Transcriber's Name Maria Priya

Applied Econometrics

Prof. Tutan Ahmed

Vinod Gupta School of Management

Indian Institute of Technology, Kharagpur

Module No. # 02

Lecture No. # 13

Probability Distribution

Hello and welcome back to the lecture on Applied Econometrics. So, in the previous lectures, we have been talking about different types of probabilities, we learned this concepts. Now, in this lecture, we will introduce a concept of probability distribution. Now, before I get into probability distribution, it is important for us to again go back to the concepts that you have learned. And one concept that you have learned is random experiment and associated variable, random variable. Now, recap what is a random experiment.

So, when I said random experiment or whenever we use the term random, so, we usually refer to the fact that there are certain chances involved, there are certain uncertainties involved, I really do not know the outcome of the experiment; so, that is when I call an experiment a random experiment. Or, the variable which is associated with that experiment, we call it a random variable. So, that means, I do not know the outcome of the experiment, or I do not know all the possible results that I am going to get uh for that random variable.

Now, when I ca uh talk about this random variables, one thing we have to remember, and that is, we are talking about uncertainties, but we never talk about equiprobable or equally likely chances. So, by random, does not mean uh they are, the probabilities are equal for all the events. Sometimes, it is confusing, we mentioned it before. uh It only means that there are certain uncertainties involved. So, what are the examples of uh random variables? Random variables could be something like, let us say I am talking about temperature distribution. Okay.

So, I do not know the temperature uh that we are going to have in say next one month. So, I am going to predict the temperature. So, that is the random variable. Or, I am going to throw a dice, or I am going to throw a coin, or maybe couple of dice and couple of coins, and I do not know what outcome I am going to get, until and unless I actually do that experiment, I

actually run that trial. Right. So, that is why this is called random variable. Now, there are 2 types of random variables, one is continuous and other is discreet.

So, let me write it down; (refer time: 02:25) discrete and continuous. So, think about the, it think about one example of discrete random variable and another example of continuous random variable. So, we have already spoken about it. So, by discrete, it means that you can actually get exact value of a particular event. So, what I mean by that; for example, uh if I talk about, you know, uh throwing 2 uh dice, or maybe tossing 2 coins. So, we always get those outcomes, right. And we always get a specific probability to that particular uh event.

Say, I want to get a probability for, say 2 head; so, I will get a probability for that. Whereas, for continuous uh random variable, uh we, it is basically, we can have a continuum of values of the variable. It is, we do not really have, uh we cannot really specify probability for a particular value; and I I will I will just explain that. So, let us say I am talking about temperature, right. So, when I talk about temperature, so, temperature could be any temperature, like, let us say the temperature in Kharagpur uh could be, you know, anything between 10 degree Celsius to 40 degree Celsius.

Now, it can be, let us say, you know, we can say 30 degree; today's temperature is 30 degree Celsius. But when you measure, when you actually see, it might be 30.001 degree Celsius, or it might be 29.998 degree Celsius. So, it is very difficult to get exact temperature, because, when I; it is basically continue of values. You can actually have a little bit of fluctuation, you know. And for for every different values; so, in this range, 10 degree Celsius to 40 degree Celsius, you can have infinite number of values, if I consider these kind of values.

So, you can go on like 99998 degree Celsius and so forth. So, that way, you really do not have; if I take any specific value, any specific value, you really do not have a probability assigned to it, because of the fact that, for any particular value like this, the probability is almost 0. So, we cannot really talk about probability for this kind of values. And that is why we will see for discrete random variable and for continuous random variable, we have to consider the probability distribution differently. Okay.

So, now we will come to the co idea probability distribution. But but but before I go to go to probability distribution, let me actually give an example. Let us say; and this kind of examples, we have done before. Let us say we have 2 dice, okay, and we are throwing these 2 dice. All right. So, we know the outcome of the throws, right? So, it can be uh any value with, from 1 to 6 for both the dice. Now, I am defining a random variable, and that random variable X, let us say; (refer time: 05:22) random variable X actually signifies the sum of values that you get from 2 dice.

So, when you throw 2 dice, you get value, a specific value on each of these dice. And then you sum it up, and that is what our random variable is. That represents the sum of the values after you throw 2 dice. And let us say uh the values that you can get from, let us say dice 1. Let me actually use a different colour; always look good if you have many colours. Dice 1. And let us say I have another dice and that is dice 2. And here, I will; these are the values you can get, 5, 6. And for dice 2 also, you can get these values 1, 2, 3, 4, 5, 6. Okay.

