Transcriber's Name Priya Prashanth

Applied Econometrics

Prof. Tutan Ahmed

Vinod Gupta School of Management

Indian Institute of Technology-Kharagpur

Lecture No. #07

Probability Theory (Contd.)

Hello and welcome back to the lecture on Applied Econometrics. So we have been talking about the different types of probability process. We spoke about classical approaches to probability. Now in this lecture, we are going to talk about the frequentist approach of probability. And we will see how they differ. Now let me ask a couple of questions before I begin.

For example, let us say if I want to understand the life expectancy of an Indian who is 25 years old. And say I am a, I am an insurance company, and I am really interested to understand the life expectancy of someone of you know of, for a for a given age, the number of years someone can live. Because I have to kind of decide the insurance premium based on that life expectancy.

So uh if I think of using classical probability rule, it will not be able to help me because I need to get some background of the, you know life expectancy of Indians. And that might actually vary if I go to England. And that might actually vary if I go to some other country, right? So for that purpose, we actually need to calculate probability using historical data.

And when we calculate probability using historical data, we actually, that approach is actually called the frequentist approach of probability. We basically count the frequency, count the numbers from past data. Now actually, I will tell you story before I begin. And that story goes back to the time of Mahabharata. And we will see the examples of frequentist approach to probability back in that time.

You might have heard the name of King Nala and King Nala is a you know pure and righteous man. He is a king. He is married to his wife Damayanti. But there was a evil king whose name is Kali and Kali is his neighborhood you know king of the neighborhood neighborhood kingdom. Kali was jealous and Kali wanted to snatch Nala of all the things he had.

But he never had a chance because, you know Nala is a righteous man, so he is a pure man. Now there was something in Nalas character that was a you know his weakness and that was gambling. So Nala used to gamble, right? Kali knew it and Kali wanted to actually defeat Nala on gambling. So Kali invited Nala in one one day to his palace, with all you know due respect, and he invited Nala to actually play a game of gamble with him.

And Kali is a shrewd man at what Kali did is Kali gradually, you know enticed Nala into this game. He kind of made him like totally, you know forgetting his reality. So what he did is he made Nala win in the first bet, in the second bet, in the third bet. And deliberately, Kali started losing his you know like elephants, his horses, his army, almost everything. So when Kali was almost, you know uh Kali lost everything except his kingdom.

So the last bet Kali said, well Nala, you are doing really well in this game. And I want uh to do I want to take a last chance and basically I want to bet my kingdom. And of course you also have to bet your kingdom and your wife. Now Nala being enticed to that he actually ended up betting his kingdom. And what happened is in the last game, in the last game, Kali won and Nala lost everything in that last game.

Losing everything uh Nala actually started wondering he has nothing. He has, you know he has no home. He is traveling in jungle. No identity anymore. Nala happened to like it happened for many years. And it was lot of you know tragedy in his life how he spent his time. After many years Nala actually met a man named Rituparna. Now Rituparna is a man who knew how to play gamble.

Rituparna somehow befriended Nala and Rituparna actually told Nala a secret and he told he whispered that, you know there is a way I can actually ensure if I win gamble or not. Nala did not believe that because gambling is pure gambling. It is a matter of chance. Rituparna then asked Nala can you tell me how many leaves are there in the tree? And Nala said it is it is very difficult.

I have to count the number of leaves in the tree and Rituparna said no. I can tell you the number of leaves in the tree without even counting. Rituparna said how is, uh Nala said I mean Nala did not believe him and Rituparna actually, you know given number to Nala that this many leaves are in the tree. Well, Nala did not believe that and it was in the evening.

What Nala did is that he actually started counting leaves of the tree throughout the whole night and in the morning, uh you know when, you know Rituparna got up, Nala counted all the tree, all the leaves in the tree, and Nala came back. And Nala said, he was surprised, he said, you know Rituparna's prediction is almost pretty close to that. So how Rituparna did that?

So how Rituparna did is very simple. So Rituparna simply took one say stem of that tree, and he counted a bunch of, you know like, leaves on that stem. And he made a sense of how many stems are there in the tree. And he made a prediction. So this is essentially, you can say, you know some evidences of, uh you know how people used that frequentist approach to calculate, you know the the very early days of calculating probability.

And this is one story that mathematical historian, his name is Ian Hacking, he was actually trying to figure out the genesis of uh you know probability or the frequentist approach. And he found this evidence in Mahabharata. Now this is a story, but actually, it is it got a shape in commercial life, back in 1666 onward, like when we had London fire, Hamburg fire.

