**Transcriber's Name Priya Prashanth**

**Applied Econometrics**

**Prof. Tutan Ahmed**

**Vinod Gupta School of Management**

**Indian Institute of Technology-Kharagpur**

**Lecture No. #10**

## Bayesian Theorem

Hello and welcome back to the lecture on Applied Econometrics. We have been talking about a very interesting topic called Bayesian theorem. And we have seen the wide ranging application of Bayesian theorem. We try to solve couple of problems, uh basically first problem we tried to solve is how, you know like, on on the likelihood of one having COVID disease.

So essentially, we did not use any formula if you have noted. So I am kind of, you know deliberately avoiding formula because the formula looks a bit, you know daunting, it is bit heavy. So I really do not want to get into the formula first and make you scared. But my idea was to give you, you know, the concept of Bayesian theorem with using some graphs and charts okay, and some diagrams.

So I hope that you have, you have managed to grasp some of the basic notions of Bayesian theorem. Now what I am going to do now is to actually derive the formula using this diagrams and graphs, okay. So let us, let us do that. And I will be using the same problem of uh you know, the likelihood of COVID infection that we have just solved. So that is the agenda of this lecture.

And going forward, we will do another problem on Bayes theorem, so that the concept is pretty clear to us. So let us uh work on this. (refer time: 01:55)

Now you remember that we kind of what we are trying to actually estimate is that probability of someone is actually having COVID given that I am tested positive, okay. So this is what I want to achieve all right? Whereas the information I had is that I am actually tested positive. Now I do not know whether I have COVID or not.

So that is the basically, problem I have. Now what information I have? So let us try to recollect the informations we had and try to sort of write them in terms of probability. So let us look at what we had at the beginning. (refer time: 02:33)

So we spoke about some priors, right. So we said that the prior here is one person in every 1000 person has COVID. So that means I have a probability. So let us say probability of COVID in the population is say 1 by 1000 because 1 in every 1000 individual has COVID. Now that means probability of not COVID let us say probability of no COVID is 1 or then it is 999 by 1000.

So every 999 of the 1000 people do not have COVID. So I can write probability of no COVID is this. Now I was also given some uh information about the test accuracy. And let us say the probability of the test accuracy or probability of one having a positive given that the person has COVID. So one person has COVID and the test will tell us that the person actually has COVID which means the test accuracy is 99%.

So which is 99 by 100 okay. Again in the same way we can also say that probability that the person has is tested positive, whereas the person has no COVID, if a person has no COVID, but the person is tested positive that is 1 by 100. How come I got this two number? So if again go back to this diagram. (refer time: 04:06)

If I if I am actually, uh I actually have COVID and because the test accuracy is 99%. So I have a chance of 99% being detected as COVID. Whereas, if I have no COVID the probability that I will be detected positive is 1 by 100. So this these are the information I already have, okay.

Now I what I have to do here is I have to kind of create a mathematical formula to sort of explain, uh you know, I actually actually have to derive the, you know, formula using all these different probability values, right? So let us again, look back, go back to the diagram, okay. So always diagrams and graphs are the best way to sort of, you know, have a sense of how it is actually happening.

We will get to understand the mechanism pretty nicely. So what we had is that we had here, let us say the B okay. So B means one having COVID, one actually having COVID and being detected positively, right. So that is essentially my probability of COVID, you know actually having COVID when I am detected positive.

So my all possible cases are I am detected positive and my numerator is going to be that I actually have COVID given that I am detected positive. (refer time: 05:26)

So this is this is the case I actually have COVID given that I am detected positive right. So this this is what is going to be the numerator, right. This this this part. So to achieve that what I will do is, so I will have say first this is my probability of COVID. Let me say probability of COVID into probability of COVID into probability of COVID, probability of sorry positive given COVID and total number of cases N okay.

So this is essentially my B, the B that I have used when I actually use this diagram. So this is my uh basically the right cases, I mean the the actual cases where I have COVID and I am undetected positive, okay. Whereas, I need to get the other value of A which is probability that I am detected positive but I do not have COVID. So probability of no COVID, no COVID but I am detected positive.

So that is probability of positive given no COVID, all right, and then with the total number of cases. So that will give me my A alright. So if I plot them, let me use a different color, here if I plot them so my, probe so that is essentially I want to get probability of COVID given I am positive is is is my B by A plus B right. So what I will get is this essentially means B is this one N into probability of COVID into probability of positive given COVID and in the denominator I have the you know the same B and A.

So same let me first write down the A. A is N into probability of not COVID, not no COVID into probability of P, which is detected as positive given no COVID plus the whole thing in the numerator. N into P COVID into P given positive given COVID, okay. So again let me reduce the size further. So that means, I will sort of N will cancel out from numerator and denominator.

So I will have P COVID which is 1 by 1000. 1 by 1000 into probability of detected as positive given COVID which is probability of positive uh detected positive given COVID is 99 by 100. And in the denominator I have probability of no COVID which is 999 by 1000. And I will have probability of detected positive given no COVID which is 1 by 100 plus the whole B, which is 1 by 1000 into 99 by 100.

And now, if I just calculate it, what I will have in the numerator is 99. So basically we will cancel out 10 to the power 5 from both you know numerator and denominator. So what I will be left with is this 99 by 99 999 plus 99. So which is essentially which essentially is going to be uh we can actually calculate it. It should be something around let me write down 99 uh 1098, we calculated already.

So this value was, this value was this, this value was this 9.02, okay, this value was 9.02. So we will essentially get the same value of course 9.02%. So this is the notation that we have developed. So let us now generalize the notation. Given the problem we have solved and different notations we have used now we can generalize it.

So let us do that. So what we have finally identified is that probability of COVID given it is positive is essentially probability of COVID into probability of positive given COVID by the same, let me actually reduce the size a little bit, same probable probability of no COVID into probability of positive given no COVID plus what we have in the numerator probability of COVID into probability of positive, given COVID.

So this is the formula. So you can basically use A and B to sort of identify those events. So I have just, you know, illustrated this formula using the problem that we have solved. So that is how we actually use the formula. I actually can derive the formula from the examples that I have solved. So we have seen the graphical representation, how we can understand the Bayesian theorem and how we can apply it.

And I have also derived a formula from the graphical understanding of Bayesian theorem. Now so that is the introduction to Bayesian theorem and we have done a little bit hands on. Now the interesting part will come when we will try to see the repeated trial and how do we really update the probability and that is what we are going to see in the next lecture.