials, and the probability of “success” is 0.5. Plotted as a histogram, this distribution looks like you see here – a peak at 5, with very small chances of getting 0 or 10 heads.

One question that we can answer with this distribution is: “what is the probability of getting more than 7 heads?” Written out mathematically, this is the probability that the random variable, X, is greater than or equal to 7.

And, as we discussed earlier, we can answer this by calculating the probability of getting 7 heads, the probability of getting 8 heads, the probability of getting 9 heads, and the probability of getting 10 heads.

As we discuss further in the next lesson, it turns out that this is also the same as adding together the area of the bars inside of this box – because they have a width of 1, the area of these bars is equal to the probability of getting a particular number of heads.

Slide 9 – Summary

In this lesson, we've seen how probability distributions can help us understand or calculate the likelihood of a particular outcome or event.

Theoretical distributions, like the ones that we have introduced here, can save us from having to calculate the probability of each individual outcome by hand, like we did for the example of rolling two dice. Instead, we can use the formula for the probability distribution to calculate the probability of whatever events we are interested in.

But, we also saw how the distribution can help us calculate the probability of a range of events or outcomes – something that will be very useful for understanding continuous probability distributions.

Slide 10 – Additional resources

You can read more about the topics we've discussed here in the textbooks – Illowsky and Dean, Chapter 4; Caswell, Chapter 13; or Weiss, Chapter 5.4.

I've included links to three different videos here that explore some of these topics in more detail – the first two discuss Pascal's triangle, which is a triangular array made up of the binomial coefficients – it has a lot of very interesting and useful properties, so if math puzzles are something that interests you, check these videos out.

The other link is to a video about the probability of correctly guessing 14 coin tosses in a row, something that happened between 1998 and 2011 in the Super Bowl, the championship game for real football.

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!