Applied Econometrics

Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Module - 3
Lecture - 27
Properties of Estimator

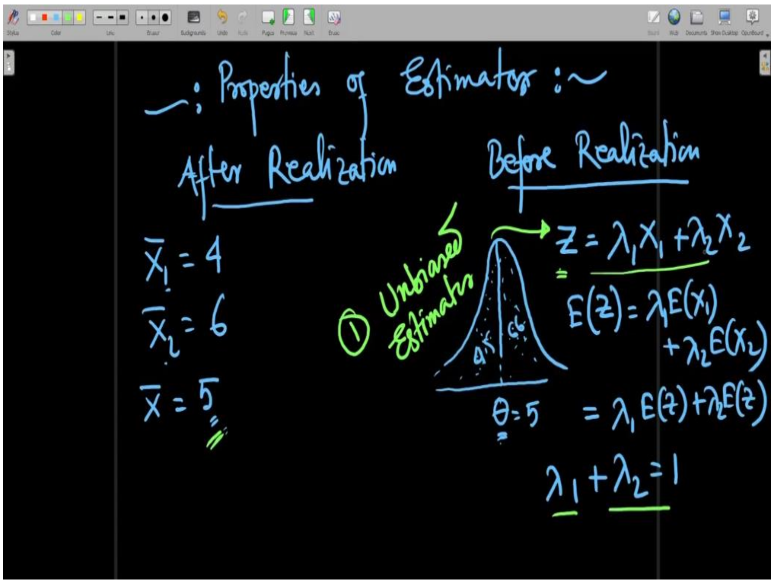
Hello and welcome back to the lecture on Applied Econometrics. So, in this lecture, we are going to talk about the properties of estimator. Now, estimator, as we have seen, is something that is a sort of a mechanism to reach from sample to population. So, we would never know the population characteristics; so, all you can do is, we can get some sample characteristics from the data we have drawn from the sample, and we basically use that mechanism to go from the sample characteristics to population characteristics.

So, that is what we call estimator. Now, this mechanism of estimator, it is not without any properties; it has certain properties and we will see, these properties are really important, because these properties are very helpful when we are deriving a regression equation and the coefficient. So, under what condition we should run a regression or what are the things we should take care when we run a regression, that will be coming from the properties of the estimators.

So, we need to sort of be familiar with the properties of estimator and why they are important. So, let us start with this. We have talked about the double structure of variable. So, if we have a random variable and if we run an experiment, so, the random variable can generate a potential outcome before you actually run the experiment. So, you have a distribution of the potential outcome.

So, that is something, is a before realisation thing; but when you actually run the experiment, when you actually get the result, that is something you get after realisation. So, using that idea, let me again use that idea to explain the properties of estimator.

(Refer Slide Time: 02:02)



So, let us say I start with after realization; it is easier to understand. So, let us say, I am interested to know the class quiz marks. Let us say I am actually having a very big class and I really do not have a system where I am keeping all the details. So, I just want to get an estimate of what is the average performance or what is the performance of my students like of the whole average performance of the whole population, but I do not have a means to sort of get all the details.

So, let us say I draw a sample of size n; it is a test out of 10, and I get the sample mean X 1 bar is equal to 4. Then I draw another sample and I see, the X 2 bar is equal to 6. Now, I could draw many number of times, but I do not want to do that. So, based on these two, I want to estimate the population characteristics, which is the mean, the average score of all the students in the class; and that is, let us say it is X bar.

And I look at these two values and I basically take an average, which is like, 4 and 6 average, this is basically 5. Now, how do I come to this number 5? What I have actually done is, I have assigned equal weight for X 1 and X 2; I have taken 0.5 as a weight for X 1 and X 2. How did I come to this number? So, that is what we need to understand. Now, this is the case of after realisation, I use this example just for illustration purpose, because it is easier to understand.

Now, before realisation is something that will help us to actually generalise the case. So, it is essentially, all the potential outcomes we are considering. Now, we say that for an unbiased estimator, the expected value of sample mean is equal to population mean. So, here, basically the average that we have taken, this is the expected value of sample mean that is equal to population mean. Now, why do we say that?

Because we know that we will never have the true population mean in my sample distributions. So, if I have all these different observations, they can be equal to the population mean only by accident. They can never be equal to the population mean; I mean, they can be, but that is by accident; but usually what you will see, all the numbers are, all the observations are actually spread across the true population mean.

So, if in the example, the true population mean is actually 5, so, I will have different samples I draw, some of the observation I will have like 4.5, somewhere I will have 6.6 and so forth; but everything will be pretty close to the number 5. So, that is what is the idea of; when we draw sample means and we actually go, try to infer the population mean; and the reason we take expectation is because of this, because I will never have the true population parameter, but I will have all the observations pretty close to the population parameter.

