Applied Time-Series Analysis

Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture - 94
Lecture 41A - Estimation Methods 1 -1

Very good morning, today we will begin with the methods of estimation; as I mentioned early on will look at primarily four methods, I would say four classes of methods, four different philosophies for estimation; method of moments and the least squares approach, the maximum likelihood approach which we have discussed mostly and then the Bayesian approach to estimation. Before we begin, I just wanted to make a couple of points with regards to what we have discussed until yesterday hypotheses testing.

As I said yesterday among with three approaches to hypotheses testing; my favorite one is the confidence interval approach and the reason is very simple. Suppose I have to redo this hypothesis test for a different postulated value. So, yesterday the value that we are postulated was 90 degree Celsius. Suppose I were to ask if 91; if I have to postulate the average to be 91 would such a hypotheses be rejected and if you were to use the critical value approach or even the p value approach one, you have to recalculate the observed statistic and then of course, compare it to the critical value or your alpha whereas, with the confidence interval approach the beautiful thing is the confidence interval is for the truth and of course, the confidence interval will only change; if the data changes and because the confidence interval has been calculated from your sample mean and sigma and root N if any of this changes a confidence interval change, but if the postulated it has got nothing to do with the postulated truth.

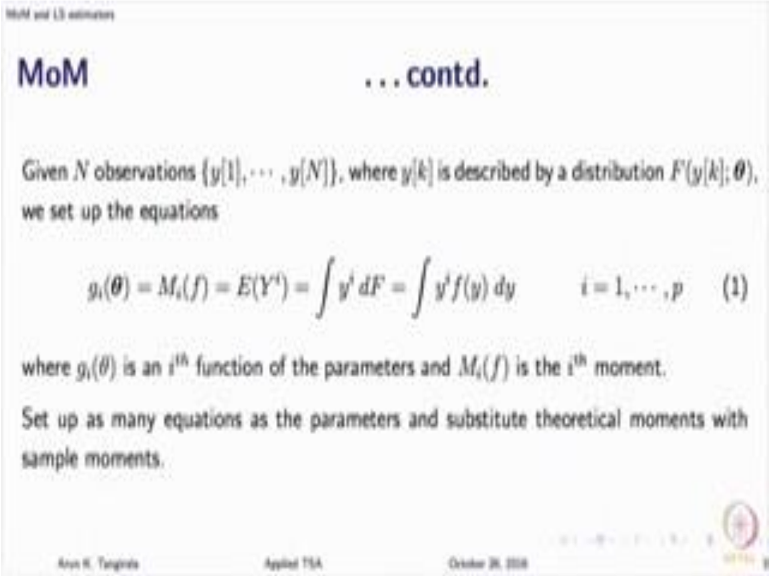
And therefore even as I change my postulate, my confidence interval remains fixed and I keep comparing; I keep checking if the postulated value is falling within the interval. So, if I were to ask if 91 is a postulated value, you can straight away look at the confidence interval and say well that hypotheses also has to be rejected and so on. Now having said that a word of caution is the confidence interval is not an interval of possible truths that is one of the misinterpretations that people generally tend to make, it is not an interval of plausible truths. It is just an interval in which the truth is we do not know where the truth is and sometimes people would believe that the truth is at one of the extremes no, we do not even know where the truth is; the truth could be midway in the interval or maybe at one of the extremes close one or the extremes and so on.

So, therefore the confidence interval should be just taken as an interval in which the truth falls, but when it comes to hypotheses and when it comes to hypotheses testing; it is a very useful approach. Therefore, I would strongly recommend using the confidence interval approach of course, the choice is yours. So, I just wanted to make this point and the other point is; you can conduct all this hypotheses tests and construct confidence intervals and so on using r, the relevant commands are given in the video lectures for example, t dot test would conduct a test for you for hypotheses concerning means then you have where dot test and so on; all those routines are mentioned in the video lectures, when I upload the slides; I will have a summary of the routines in r. Of course, when we talk about, when I show you how to estimate parameters using least squares in r then I will also show you perhaps how to compute the confidence intervals using some built in commands.