Now, to get the va, to sort of uh get X, the sum of the values of 2 dice; so, what I will get here; so, if I get 1 and 1, let me write down a different colour; 1 and 1, so, I will get 2 here; here 2 and 1, I will get 3; 4; 5; 6; 7. Here, I will have 3, 4, 5, 6, 7, 8. Here, I will have 4, 5, 6, 7, 8, 9. Then I will have 5, 6, 7, 8, 9, 10; 6, 7, 8, 9, 10, 11. And then, I will have 7, 8, 9, 10, 11, 12. So, I will have all these different values, which is basically uh representing the values of, the different values that X can possibly take.

Now, if I want to constitute; so, from here, I can very easily constitute a frequency uh distribution. And from there, we will see the probabilities associated with each of these outcomes. So, let me do that. So, let us just count. So, how many twos are there? How many twos are there? So, basically, let us say; uhm let me actually create a new page. (refer time: 07:51) So, here I will. So, the values of X here are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12. Right?

Now, if I go back and see, how many ones are there; from here, you can you can count, you can see, how many ones are there? There is no 1, right? So, I can write, this is equal to 0. How many twos are there? I have only one 2. Okay. And how many threes are there? 2 threes. How many fours are there? 3 fours. How many fives are there? 4 fives. How many 6 are there? 5 sixes. So, let me write these numbers down. Okay. Yellow; 0, 1, 2, 3, 4. And what about 6? We had 5 sixes; 5. And how many sevens are there?

You will see, we have 6 sevens; 2, 3, 4, 5, 6. And then, I will see 5 eights, 4 nines, 3 tens, 2 elevens and one 12. Okay. See, if I just write it down here, 6; oops; 6, 5, 4, 3, 2, 1. Okay. So, these are the frequencies I got. The frequencies, if I count the frequencies; so, this is, these are the number of frequencies, these are the number of times you can see this particular outcomes. Okay. Now, I, now I want to calculate the probabilities. Now, I want to calculate the probabilities. And we know that uh there are; altogether, how many outcomes are here?

So, altogether, there are 6 and 6, so, 36 outcomes. So, N is 36. So, when I have N is 36. So, then, I will have probability values 1 by N, right? So, it is going to be; so, there is no probability for 0. So, it is 1 by 36, 2 by 36, 3 by 36, 4 by 36, 5 by 36, 6 by 36. Again, 5 by 36,

4 by 36, 3 by 36, 2 by 36, and finally 1 by 36. Okay. So, these are the probabilities that you have got. And if you add them up, you will see, it is going to be 1. So, basically, if you add them up, I am going to get 1 plus 2, 3; plus 3, 6; plus 4, 10; 15; 21; and 5 is 26; plus 4, 30; 33; 35; 36.

So, 36 by 36 is 1. So, our calculation is correct. So, what you have got; basically, the reason we did this exercise is to sort of give you an idea about what is the frequency distribution or what is the probability distribution. So, for all these outcomes, we kind of ran an experiment uh where we threw 2 dice and we got all these possible, we know all these possible outcomes; we have their frequencies; we got their frequencies and we also calculated the probability distribution.

So, these are basically uh how you can see a probability distribution and a frequency distribution. Okay. Now, that we have to sort of; when we talk about probability and frequency, we have to sort of get some ideas clear, some uh basic uh ideas. And that is uh, we will see, we are repeating, we are sort of encountering these kind of uh concepts; we have to clarify these kind of concepts time and again. And that is, we will talk about a double structure of variable. And we are just going to come there.

But before that, we have to sort of understand a fact that, when you are talking about a frequency distribution and the probability distribution, certain things are different. Where where are they are different? So, when we are talking about frequency distribution, we are basically, you know, already sort of running the experiment, we are getting the results, right? We already have the results in our hands. So, in this case, uh we have sort of, you know, like, constituted this sample space, it is uh sort of under the paradigm of classical uh probability theory, where you really do not need to run an experiment, but you always know, given a certain condition, what are the outcomes going to be.

uh But we will see in a while, when you do another, another example where there we will actually, you know, use some historical data to constitute the frequencies. Okay. And then, from there, we will get the probabilities, and sort of probability distribution. Okay. So, we will see that in details. But basically, the idea is, when I talk about frequencies, we already have the numbers in our hand. So, when we already have the numbers in our hand, it means that you have run the experiment, you have got the facts on the table. Right.

So, you do not, you do not have to sort of think about a future potential distribution, you already have the numbers. And that is basically the difference between frequency distribution and the probability distribution. So, from frequency distribution, you constitute the probability distribution. And that actually helps you to sort of for for future events or sort of

to get the potential distribution; you sort of make a prediction from the probability distribution. uh And that is where these two are uh quite different.

And in this neck lec, next lecture, we are actually going to talk about the differences between probability distribution, frequency distribution, and we are going to see with an example, how we constitute a probability distribution from a frequency distribution. So, with that, we end the lecture here. And in the next lecture, we are going to see in details, how the concepts are different.