Transcriber's Name Priya Prashanth

Applied Econometrics

Prof. Tutan Ahmed

Vinod Gupta School of Management

Indian Institute of Technology-Kharagpur

Lecture No. #12

Bayesian Theorem (Contd.): Example of Diamond Identification

Hello and welcome back to the lecture on Applied Econometrics. We have been talking about Bayes theorem and in the previous lectures we have seen actually how to use Bayes theorem. And we did solve some problems, particularly for a COVID patient who is repeatedly getting trialed.

And actually we are getting going to see how the probability of the, of our belief that the person is actually having COVID is updated with repeated trials, right? And we did try to understand the Bayesian theorem and its application we using, you know different graphical illustrations. We also derive mathematical formula from there.

And we also have seen with repeated trials, how the probability value of our belief is actually updated, right? So let us use the same concepts here. And we, you know we will try to solve another problem so that our concept about the Bayes theorem is, you know sort of, we are fluent with this. And this problem, let us say is about identifying a diamond.

Let us say you have read in a novel that diamonds are available in African you know river basin, okay. And you are adventurous person, and you decided that you will actually go to Africa to find out diamond. Let us say you have gone to Congo. And let us say you have met the chieftains there and you got the information from the chieftain that yes he has welcomed you for your bravery.

And he said that alright son, you can actually go and find the diamond, but you know I will give you an information. And that is in every once every 100 stone, you can find one diamond. So if you pick up 100 stone, you will have one diamond, right? So that is what he has said. Now you are, you know advanced guy, and you actually, you know have all these equipments with you, and you have taken an equipment that identifies diamond.

And let us say that equipment has 95% accuracy, okay? Now let us actually write down from this informations, whatever we have so far, let us try to write down uh, you know different different probability terms that we are going to use to calculate the probability of finding a diamond. So first thing is the prior. (refer time: 02:29)

Prior, so the chieftain has already given you the prior. So prior is probability of diamond is that out of every 100 stones, you have one diamond, so probability of diamond is 1 by 100. The second probability that you also got from the chieftain is the probability of not diamond. So all the stones you pick up what is the probability that if you pick up one that is not a diamond, so that is going to be simple 99 by 100.

Now you have your equipment that identifies if a stone is a diamond or not. So you have like different criteria, like you know sharpness, cut and so you know brightness and whatever. So those criteria is actually helping that equipment to identify something as diamond. But equipment being an equipment, it has also it has an error chance and that is let us say 95% accuracy you have the equipment.

Now if I write that in terms of probability, how I will write is probability that you actually get a positive result given that it is diamond. So that is that means if something if or if your stone is actually diamond and you get a positive result, so that is a that chance is 95% right? Whereas, if it is a stone and you actually get a positive result, probability of positive given that it is not diamond, so because that is the error part.

So that is the 5 of 100 okay? So these are the probability values you already have, okay? Now let us try to recollect the concept of you know either you can use the area chart, area area plot or you can use the uh graph plot that we have seen in the previous examples and we can let me again show you. (refer time: 04:17) So this is the graph plot.(refer time: 04:19)

And this is the area plot to identify the probability value in this case. So let us say my probability value in this case is going to be I will, what I will do here I will basically create that area again and this area is going to be so let us say probability that is 100. So if I you know have a you know sort of sample or population of 1000 so I will have 10 diamond and 990 non-diamond. So D and ND, okay.

And out of which so if I actually run the test, since the test has 95% accuracy, so what I will have this part after the test which is 9.5 will be identified as diamond, right? And this part because of the test inaccuracy, I will identify this part to be diamond and this is going to be

990 into 5 by 100 and if I actually multiply this, so it is going to be so let us say 2 and it is 49.5 This part is 49.5.

So this is this is not diamond, but you because of the error in your machine you actually identify this part to be a diamond. So that means, the probability of diamond, actual diamond, given that your test is positive, this is what we are interested in, is going to be your this part, this is the true diamond you get after the positive test and which is nothing but 9.5.

And in the denominator you have 9.5 plus 49.5 which means 9.5 and here you add 59. And what is the probability what is the value here? Let us actually calculate that and I will actually calculate. So let me do. So I what I need to estimate is 9.5 by 59 is going to be 16.1. 9.5 by 59 is 16 is 16.1%.

So this is the probability actually that despite being identified as diamond, you know this is the, only only there is a 16% chance that yours, what you have identified as diamond is diamond. And the whole thing is coming from the prior the population characteristics and the test accuracy, right? So that is how we have learned previously how to estimate or the Bayesian probability.