So you know the city was kind of demolished in fire, and people sort of, you know started to rebuild the city. And that was also the genesis of the insurance industries. Because, you know the people, the business people, they actually found an opportunity that, you know if you can insure people for all these kinds of risks, so they actually would pay you a premium, and you can actually, you know build a business around it.

Now that was the beginning of actural actuarial science. I mean, the example I gave at the beginning, I want to understand the life expectancy of someone, you know because of a given age, because I want to charge a premium on on, you know in case on his life, basically. If he dies, I have to pay back some money to his family. So these are basically the examples of the frequentist approach to probability.

And as you see, I mean, the all the classical, you know the rules of probability that we have learned after classical approach, they are all valid here. Only thing is that we are deriving the probability out of data. So with this, I will actually do an example. (refer time: 07:26)

And this is the data set that I have and this is I have taken from somewhere and let us say, we are talking about a town, XYZ town. And this is a data for COVID related, you know death in that town, let us say. I have this data for different age groups, and I have number of deaths in that age group.

I also have a number of deaths with if they have any condition like the morbidity condition like comorbidity condition, or they do not have any comorbidity condition, or I do not know why exactly they die. I mean, I do not have any information about their morbidity. I mean, if they had morbidity or not. So let us say this is the table and I actually want to uh answer some questions.

So let us say the question number one is that what is the, you know let us let me write down the questions actually. (refer time: 08:22)

What is the probability, let us, question number 1, what is the probability probability that someone who died in the age group of 64, I think I had 64 to 75 with underlying condition. So I do not have any formula like any classical approach formula that I can use directly. But what I can do is I can actually use the data that I have provided to actually uh calculate that, okay.

Similarly, I will just ask couple of more question so that our, we are conceptually clear on this topic. So the second question is that uh a person died due to Covid in XYZ town, town. What is the probability, let us say what is the probability that he or she, he or she is in the age bracket in the age of 64 to 75, let us say. So these are different questions. But we can essentially use that data and we can actually get the result.

Similarly, we can actually, you know let us, let us just answer these two questions, okay. So what how what do we do here? Simply the first case we have someone who died at the age group of 64 to 75 with underlying condition. So first thing what is given here is the age group. So I have to actually focus on the age group, right. Age group is 64 to 75. So this is the age group and with underling condition, right.

The underlying condition is, you know uh is 1272. So how do I write the probability? How do I write the probability? (refer time: 10:51)

I write so for question number 1 is, so probability that probability that let us say let me write down the denominator first. Probability that he is age group 64 to 75. And on the numerator probability underlying condition intersection age group 64 to 75, right. So the number is we already have the number here. It is just you have to plug the numbers here. So the total number of death is total number of death is 5151.

Total number of death is 5151. No sorry, the age group the total number of just a second what is the question again? Okay, yeah, so died in the age group of 64 to 75 that means total number is 1683, right? 1683 and underlying condition, underlying condition would be 1272. So I have to just get the ratios. Essentially, we have to understand this concept of how, you know I am basically intersecting the you know different conditions, right?

So that is my question number 1. The question number 2, I asked a person died due to COVID in XYZ town, what is the probability that he is in the age group of 64 to 75. Now a person has already died, right? So I am already having the information about the death and which is 6839. So I am talking about the entire you know population or all the samples of people who died.

So let us say probability of say people died due to COVID in XYZ town, alright? And in the numerator, I will have I will have uh those who died in in the age group of 64 to 70 uh 65 to 74. So that will be people died in the age group of 64 to 75, now is 65 to 74. You can also think that this is number of people died in the age group of this intersection, total number of people died in the, total number of people died in XYZ town, right?

So it it is the same basically, because it will be the same number here, right? So the answer will be number of people died in the age group of 64 to 75 is going to be how many people? It is uh 65 to 74, it is 1683 and total 839. So essentially it would be in your denominator you will have 6839 and in your numerator, you will 6839 and here you will have 1683, right?

So this is how you based upon the question asked, so you always kind of compute the probability using the frequency that you have obtained. So this is how you actually, you know use a frequentist approach of probability. So with this we will end the lecture on frequentist approach. And in the next lecture, we are going to cover a very interesting topic called Bayesian probability, where we are where we will talk about the posterior probability. All right. So with this, we end the lecture here.