Slide 1 – Title Slide

Hello and welcome to Week 7, Part 1 of EGM101: Introduction to Probability. In this lesson, we'll start to learn the basics of probability, the branch of mathematics that we use for inferential statistics.

Slide 2 – Week 7 Outline

Over the remaining lessons this week, we'll cover a number of core concepts of probability, including discrete and continuous probability distributions, before wrapping up by discussing the Central Limit Theorem, one of the most important theorems in all of statistics and probability.

Slide 3 – Question

Before we get started, though, we begin with a question. Let's say that we know that on average, salmon in our favorite stream are around 70 cm long, with a standard deviation of 15 cm.

Now, if we go fishing in our favorite stream, what is the likelihood of catching a fish 100 cm or longer?

In other words, we want to know the probability of catching a fish that is 30 cm larger than the "average" fish in the stream.

This is not a question that we can answer yet, so don't worry if you're not even sure where to begin. By the end of this week's lesson, though, you should be able to understand what this question is asking, and how you would go about finding the answer to it.

Slide 4 – Inferential Statistics and Uncertainty

In Week 5 at the beginning of this section of the module, we covered a number of definitions. First, we learned that "descriptive" statistics is how we organize and summarize our data, while "inferential" statistics is how we can draw conclusions from "good" data.

So far, we've mainly focused on descriptive statistics, but now we'll start to learn more about inferential statistics. Because inferential statistics involves drawing conclusions about a population from a comparatively small amount of information, there is a degree of uncertainty involved, and we need to be able to understand and describe this uncertainty.

In mathematics, probability theory is what we use to understand and describe uncertainty and evaluate the likelihood of different outcomes. In other words, it is the foundation of inferential statistics. Thus, before we can jump in to learn about inferential statistics, we need to cover the basics of probability.

Slide 5 – Some Definitions

As is tradition, we start with some definitions, beginning with a statistical experiment – in probability theory, this is the process by which an observation is made. This could be, for example, measurements, laboratory experiments, clinical trials, or something like a coin flip – as long as the process is repeatable and the result is not certain, it is a statistical experiment.

In the context of probability, we refer to the result of an experiment as the “outcome” of the experiment. So, when we flip a coin, the way that it lands, heads or tails, is the outcome of the experiment.

The set of all possible outcomes of an experiment is known as the sample space of the experiment, normally denoted using an uppercase S. Here, we see that the sample space of a single coin flip is either heads, denoted by H, or tails, denoted by T. Note also that when we write out the sample space, we use curly brackets.

If we flip the coin two times, the sample space looks like this: we might end up with two heads, two tails, a head first and then a tail, or a tail first and then a head.

An event is any subset of the sample space of the experiment – this can be any individual outcome, such as getting two heads on two coin flips, or it can be a combination of outcomes, such as getting either two heads or two tails on two successive coin flips.

Finally, when we talk about the probability of an event, this specifically means the long-term relative frequency of that event. When we introduced frequency back in Week 5, we said that it was the number of times that an observation, or a range of observations, appeared in the data – and that is still true here.

As we will discuss further this week, if we know the sample space of an experiment, we can easily calculate relative frequency of the individual outcomes that appear in the sample space. So, the probability of getting two heads in a row when we flip a coin is just the number of times two heads appears in our sample space, divided by the total number of outcomes in the sample space: 1 divided by 4, or 0.25. Note here that we denote the probability of an event using an uppercase P.

You should also remember that relative frequency is always between 0 and 1 (or 0 and 100%), which means that probability must also be between 0 and 1 (or 0 and 100%).

If the probability of some event A is equal to zero, we say that it is impossible for that event to occur – it does not appear in the sample space.

On the other hand, if the probability of some event A is equal to 1, then we say that it is certain that that event will occur – there is no way that the outcome an experiment is not in A.

Slide 6 – Deterministic, Probabilistic, and Random

When we’re talking about the outcome of an experiment, we have a number of different terms that we can use to describe how we get to that outcome, beginning with deterministic. In a deterministic experiment or process, the outcome is determined beforehand – there is no uncertainty in how we get to the outcome.

If there is some level of uncertainty in how we get to a particular outcome, then we say that it is a probabilistic or random experiment – in other words, the outcome is unpredictable or due to chance, and we cannot know what will happen beforehand.

Remember that in Week 5, when we began talking about sampling, we said that in a “random” sample, each member of the population has an equal chance of being selected. Another way to say this is that a random sample is fair: each member of the population, or each outcome of the experiment, has the same equal likelihood of occurring.

