NCRM Intro to BDA 3 Bayes rule

Let's take a look at another example. This time, we are going to take a story, formulate a model of data generation based on this story, formulate a likelihood function, specify the prior beliefs, and update those beliefs using the likelihood function that we have just constructed.

So suppose we are sitting at home, and we want to know whether it's cold, warm or hot outside. One of the ways of doing this is by looking through the window and observing how the people outside are dressed. For the sake of simplicity, let's assume that there are only three types of weather cold, warm and hot. Let's also assume that there are only three type, two types of outfit that people can wear, they can either wear coats, or they can wear sandals, they cannot wear coats with sandals or wear neither coats or sandals.

So we know that if it is cold, people are more likely to wear coats than wear sandals. When it is warm, people are wearing coats at sandals with roughly the same probability. And if it is hot, people are more likely to wear sandals than to wear coats, I'm going to assign the probability of 0.7 to the event that people are going to wear coats when it's cold. And I'm going to assign the probability of 0.2 to the event that people are going to ware coats in case it is hot outside, so the conditional probabilities. So we glance outside, and we see that there are four persons who are wearing coats, and there is one person wearing sandals. Okay, now that we have put down all the details of our story, we can start building our model of data generation.

And so in the story, the main parameter, the main unknown is the weather. And we know that it can take only three values, it can be cold, warm or hot. The data can be summarised in with two numbers. I'm going to call this numbers y and n. n is the number of the total number of observations that we've made by looking through the window, it's in this case is going to be equal to five. And y is the number of people who are wearing coats, it's values going to be four. One way of modelling the data generation process that generated these two numbers is by simply saying that the number of people who are wearing coats, y follows a binomial distribution with a probability equal to lambda, which depends on the weather. And the number of observations, the number of trials equal to n. Lambda in this case is equal to 0.7. If the weather parameter is equal to cold, it is equal to 0.5. If weather is equal is warm, and it is equal to 0.2, if weather is hot.

Given this model of data generation, we can now construct it the likelihood functions, electrical function is going to produce three numbers, one number for the case when the argument the main argument, the weather is equal to cold, one value for the case when the weather argument is equal to warm and a one number for the case or when the weather is hot. Here down the street, you can see the directly correlations. For this function for the electrical function, it is basically the application of the probability mass function for the binomial distribution.

T

In principle, you can do it easily by hand. But you could also do it using the software. For example, you could use the D by norm function in r. This likelihood function has the highest value with the value of the weather is equal to cold. And this should not be surprising, because we observed that almost more than half of the people, about 80% of the people outside are wearing coats. And so since when people are more likely to wear coats when it's cold, it's very reasonable to conclude that it is more likely to be cold outside than not. Now, now that we have these likelihood functions, we can go ahead and try applying twice. Define the prior beliefs about the weather parameter and updating those beliefs using these likelihood functions.

Oftentimes, before we observe any actual data, we do not have any additional information about the distribution of parameter values other than what kind of values this parameter can take. In such cases, we want our analysis to reflect such lack of information, we do not want our posterior to be in any way affected by by the prior. In such cases, we want to use what's called non informative priors. Non informative priors are the priors that do not favour any specific value of the parameter. In this case, in the case of our example, with the weather, we can specify in an informative prayer by setting the probabilities of the parameter being equal to cold, the parameter being equal to warm and that parameter unikl too hot to the same value. We can do this by setting the probability of the weather being cold to one third, the probability of the weather being warm to one third, and the probability of the weather being hot to one third. This way, each of the three possible values of the weather parameter is equally likely.

Now, let's talk about the Bayes rule, and incorporate this information into the calculations of the posterior. According to the Bayes rule, the posterior probability of a particular parameter value is proportional to the prior probability of the parameter value multiplied by the likelihood associated with that parameter value. In the case of our example, it can be stated as follows. The probability of it being cold, after observing the data is proportional to the probability probability of it being cold before we observed the data, times the probability that we would observe the data the way we observed it, given that assuming that it is cold outside. We can make the same statements about the other two possible values of our weather parameter.

Now, the posterior probabilities are probabilities. And so they should range between zero and one and add up to one. Therefore, in order for us to calculate the actual posterior probabilities, what we need to do is to take these products and normalise them to a range between zero and one and to add up to one. And this is how we can do that. We can compute the products of the likelihood and the prior for each of the possible values of our parameter, sum them up all together and use that sum as the denominator. The denominator is used to divide to adjust the product of the prior and the likelihood associated with a particular parameter value.

To give you more intuition about these calculations, let me show you a picture. Imagine that we are sitting in a room and we have not observed any of the data yet. This rectangle here represents the universe of all possibilities. We can divide that rectangle represented the universe of four possibilities into three stripes, three events, the top stripe will correspond to the event that it is cold outside, the intermediate stripe will correspond to the event that it is warm outside. And the bottom strap will correspond to the event that it is hot outside. The shares of the area of each stripe in the total area of this big rectangle is the probability, is the prior probability associated with that parameter. Because we want to be agnostic about the weather outside, we set the priors to be the same and so the shares are of these stripes are the same on this picture. Now we can use our knowledge about the rules that people apply to deciding what to wear, given the weather outside to break down those rectangles, those stripes even further.

Before observing any data, the probability that the person wears the coat and it is cold outside can be represented by this segment, the segment is part of the top stripe. The proportion of this segment in the top stripe is equal to the probability with which a person would wear a coat if it is cold outside according to the remaining segment of that top stripe corresponds to the event when it is cold outside and the person is wearing sandals. The share of this remaining segment in the total area of the top stripe corresponds to the probability that a person would wear sandals given that it is cold outside.

Similarly, we can break down each of the other two stripes into two smaller pieces, a piece that corresponds to the event that the person is wearing a coat and the piece of corresponds to the event that the person is wearing, sandals. Now, all these calculations can be done before we actually observe any data. So we can break down the universe of all possibilities into the events according to the our beliefs about it being cold outside and about and the rule that the people would adopt as they decide what to wear, given the weather outside.

Now, once we observe the data, once we observe a person wearing something, we can get rid of all of the rectangles are all the segments of that of this big rectangle that have not happened. If for example, we have observed that a person is wearing a coat, we can remove this segment, this segment and this segment, which will give us the following weirdly shaped picture. Now, the share of that top segment, or the the event that is cold outside and the person is wearing a coat in the total area of this funny looking picture is the posterior probability of it being cold outside. This is what we should believe is the probability that it is cold outside, once we have observed a person wearing a coat.

Similar logic applies when we do not observe only one person, but we rather we observe a bunch of data points. Similarly, to obtain the posterior probability of delivery with a weather being cold, we need to compute the area of the top segment and divided by the total area of this funny looking shape. This is what's happening here. And this is exactly the Bayes rule.

Now let's do some calculations and actually produce the posterior probabilities of it being cold outside, warm outside and hot outside given the values for the likelihood function that we have, and our non informative priors. To do that, we need to multiply the prior probability of it being cold outside by the electrical associated with it being cold outside and write it down in this row. Prior term likelihood, we need to multiply the probability of it being warm outside with the likelihood associated with being warm outside and we need to multiply the probability of it being hot outside by the likelihood associated with it being hot outside. Once we have these products, we can normalise them to obtain the posterior probabilities, the posterior beliefs and here are the results of such calculations.

T