So, with this few words let me actually now get going with the method of a different methods estimation and we will begin with the method of moments. The method of moments is something that we have seen before, I have explained the principle before the idea is very straight forward, you write down the theoretical relations between the moments of the joint pdf and the parameters that you are estimating. So, the parameter estimation problem is straight forward; we have already stated the problem of parameter estimation. So, there are some parameters that have to be estimated; we write down the relationship between the moments of the pdf and the parameters and then assume that the estimates also satisfy these relations, which is a big assumption; they do not have to because they estimates over all always in error; what we begin with this relationship between the true moments and the parameters.

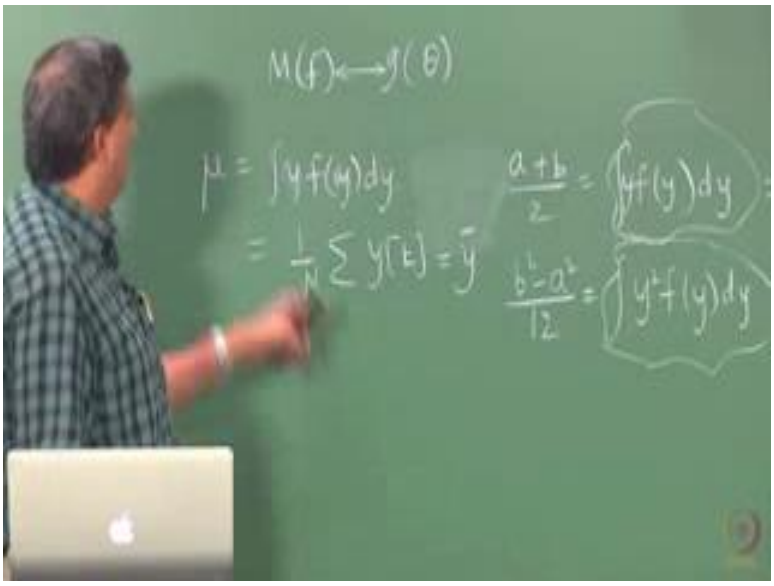
Now; obviously this means that right from step one the method of moments assumes that the data that has been given to you is coming from a random process. When we move on to least squares, we will say that we do not have to assume randomness in the data at all. Whereas, method of moments right from step one assumes that the data comes out of a random process pretty much like they likelihood approach alright.

(Refer Slide Time: 06:07)



So to give you the math behind it, essentially what you do is you have M I which is I th moment of the pdf and you have g of theta which is some function of the parameters. So, you necessarily do not get a form like this when I say that you have to strike a relation between the parameters and the moments.

(Refer Slide Time: 06:28)



What you are essentially doing is there is a moment of the pdf I th moment or N th moment and you are striking a relationship between these two quantities. Sometimes is g of theta may be theta itself, as a simple example; suppose I am looking at estimation of mean I am given some series a standard problem that we have been discussing. The relevant moment would be the first moment x times f of x; dx; suppose I am or let me use y here y times f of y; dy is the first moment we know that is nothing, but mu.

Since the mu is unknown, we would generally prefer to write mu on the left hand side. Now this is a theoretical relation between the first moment and the parameter of interest alright, now this parameter of interest is mu; in some pdf the parameter of interest may be something else which need not be mu, so you will have to rewrite this in terms of mu. Suppose I am estimating mean of a uniform distribution, so what are the parameters of interest in a and b; the intervals, so I would have to rewrite this equation here in terms of a and b. So, I would say a plus b by 2 is integral y; f of y; dy; obviously, this single equation alone is not enough to get me the estimates of a and b. I would need to invoke a second equation and what would be the second equation. Should be no hesitation if you have an answer, what would be the second equation in terms of moments.

