NCRM Intro to BDA 2 DGP and Likelihood

In a moment, we are going to see how we can use data. To update our initial beliefs about the unknown scenario model, primarily about the model parameters. We are going to use what's called biases rule or bias theorem to arrive at the posterior distribution. Before we can do that, we need to briefly discuss the concepts of data generating process and likelihood function.

These two concepts are critical for understanding how we can connect data with the model and the model parameters. Well, the data generating process is a model that describes how our data could have been generated. It is a chain of hypothetical events that lead to each and every observation in our data set being observed. This chain of events can include both deterministic steps and random steps.

Deterministic steps are the steps that for each combination of inputs return a unique value for the output. Random steps are the steps that for each combination of inputs, return a distribution of possible values of the output. Data generating process can include the unknown parameters. Typically, when we have unknowns, when have parameters in a DGP we seek to use data to assess what the values of these parameters are, in order for us to be able to draw predictions to be able to characterise the effect of variable and to be able to extract other quantities compute other quantities of interest from using our model.

Now, likelihood function is in a way a summary of the data generating process. The electric function is a function of model parameters, and it is specific to the DGP and to the data that we have on hand. The electrical function is typically computed as the probability that the data that we actually observe could have been generated using our model with the specific values with the given the variance of the model parameters. To see the DGP and the likelihood function at work, we are going to take a look at an example. And we are going to look at the logistic regression.

Logistic regression, logic is a model that is used to connect independent variables with the binary outcome variables. A binary outcome variable is a variable that takes the values of zero, or one. One way to think about the data generating process behind the logistic regression is to think of it as consisting of two steps one deterministic step and one random step, one stochastic step. The first step, the values of the independent variables specific to each observation in our data, together with the values of the coefficients, determine the value of theta, for each observation that we have in the data set. To mimic this process, we could take the vector of the values of the independent variables for each observation, multiply them by the vector of coefficients, and apply the inverse logit function. The random step at the stochastic step nature draws the actual observed values of the dependent variable from Bernoulli distributions with the probability parameters equal to the status. The value of the independent variables and the value of the dependent variable are known and observed. The beta decantations, as well as the parameters of the Bernoulli distribution used to draw the values of the dependent variable are not observed they are unknown. So let's try computing constructing a likelihood function using this information.

In addition to this description of the data generating process, we are going to need a hypothetical data set. So here on the left side of the slide, you can see hypothetical data set with two independent variables and one dependent variable and three observations. The variable x-1 the value of one, two and zero, the variable x-2 the second independent variable takes the value of negative one, one at three, and the variable, the dependent variable has the values one, zero and one.

Let's follow through this data generating process until we arrive at the probability that we could in fact, observe the observed value of the dependent variable in each of the cases in each of the rows of our dataset. And the first step, we can compute theta for each row that we have in our data set. To do that, we are going to need to take the values of independent variables and multiply them by the parameters beta one and beta beta two, we're also going to add intercept, zero to the sub. Once we do that, we can apply the inverse logit function. And this will give us the formula for theta I for each row in our data set.

As we know from the descriptions of the data generating process, nature draws the observed the actual observed values of the dependent variable from Bernoulli distributions to be the probability parameters equal to those betas. Therefore, the probability that we could observe the value of one is our the value of the dependent variable is going to be equal to the setup. And for each observation, the probability that nature will return the value of zero as the value for a dependent variable is equal to one minus theta. And so if we actually observe the value of zero, then the probability of observing the actual observed value of the dependent variable is going to be equal to theta i, if we actually observed one, and is going to be equal to one minus theta i, if we observe the value of zero.

The likelihood function is constructed as the probability that we could observe the data set that we actually observed, given the variance of the parameters. And so because we drew our observations independently, that likelihood function is going to be equal to the product of the probabilities or the value of the dependent variable in each observation, given the values of the independent variables, and beta taken separately for each row. Once we combine this information, we can arrive at the expression that you can see in the last row on the slide, and the seas, as you can see, the likelihood function, it is a function of of betas, and only of betas.

So what can we do with this function? Well, if we were to use the maximum likelihood estimation approach, we would try finding betas such they maximise the likelihood this likelihood function. In the Bayesian context, we are going to use the likelihood function in a different way. As discussed earlier, in the workflow of Bayesian analysis, we are going to start by treating each parameter as a variable. We are going to figure out which values they can take and assign a distribution, a probability distribution to each of those parameters. That's going to be the prior distribution. Then, according to the biases rule, the posterior distribution is just a combination of the likelihood and the prior according to this formula.