And when I actually take an expectation, which is basically an average, I get the true population parameter. Now, let us put it in a formula. So, let us say my estimator is Z; and when my estimator is Z, I can write it down as; let us say, I do not know the weight of my first sample mean, second sample mean; so, let me write down as lambda 1 X 1, lambda 2 X 2. Let us say, for the sake of simplicity, I am basically taking one observation instead of sample mean.

So, my X 1 is nothing but the sample mean, but I am just writing as X 1, not X 1 bar. Now, if I have this equation, so, I have to understand the expected value of Z, because that is what I am trying to get. So, the expected value of Z is equal to lambda 1, expected value of X 1 plus lambda 2 expected value of X 2. Now, I know, for an unbiased estimator, my expected value of Z and expected value of X 1 are going to be the same.

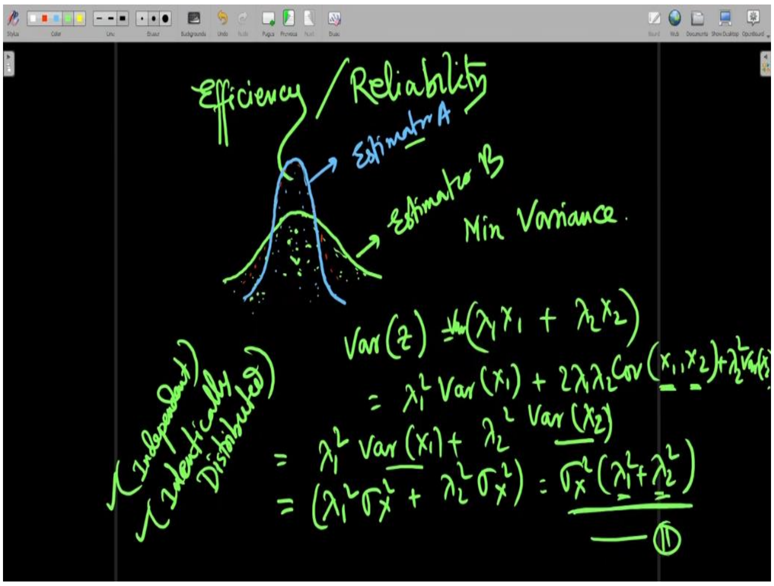
So, I will write lambda 1 expected value of Z plus lambda 2 expected value of Z. So, which essentially would mean my lambda 1 plus my lambda 2 is equal to 1. Now, let me write down; this is an unbiased estimator I am talking about, which is in this case Z. Now, for an unbiased estimator, what we see is, my lambda 1 and lambda 2 is equal to 1. Now, the moment I get this equation, I can have many values for lambda 1; many values for lambda 2;

lambda 1 can take 0.2; lambda 2 can take 0.8; lambda 1 can take 0.3; lambda 2 can take 0.7 and so forth.

So, there are infinite number of possibilities. Now, if I use all these infinite number of possibilities in this equation, I am going to get different value for X bar. So, they are going to be sort of different; I am basically, the moment I am putting different values of lambda 1 and lambda 2, my mechanism, the mechanism, the estimator that I said is going to be different. So, then, there are so many different estimators now.

So, which one to choose? That is the question. To decide on this, we have to consider the other properties of estimators. The first property is that we have to have an unbiased estimator; that is known; but that is not sufficient, so, we have to consider other properties. And the other property is that to have efficiency or reliability.

(Refer Slide Time: 08:00)



And let me explain what I mean by efficiency or reliability. Now, by efficiency or reliability, it means that the observations that I draw from the sample, need to be as close as possible to the true population parameter. So, what it means is, if I have my true population parameter somewhere here, I will want my distribution to be something like this; let us say a different colour. It could be something like this. I can have other sort of distributions of estimator.

So, these are all; or let me actually make it thick, so that these are distributions for estimator. Let us say this is for estimator A, and the green one is for estimator B. Now, what I will want is, the observations that I get, they should be pretty close to the population distribution, they should be as close as possible. So, that is what is represented by my estimator A. Whereas, for estimator B, I will see that the observations are spread across, their variance is very high.

So, essentially, it means that most of my observations from estimator A are coming to true population parameter. Whereas, for observations from estimator B are actually quite away from the true population parameter. So, we can take an example; but it does not mean that all my observations of estimator B are actually close to the true population parameter. So, there could be something here; there could be some values here; whereas, some observations under estimator B could be actually very close to the population parameter.

So, it is in general the case that most of the observations in estimator A are close to the true population parameter, whereas, most of the observations in estimator B are actually spread across basically. So, we can take an example; let us say the car accident, and if the person is wearing seatbelt. Now, in most of the cases, you will see that, if a person is wearing seatbelt, there was no accident.