Student: (Refer Time: 08:25)

Second moment right; you can pick any moment, it does not have to be the second moment, but the natural choice would be the second moment and what would be on the left hand side.

Student: (Refer Time: 08:41)

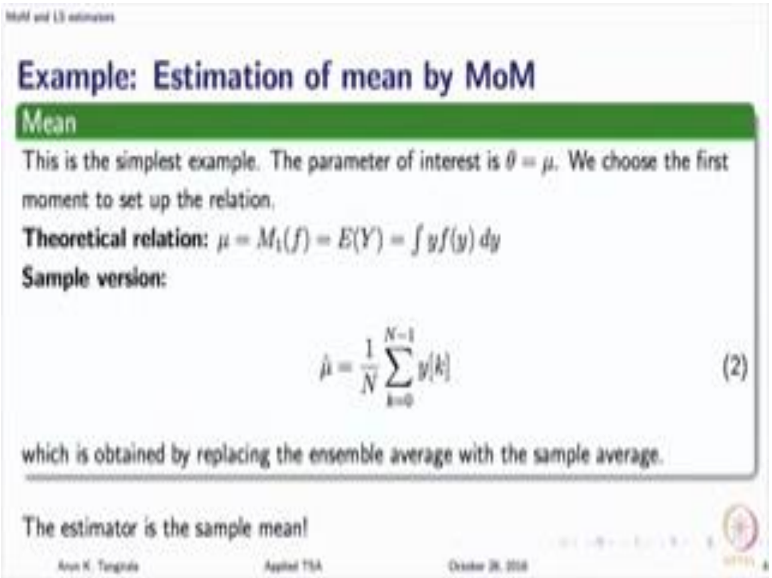
B square minus by.

Student: (Refer Time: 08:50)

Twelve, but we are not yet ready to use this equation until we replace them; these right hand side with the sample versions. What do you mean by sample versions, you would replace this for example, on the right hand side you would replace this with the sample mean that is rather than saying sample mean, you would replace them for example, with this, the first equation that is the sample average. So, you are going to replace ensemble averages with sample averages very simple. This is an ensemble average, across the possibilities and here you have the sample time average. So, replace ensemble average with the time averages that is the basic idea. Step one write the theoretical relations, step 2 replace the theoretical expect averages or ensemble averages with the time averages.

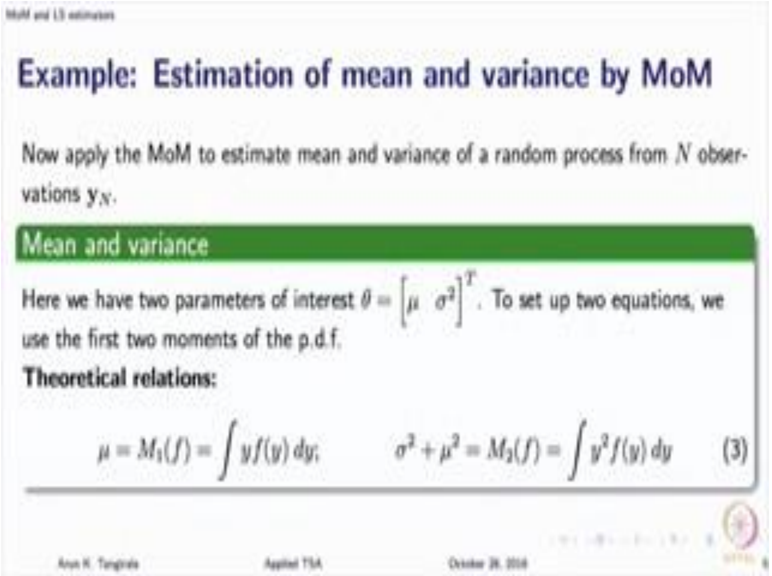
So when you do that, you will get two equations and you can solve for a and b; even it comes to Gaussian parameters straight away you get the estimate to be 1 over N sigma y k which is nothing, but y bar. So, first identify the parameters of interest write the theoretical relations, if you are estimating parameters of a Gaussian pdf; the parameters are mu and sigma and then you write the theoretical relations; these equations, these moment expressions do not change with the pdf there is just integral y f of y dy and of course, what I have not mentioned here is that you are looking at a joint pdf, but I just suppress that notation. So, straight away you get mu whereas, here you have to solve them simultaneously to get your a and b, any questions on this approach very straight forward.

