Applied Econometrics

Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology-Madras

Lecture - 30
Course outline for Applied Econometrics

Qualitative Response Model-Part II

So, estimation of logit model that we are going to discuss, estimation of logit model. So, using how we are going to estimate using maximum likelihood function or maximum likelihood estimates. So, let us say I am writing maximum likelihood estimation, estimation, which in short known as MLE, maximum likelihood estimation that we are going to use. But before we talk about maximum likelihood estimation, we need to know what is a likelihood function, what is a likelihood function. This is very important likelihood function. Can you think of what is a likelihood function? The concept of likelihood function is actually closely related with probability distribution or probability density function.

So, if you know probability distribution function or probability density function, then you can easily understand the concept of probability likelihood function. Now, what is a probability distribution function? In any probability distribution function, if you think of let us say, binomial, Poisson or any other type of probability distribution function. Suppose you are thinking of the simple probability distribution function, the binomial function. And the probability distribution function, how it is characterized? Let us say I am writing PDF, probability distribution or probability density function.

PDF is characterized by a observed thing given the parameters. I will explain this. When I am specifying a particular probability density distribution function, let us say I can find out from there, if I toss a coin, let us say 15 times, what is the probability of having let us say 5 heads, 5 heads. So, that means from PDF, let us say from binomial distribution function, which is given by two parameters n and p. So, I can answer this type of question.

If a coin is tossed in times that means, let us say, 15 times what is the probability of getting 5 heads. Let us say this getting a head we are calling it, this is success. That you can find out if the probability distribution function is given. Now, if I reverse this question, so if I reverse this question, here the parameters are known and you are trying to find out the probability of getting 5 heads out of 15 trials. So, you are tossing the coin 15 times that means, number of trials is 15 and you can find out what is the probability of getting 5 successes or 5 heads in 15 trials.

What is the reverse of this problem, can you think of the reverse problem? Reverse problem, reverse of this that means, I know in my sample, how many trials I have made, let us say 15. That means, let us say I am asking 15 individuals, so my sample size n is 15. So, there are the reverse problem is n is known, sample size is known, known. I also know that in that out of these 15 individuals, how many people they have car? Here, I was trying to find out that, how many heads I will get that means, what is the number of success? Here, if I define that owning a car is actually a success of that trial that means, I already know out of 15 individuals, how many of them are actually having a car? Then what I do not know is actually the parameters, parameters are unknown. So, that means, in a sample size of 15 individuals, 5 individuals are having a car that is my sample and I have already observed that sample.

Here in terms of O, O is actually the observed thing, which in probability distribution function you are trying to find out, probability of getting 5 success that is denoted by O. And what is known to you, is θ, which is the set of parameters. So, here, 0 in sorry O indicates observed thing and θ basically indicates set of parameters. In binomial distribution function, this P, the parameter is actually known for a single toss, it is half that is known to you. So, given θ, set of parameters in probability distribution function, you are trying to find out probability of O that means probability of 5 heads, but in a likelihood function, I already know that, that is known to me, what I try to find out is θ .

So, this is unknown, that means parameter are unknown, here this is known. That is why I say that likelihood function is just the reverse of probability density or distribution function. In likelihood function, I would like to get the parameter, because a set of parameter will only define the PDF, so that, I can observe that sample with maximum probability, that is why it is called maximum likelihood function. That means, in short, what I want, I already observed a sample, where out of 15 individuals, 5 individuals are owning a car, you give me the set of parameters that will maximize the likelihood of observing that particular sample, because from a given population, I could have observed some other samples, some other samples of 15 individuals and in that sample, it would have happened that out of 15, 2 individuals are having a car or 14 individuals having a car or 10 individuals having a car. So, if you change the parameter, you will get different type of probability function, distribution function.

So, I would like to get that set of parameter that will maximize the likelihood of observing that particular sample. And that is the technique what we apply here in maximum likelihood method, instead of minimizing the error term that we used to do in ordinary least square method. So, this approach is little different, maximum likelihood estimates is different from OLS in that regard. In OLS to get the parameters, we minimize some of the error terms, that means, summation $\sum_{\widehat{ui}} 2$. Here, we are maximizing the likelihood.

What is the likelihood? Likelihood of observing that particular sample, where out of 15 individuals, I already observed that 5 individuals they own a car. So, that is just the reverse of probability distribution function. In probability distribution function, given θ, I would like to estimate the probability of a observed thing. And what is the observed thing? That probability of 5 heads that I would like to observe. Here in likelihood function, I have already observed the O, but I would like to get θ set of parameters that will maximize the probability of observing O, that means, probability of observing a particular sample.