But there could be cases where the person is actually wearing seatbelt, but still there is accident or death let us say, the accidental death; so, there are still accidental death. So, that is what, you may not expect that all the observations are going to be near to the true population parameter, but there could be some observations which would be a little away from the true population parameter.

Now, having said that, if I have to achieve this efficiency reliability, so, I have to have a minimum variance. So, essentially, what we try to do is, we actually minimise the variance. And if I want to minimise the variance, what I have to do is, I have to minimise the variance of my estimator. So, my estimator is this, and I have to minimise the variance. So, let us say I write variance of Z; I write it as lambda 1 X 1 plus lambda 2 X 2 and I use Var of this.

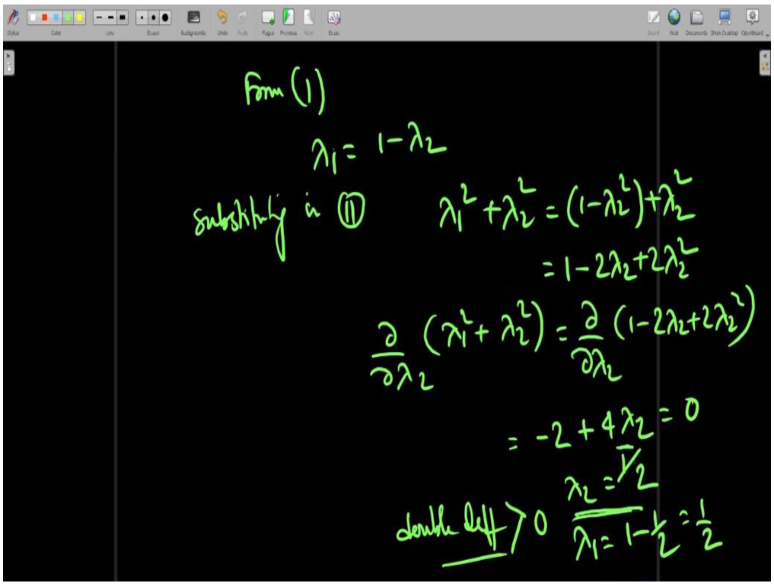
The moment I expand it, it becomes lambda 1 square Var X 1 plus 2 lambda 1 lambda 2 Covar X 1 X 2 and then lambda 2 square and Var X 2. Now, how can I then further simplify this equation? So, if I have this Covar X 1 X 2, it will be equal to 0. And the reason is, I have assumed that my samples are independent. So, they are not dependent on each other. So, the moment I have the samples independent, I will have my covariance term 0, which would mean I will have lambda 1 square Var X 1 plus lambda 2 square Var X 2.

Now, recall the fact that, all this, in this random variable that we are sort of drawing, in the experiment we are doing, are basically identically distributed. They are independent; so, that is one criteria. They are also identically distributed. Basically, they are IID random variable. So, this is an assumption we have made, because it will make our life easy. So, the moment I have them as identically distributed, I can simply say some; so, all the variances are going to be the same; so, let us say I write sigma X.

So, I will have lambda 1 square sigma X square plus lambda 2 square sigma X square. So, which would mean sigma X square lambda 1 square plus lambda 2 square. So, I got an expression of the variance. Now, I have to find out how can I minimise the variance. So, I have to sort of find out the value of lambda 1 and value of lambda 2 that will help me to minimise the variance.

Now, I will use some of the equations that I have already derived. So, let us say this is my equation 1 and this is my equation 2. Now, if I want to minimise this variance, so, I have to sort of get the minimum value of lambda 1 square plus lambda 2 square.

(Refer Slide Time: 14:02)



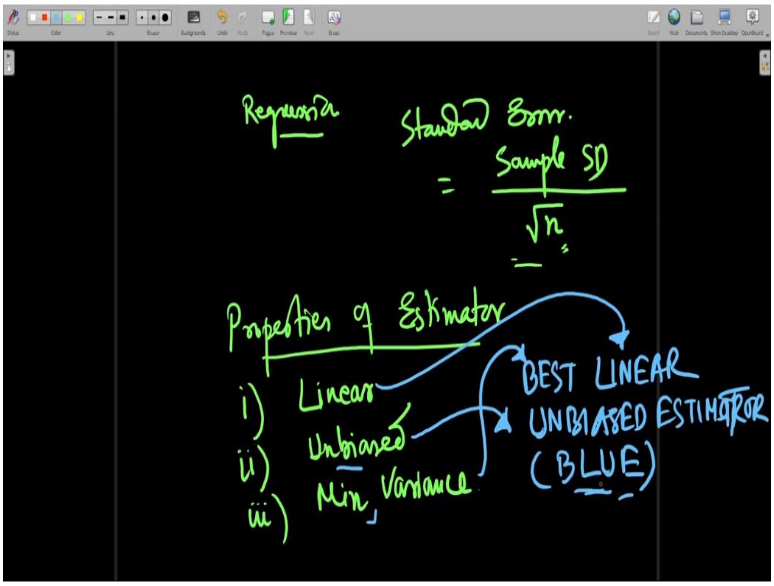
Now, from equation 1, I know lambda 1 is equal to 1 minus lambda 2. Now, substituting in equation 2, what I get is lambda 1 square plus lambda 2 square is equal to lambda 1 square or basically I can write lambda 1 minus lambda 2 square plus lambda 2 square. And if I expand it, I will have 1 minus 2 lambda 2 plus 2 lambda 2 square. Now, I want to sort of get the minimum value of it. And to get the minimum value of it, I have to basically minimise it.