(Refer Slide Time: 11:03)



So, this is the example that we just discussed estimation of mean by method of moments.

(Refer Slide Time: 11:11)



And if you want to estimate mean and variance for example, of a Gaussian pdf then you would require to write two equations; the first equation both equations we have written on the board, the only difference between what we have written on the board and this example is the parameter are different. The parameters of interest are mu and sigma whereas, for the uniform pdf, the parameters are a and b; any questions?

It is a very straight forward idea; there is nothing complicated here; however, the problem is with the uniqueness, I can use any moment equation there is no restriction. I just need as many relations as a number of parameters, so I could write in this case we have used first and second moment, I can relate mu and sigma through may be fourth moment of a Gaussian pdf; if I am looking into Gaussian pdf, I can express the fourth moment in terms of mu and sigma there is a relation, I could use that relation instead of the second moment. Do you think that will give you different estimate yes or no. So, in general the problem with method of moments is lack of uniqueness and depending on which moments you have used the estimates can change and as a result of which their properties can also change the bias, variance and so on.

So, for example, here if you are looking at Gaussian pdf already this mu is going to be estimated by sample mean. The sample mean is a beautiful thing, we have already shown sample mean is an estimated the least square sense in MLE sense, now also in the method of moment sense. The only problem with this method of moments of course, is uniqueness, but the other problem is I do not know in what sense it is optimal right at least with least squares and m l e; I know that I am solving an optimization problem explicitly whereas, here with the method of moments, I am not solving an optimization problem explicitly I am not trying to minimize some sum square errors or maximize likelihood or even maximizing efficiency nothing I am just putting up a bunch of equations and solving and the answer turns out to be sample mean and if you are also looking at estimation of variance.

(Refer Slide Time: 13:36)

What is the expression that you would get for; so mu hat would be this and sigma square we have written the theoretical relation. If you have to use that theoretical relation that you see on the screen, what would be the estimate of sigma hat sigma square; sorry.

Student: (Refer Time: 13:58)

Expect (Refer Time: 14:00)

Student: summation of (Refer Time: 14:04)

Y minus y bar.

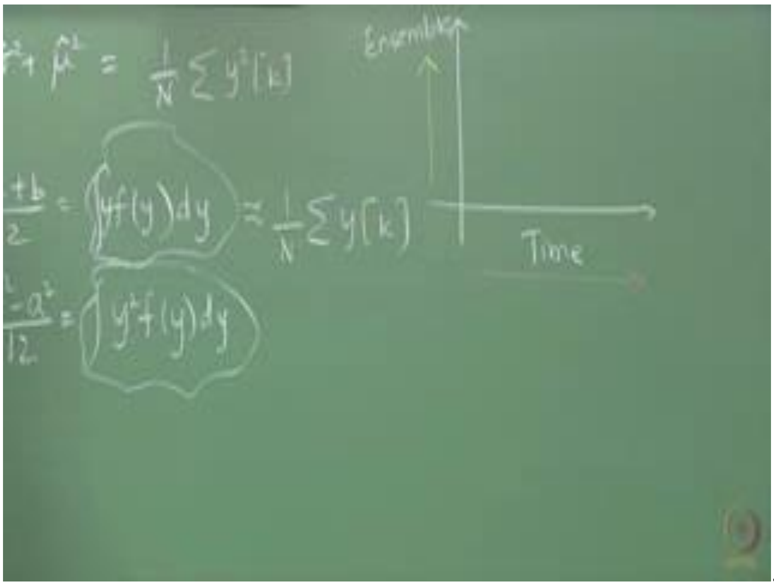
Student: (Refer Time: 14:08)

Ah by what factor that is all.