You can also use the same concept of like you know tree branch like the graph graph plot where you will have say you know the population where you actually have diamond which is say 9 out sorry actually have diamond which is going to be let us say this is the diamond part. So 10 and here it is not diamond is 990 okay. And then you do the test and one comes out to be positive, one comes out to be negative.

Here also the same thing. So if the test is positive, so in in, this is the true cases of diamond. These are all diamonds, and these are all not diamonds, if it is you know positive. So because your test accuracy is 90% 95%. So you will have 9.5 here. And if it is negative, which is wrong, here for the first case, you are going to have 0.5. Whereas here if you have a positive, it is wrong, because they are actually not diamond.

So that is going to be 990 which is 990 into 5. This is going to be 990 into 5 by 100, which is we have already calculated, which is 49.5, which is 49.5 which is 49.5. And this part is going to be the rest, which is 990 into 95 divided by 100 okay? So that is basically, if I just subtract it, it is going to be 940.5. I think I am correct in my calculation. So these are the values I get.

And if I plug in these values here, if I plug in these values, like this one, I am concerned about this positive values. So if I plug this is my B this is my A and if I do if I do B by A plus B we will get the value here, 16.1%. Basically this is the same problem. It is just the way you address is a little different just but the conceptually they are basically the same thing. Now we will do the trial part.

So this is the first you know time you do the test and first time you calculate the Bayesian probability. Now what how I am going to update my belief. So that is what we are going to see, trial. (refer time: 09:19)

Now when I am going to do the trial, what are the things that are going to change? So when I am going to do the trial, only thing that is changing here is this. The population here is going to be this. I am no longer bothered about this people here, okay. I am bothered about this true population which is this and the other population who are wrongly identified to have, to be uh wrongly identified to be diamond.

So that means my actual population my N is going to be the true diamond and the falsely identified diamond which is 49.5 okay, so which is 59.5. Now if I want to, I did not, so if I want to get the probability of diamonds, so that is going to be now 10 by 59.5. This is updated. Earlier I had this 1 by 100. Now it is 10 by 59.5. And probability of not diamond, probability of not diamond is going to be 49.5 by 59.5 all right?

So these are the things I actually get updated. The other things the test accuracy, they remain the same, because nothing is changing for instrument. So it is going to be probability of positive, given it is diamond. So basically correctly identify that is 95 by 100. And probability of negative negative given it is not sorry, probability of positive probability of positive given it is not diamond is going to be 5 by 100 because that is the mistake, that is the error that you do because of your test, right?

Now what I have to find is I am just going to plug these values into my formula, and I will see what is the updated probability. Now what I have to find is that, my purpose here is I will just minimize the screen a little bit. So what I want to find is that probability and I will probability of diamond given the test is positive, okay? So that is what I am trying to understand, is going to be basically P(D) probability of positive given D.

And in the denominator, I have P(D) into probability of positive given D plus P(ND) into probability of positive given ND. So that is what I simply try to understand. And let me write it from here. So P(D) is going to be, let me use actually a thinner line, 10 by 59.5 into probability of this P when D is 95 by 100. And here again, I will write the same thing 59.5 into 95 by 100.

And here, what I will have is 49.5 by 59.5. Because that is my probability of ND. And probability of positive given ND is just 5 by 100 okay. So now it is a simple, just a calculation. And all the denominators in the numerator and denominator will actually cancel out because they are same. So what we will have is 950 in the numerator. And the denominator, you will have there 950.

And this part is going to be 49.5 into 5. So which means you will have I think 247.5. Now if you calculate that, what you get is, what you get here is, let me write a different color now, 950 because that is the result I am going to get. And this is going to be 1197.5. And if I actually calculate that, we actually did that 950 by 11 97.5, which would mean 79.3%. 70, I will use a bold color here because that is the result 79.3% 79.3%.

Essentially so that means my probability is actually increasing from 16.1% to 79.3%. So that means, with repeated trial, if my results are coming out to be the same, the positive, then my probability is actually jumping significantly. The quantum jump is quite significant. And that is the beauty of Bayes theorem. So the moment I actually get repeated confirmation, my probability value increases quite significantly.

So that is how we actually see how our belief is updated with the uh you know repeated trials. And that is where the beauty lies in Bayes theorem. And there are you know significant uh you know applications as we said at the beginning of this lecture of Bayes theorem. And this is this is simply the, you know basic building block of that, and there are, you know plenty to learn about Bayes theorem.

If you are interested, you can actually explore that. I talked about Bayesian network. So with this, we will end this lecture on Bayes theorem. And in the later lectures we are not we are not going to use a lot of Bayes theorem. Somewhere we will use some some concepts but not quite significantly. We will remain in the realm of mainly frequentist approach. So with this, we end the lecture here and thank you.