And if I do that, so, essentially 1 minus 2 lambda 2 plus 2 lambda 2 square or essentially it would mean that minus of 2 plus 4 lambda 2 has to be equal to 0, because I am doing minima. So, that would mean that my lambda 2 is equal to 2. Now, if you actually; how do you know it is a minima? Because, if you double differentiate; double differentiate is basically positive, right; if you double differentiate, it is going to be 4; so, which is positive.

Sorry, lambda 2 is half. So, essentially, I get my minima at lambda 2 is equal to half. Now, if I substitute; so, essentially, what it means is that, for the case of all the lambdas; so, essentially, lambda 1 is also going to be half; 1 minus half is half. So, what it shows is that, for equal weights, if I have equal weights, then I will be able to achieve the minimum variance condition. Now, that is something that is very important finding.

So, for all the different values of estimator that we have got, you have to assign equal weight to them to basically get to the population parameter that you want to estimate. Now, that is well understood from this derivation. Now, the question that will come is, well, where in regression equation we really use this lambda 1 and lambda 2? I have never seen a lambda 1 and lambda 2 in regression equation. Well, so we need to explain that.

(Refer Slide Time: 16:34)



So, in a regression equation, you will see, what we do is, sometimes, we estimate the standard error. So, when we estimate the standard error, how we do is, we actually take the sample standard deviation. We take sample standard deviation because we usually do not have population standard deviation known. So, we use sample standard deviation and we do a root n, because we really do not want to draw the sample time and again.

So, essentially, we get all the observations in the sample and get the sample standard deviation and divide by root n, and which is approximately equal to drawing samples multiple times and actually getting a sort of a group estimate like here, this standard deviation. So, you simply, by dividing root n, you get that. So, here the assumption is, you are assigning equal weight to all of the standard deviation values you are getting, which you are actually not doing, but the assumption is that.

And that assumption is actually supported by the fact that all the lambdas, all the weights of the different estimators are equal. So, only if that is satisfied, you can actually say that the approximation you have done, that is correct. So, that is how we actually use the condition of minimum variance to, or the weight, from where you have actually got the weight of the, different values of the estimator to estimate the population parameter, they are actually going to be the same.

Now, with this, we can actually say the properties of the estimators; we can summarise the properties of the estimators. So, I will say that the properties of estimator; so, let me write down; properties of estimator: It has to be a linear; we are talking about a linear regression here; so, it has to be linear. It has to be unbiased; and we have just seen; it has to be unbiased. And it has to be minimum variance.

And the moment it satisfied all these conditions, we call the estimator as Best Linear Unbiased Estimator or in a word, in abbreviation, we can write it B L U E or BLUE. So, essentially, if I have to have the best kind of estimator in my regression equation, I have to have this property satisfied. So, what I mean by each of this term? So, by linear, I mean the linear in parameter that we have explained.

By unbiased, we have just seen unbiased, what it means, it means the expected value of the sample mean has to be equal to population mean. And the minimum variance, so, the minimum variance condition is called the best because, if your unbiased condition is satisfied, you can get a plethora of possibilities for different weights that you can put for different values of the sample; but then, in order to get the minimum variance, you have to satisfy certain condition.

And if you satisfy all these conditions, your estimator is going to be BLUE, Best Linear Unbiased Estimator. And we will see under this BLUE property of estimator, we will actually be able to outline all the different properties of the estimator, which we call the Gauss Markov assumptions. So, those assumptions are, you have to check when you are running a regression equation, whether those assumptions are satisfied.

If they are satisfied, then you will be confident about your regression equation that you have done it correctly. So, that is why this properties of estimator are really important, and that is the implication of the properties of estimator. We have to get the Best Linear Unbiased Estimator and we will see that in regression equation, they are going to be extremely critical. So, with this, we will wrap up the lecture on the properties of estimators.