Student: by N minus (Refer Time: 14:10)

N minus 1; do you want to work it out, you have to solve two equations; one equation is already telling you what is mu hat; you have to plug in the value of the expression for mu hat into the second equation. So, the equation that you have to solve is sigma square plus mu square well now you are going to replace with the estimates because you are going to use time averages on the right hand square and here right what is the time average of integral y square f of y dy; 1 over N sigma y square. What we mean by time average is whatever I am averaging in the ensemble, what I am averaging, y square remember f of y integral y square f of y dy what is it; it is an average of y square, but in the outcomes space that is what we call as ensemble average.

(Refer Slide Time: 15:13)



.

So, instead of looking at this ensemble direction, we are looking at the sample or the time direction. So, here is your ensemble or realization; the theoretical relation looks at this direction here whereas, what we use in practice is the time direction. There are green chalk I do not know how green would look on a green board; I will try it out somewhere. So, what we are looking at here is in the time direction because this is all we have alright. So, I have here this equation; I have already have mu hat with me; so what is expression for sigma square hat.

Student: (Refer Time: 16:11)

One mine

Student: (Refer Time: 16:14)

Put together, you should expect to see something of the form sigma y k minus y bar square what about the factor in front of the summation.

Student: (Refer Time: 16:35)

1 by N; 1 by N minus 1?; You will get one by N minus 1?; N minus one where you get from?

Student: (Refer Time: 16:42)

You get 1 by N, so the factor is 1 over N; what kind of an estimator is this in what sense is it optimal.

Student: (Refer Time: 16:59)

Where did we get this kind of an estimate?

Student: (Refer Time: 17:02)

MLE from the maximum likelihood approach we got this estimate correct. So, the method of moments seems to be giving an MLE type of estimate which means we already know this is a bias estimates. So, which also tells us method of moments can give you bias estimates, but we know it is asymptotically unbiased as well.

So, what we learned is that the method of moments forces the time averages to satisfy the theoretical averages which itself may seem in error, but by some magic it is giving you optimal estimates right. Although we are not solving an optimization problem exactly, we are ending up with optimal estimates. In fact, it is true in general you can show that the method of moments is optimal in some sense; although you are not solving, you are not formulating an optimization problem, but the problem with method of moments is lack of uniqueness. I may use a bunch of relations and some other person can use another bunch of relations. So, there is this choice that prevails and one has to figure out what choice of moments you have to use of course, you know people have studied then there are some standard approaches.

When it comes to applying this method to estimating parameters of time series models., slowly we are talking about that; we have already seen that where have we seen the method of moments being applied in estimating time series models sorry.

Student: (Refer Time: 18:21)

Yes, the Yule Walker equation; you first derive the relations for the between the autocovariance of the process and the model parameters and we have already spoken about it those relations can be used for two different purposes; one is given the model parameters I can compute the theoretical auto-covariances which is what ARMA ACF does and the other way around given the ACFs, I can calculate the model parameters. How do I do that? I use the method of moments idea; I assume that those theoretical relations that we have derive are also satisfied by the estimates of ACF.

So, essentially I am going to replace the theoretical averages, ACFs are also theoretical averages only; any theoretical property is an average it is an average across the ensemble because there is an expectation those theoretical averages are replaced by the estimates and that is it you get your Yule Walker equations. So, the application of method of moments to time series models is very straight forward, but the question is how good those estimates are. We have also seen for example, how to use this method of moments idea to estimate parameters of an MA 1 model and we found there are two solutions and so on.