With a fair coin, we have an equal chance of getting heads or tails each time we flip the coin. With a fair six-sided die, we have an equal chance of getting a 1, a 2, a 3, a 4, a 5, or a 6, each time we roll the die.

It’s important to note that this idea refers to individual outcomes, and not the value of that outcome. If I have 3 50p pieces and a 1 pound coin in my pocket, and I pick a coin at random, I have an equal likelihood of picking any one of the 4 coins in my pocket. This does not mean that I have an equal likelihood of selecting 50p or £1, though – the chances of me selecting 50p from my pocket is ¾, or 0.75, while the chances of me selecting £1 is ¼, or 0.25.

Slide 7 – Random Variables

Hopefully you will recall that a variable is some characteristic that will vary, or have a different value, for each member of the population. Examples of different variables that we’ve seen so far have been fish lengths and weights or monthly temperature and rainfall.

A random variable, then, is a variable where the variation in the value of the variable is random, or unpredictable. Some different examples of random variables are the number of heads after a certain number of coin flips, the value of the numbers that show when we roll dice, or the weight of a randomly caught or selected fish.

Most of the time, we are going to assume that we are working with a random variable, most of the time out of convenience. It’s important to note, though, that this is something that we do at our own risk – just because it’s true most of the time, doesn’t mean that it’s always going to be true. Fortunately, it’s true enough of the time that we usually manage to get away with it.

Slide 8 – The Expected Value

In Week 5, we introduced the idea of the arithmetic mean – the sum of all of the values in a sample or population, divided by the total number of values in the sample or the population.

In probability, instead of the arithmetic mean, we often use the expected value of a variable, written using an uppercase E.

For a variable X, the expected value is equal to the sum of each possible value of X, multiplied by the probability of that value. Another way to say this is that the expected value is just a *weighted* average of all of the possible values of X.

Because this is an average, it's a measure of central tendency – and in fact, it's numerically equivalent to the arithmetic mean.

As an example of how to calculate this, let's calculate the expected value of rolling a fair six-sided die, which I will write as a D6.

Because the die is fair, each number has the same probability: 1 over 6. We can write out the formula like this: the expected value is equal to 1 times 1/6 plus 2 times 1/6, plus 3 times 1/6, and so on.

Because each of the outcomes are equally likely, we can re-write this equation like this, as the sum of the individual values divided by the number of values – and here you can see that it is, in fact, the arithmetic mean. When we sum these values up, we get 21, and 21 over 6 is equal to 3.5

Slide 9 – The Law of Large Numbers

What does the expected value tell us, though? Remember that we said that the probability of any individual event is the long-term relative frequency of that event. In other words, if we roll the die many many times, we would expect to get a 1 around 1/6th of the time.

What if we kept track of each roll, and calculated the mean value of all of our rolls each time we rolled the die? What happens to the sample mean as we increase the number of rolls?

This brings us to the first of several important concepts in probability theory: the law of large numbers.

Put one way, the law of large numbers tells us that as we increase the number of trials of a random variable – that is, as we repeat an experiment or a measurement over and over again, the probability that the sample mean of all of the outcomes equals the expected value within some margin approaches one.

Put a different way, as the number of trials increases, the average outcome looks more and more like the expected value. As we take more and more samples of a random variable, the sample mean gets closer and closer to the expected value. Or, alternatively, we can say that the average of many measurements is more accurate than a single measurement.

As an example, let's look at the expected value for rolls of a single D6, shown in this graph as the red line. For a low number of rolls, we see that the mean value of the rolls fluctuates quite a bit – we have high variability. We start off down near 3, then up near 4.5, and back and forth. Because we only have a few measurements, each value changes the average by quite a lot.

After many rolls, though, the curve flattens out, and stays pretty close to the expected value, shown by the black dashed line. There's still some variability, but line showing the mean value is pretty smooth and doesn't change by very much as we increase the number of rolls.

Slide 10 – Summary

In this lesson, we've discussed how inferential statistics is inherently uncertain, which is why we need to learn about probability theory – the branch of mathematics that specifically deals with uncertainty.

We’ve also discussed how most of the time, we treat our observations or measurements as a random variable, though it’s always a good idea to double-check whether this is a reasonable assumption to make.

And finally, we talked about the Law of Large Numbers, one of the most important ideas in probability theory, which tells us that the average of many measurements or observations tends to be more accurate than a single observation.

Slide 11 – Additional resources

You can read more about the topics we’ve discussed here in the textbooks – Illowsky and Dean, Chapter 3.1; Caswell, Chapter 12; and Weiss, Chapter 5.1.

I’ve also provided a link to an article that covers the same topics that we have covered here, that you can also read to help your understanding.

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!