In which 5 out of 15 individuals own a car, that is actually the likelihood function. So, this reverse problem will give you likelihood function. Now, how will you apply this likelihood function in the context of a logit model estimation? So, let us say that the likelihood function is L denoted as probability of observing a sample. And as you know, probability of observing a sample means probability of observing y1, probability of observing y2, probability of observing y3, dot dot dot, probability of observing yn. So, your sample consists of n number of responses, basically, y takes the value 1 and 0, this is the response function.

So, if you observe all your yi that is basically the likelihood function. That is basically the likelihood function. And this, so that means likelihood function is the joint probability of observing y1, y2, y3, dot dot dot, yn. So, that means this is nothing but probability of y1, then probability of y2, then probability multiplied by probability of y3, then probability of observing yn. This is the likelihood function, which is nothing but the joint probability distribution function.

Now, suppose I have arranged the sample, suppose I have arranged the sample in such a way, suppose that the first is one observation or is one individuals, own a car and next in two individuals do not own a car. That is how I have actually arranged. First, that means, where n1 + n2 = to n1 + n2 equals to n, which is the total sample size. So, that means what I am saying, probability y1, y2, dot dot dot n1 up to n1, it is actually, it is actually pi, because pi is the probability of observing the sample. So, from here what I can write, equals to p1 * p2 * p3, how long this will continue up to p of n1, because, first n1 individuals will own a car that is why p1, p2, p3, dot dot dot p n1.

And what would be the next term? Next term would be 1 - p n1 + 1, because next person does not have a car that is why 1 - p. For first n1, p1, p2, p3, dot dot dot p n1, next observation, next individual does not own a car that is why 1 - p n1 + 1, then 1 - p n1 + 2, dot dot dot 1 - p n. That is how I have arranged the data. Now, if you write this expression in a concise format, then what we can write, this is nothing but product of i running from 1 to n1 pi and then this is multiplied by i running from n1 plus 1 to n 1 minus pi, 1 minus pi. And this again you can write as 1 grand product pi to the power yi into 1 minus pi to the power 1 minus yi.

Now, this step is little interesting how I could write this step. From this step to this step, you need to understand how. See, this happens, this happens because yi = 1 for the first n1 observations and if yi = 1, then this would become pi and yi equals to 0 for the next n2 observation, that is why it will become 1 - pi because this will become 0. So, ultimately this will result in pi into 1 - pi. This happens because as yi equals to 1 for first n1 observations and yi equals to 0 for next n2 observations.

Now, if you take log, then what will happen? If you take log, log of l equals to what you can write? This you can say that log(yi), sorry yi log (pi), yi log pi plus 1 minus yi log of 1 minus pi, i running from 1 to n1 and here i running from n1 plus 1 to n. Why I have done that? Because log l, why I did this? Because log l is a monotonic transformation, monotonic transformation of l, monotonic transformation of l. So, this is the thing log l equals to yi log pi and plus 1 - yi into log 1 - pi. Now, you can substitute the value of pi, what is the value of pi? Where pi = $(1/1 + \mathrm{e})^{-\alpha+\beta}$. So, you can substitute pi here and then you are trying to maximize this with respect to α and β.

So, that means I am trying to maximize this with respect to α and β. So, I want that particular set of α and β that will maximize my likelihood function. And this likelihood function is basically probability of observing that particular sample in which out of 15 individuals, 5 individuals own a card, this is the mechanism. This is the mechanism of maximum likelihood estimates for estimating the logit model. Instead of minimizing the errors, some$\sum_{\widehat{ui}}$ 2, we are actually maximizing log l with respect to α and β.

And you will get the optimum α* and β* that will maximize probability of observing that sample. So, I will take that particular $\hat{\alpha}$ and $\hat{\beta}$. So, that means $\hat{\alpha}$ and $\hat{\beta}$ then will be used for inference making. So, in OLS, I was minimizing $\sum_{\widehat{ui}}$ 2. So, OLS in case of OLS, what you are doing, you are minimizing $\sum_{\widehat{ui}}$ 2, but in MLE, what you are doing is actually maximizing log of l with respect to α and β.

That is the difference between this and this. And there are certain advantages of this

MLE. The advantage is that this α* and β* what you get after maximizing log l, that is basically they are asymptotically efficient, and they approach normality that is also asymptotically. So, they are efficient asymptotically, that means what? That means they are efficient in large sample and they approach normal, they follow the normal distribution that is also asymptotically. That means the MLE method should be applied, in large samples only.

So, whenever you are estimating logit model, you have to keep in mind since the underlying estimation strategy is maximum likelihood and $\hat{\alpha}$ and $\hat{\beta}$ they are efficient asymptotically, that means they show the efficiency property in large sample, you must have a large sample to estimate your logit model. So, we should not apply the logit model in a small sample like 3, 30 or 40, we should have minimum 200 to 250 observations for logit model to estimate. So, with this, we are closing our discussion today. Tomorrow, we will discuss about another class or another particular model of binary response model or qualitative response model and then we will also estimate this, we will see how to estimate the model using a particular data set.