Now, it turns out that when you apply the method of moments to estimating autoregressive model parameters, you get very good estimates; very good in the sense you get sufficient estimates; you get the most if you get the same efficiency as you get least squares and MLE alright, but when you use the same idea to estimating parameters of an MA model and will I will talk about it, I will list all of this points later on, but I am just giving you some preview when I use an MA method for estimating parameters of a moving average model then I run into efficiency issues, I get estimates with high error.

So, normally what is done is this Yule Walkers method, when it comes to estimating moving average model parameters, normally we use MLE, but MLE results in a non-linear optimization problem, after all you going to maximize log likelihood and that is going to result in a non-linear optimization problem. Since non-linear optimization problems require initial guesses, we use the Yule Walkers method to kick start the MLE algorithm, there are other ways to kick start MLE algorithm; this is one of the ways alright.

So, you have to remember that not every estimation method is suited for all kinds of estimation problems. MLE is has a more universal appeal to it, but method of moments does not; now having said that somewhere in 1980s; this method of moments has a long history, but in somewhere in 1980s an economist by name Hanson; he proposed what are known as generalize method of moments. In fact, he went on to actually become the Nobel prize winner for his pioneering work in econometrics and lot of methods that we learn in time series analysis come from econometrics interestingly and rightfully so because in econometrics you have all kinds of complexities that you cannot see in engineering processes and so on.

May be biological processes are the ones that can beat the econometrics in terms of complexities because econometrics involves first of all so many factors that cannot be measured and that are too complicated particularly because there is a human factor involved. A lot of the effects that you see or the observations or variables that you seeing in econometrics are driven by human decisions and humans are very weird people; they are very difficult to model; they are particularly the psychology is very difficult to model.

So, you will see all kinds of complexities and then you combine the psychological factors of many human beings, it becomes even more complex. That is why whenever you have a complex process that you do not know how to model, you turn to econometrics and there will be a solution, whether it is heteroscedasticity or whether it is fractional integrating effect whatever it is any kind or co integration and so on, you do not hear this kind of fractional integration or co integration and so on in engineering systems, but you will see that very often in econometrics.

So; obviously, people have spent a lot of time in econometrics and this particular gentleman came up with this method of generalize method of moments. The basic idea in generalized method of moments is two (Refer Time: 00:00) one; he said method in method of moments you are writing exact number of relations as a number of parameters and then you are forcing the time averages to satisfy those relation; obviously, we know there is going to be an error; I mean even if I look at this what I am saying here is the mu is going to be satisfying this and then saying that I am going to replace this on ensemble average with time average and then call that as an estimate; obviously, there is an error huge error. So, he says when there if error, it does not make sense to have exactly that many relations as a number of parameters, write more relations; when you write more relations, you have a set of over determined equations and generally how do you solve over determine equations using a least square approach right.

So, that is the key change that generalize method of moments brings with it and that change is just amazing. You write more number of relations because you have now error in every equation, what basically he pointed out was that this relation cannot be satisfied exactly when you know that certain relations cannot be satisfied exactly; have more number of relations written up and then solve a bunch of over determined equations, typically using least squares or your favorite method whatever method you have for solving over determined equations.

It turns out that it has much better properties when it comes to estimation. In fact, he went on to show that it has asymptotically the same property as MLE. So, you generalize method of moments which we are not going to discuss, but I gave you a brief idea; some write up is given in my text book, but you can go to the literature and read up more on generalize method of moments. There are packages that do that for you, but it is a beautiful idea, the only problem there again is how many relations should I write and you know how does the estimate improve with the number of relations and so on. Obviously, there has been lot of study done on that so turn to the literature to see what answers you have there.

So any question on the method of moments, you should remember method of moments does not set up an explicit optimization problem. The vanilla method sets up as many as the number of parameters and then the choice of these relation equations that you are writing are not unique, but at least for some basic problems, you can show that the result in optimal estimates. Sometimes I can give you efficient estimates depending on which parameters you are estimating, what estimation problem you are solving. So, then we move on.