**Applied Time-Series Analysis**
**Prof. Arun K. Tangirala**
**Department of Chemical Engineering**
**Indian Institute of Technology, Madras**

**Lecture – 81**
**Lecture 36A - Introduction to Estimation Theory 4**

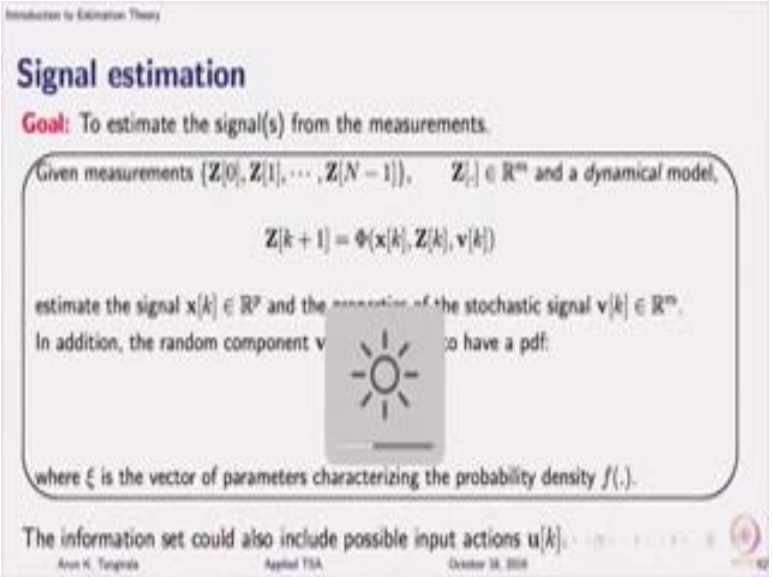
What we will do today is something quite exciting, we will learn what is known as Fishers information and then talk about the properties of estimate as bias and variance and so on. But before we do that it may be a good idea to get a preview of what are the different types of estimation problems that one encounters in the literature given the vast body of work that has a come about in the last 200 to 300 years at least in the open literature does not mean that people are not through about estimation.

Generally whatever is done is attributed first to Greeks. And I have a problem with that I think there are there was civilization before Greeks also and they have also contributed. But we will not go that past; we will talk about the literature that is available in the last 200 to 300 years. And as I said this is one of the oldest problems that have been studied. And people have found that there are equivalences between these different types of estimation problems. Nevertheless it does merit to talk about this difference classes of estimation problems that have emanated from different fields.

And yesterdays example gave us some flavour of what is a signal estimation problem, what is a state estimation and parameter estimation; we will just briefly expand on that and move forward, we are not going to really go in to full technical details; I just want to give you a feel of these different types of estimation problems.

So, if you look at the signal estimation problem the statement is as follows.

(Refer Slide Time: 01:52)




The goal is to estimate signal from measurement; that is a classic a objective that does not change over a period of time. But always you have to ask in estimation what is given to me? Remember we talked about that yesterday. So, here we have given measurements and this z could be sorry, z could be a vector or could be a scalar at each instant.

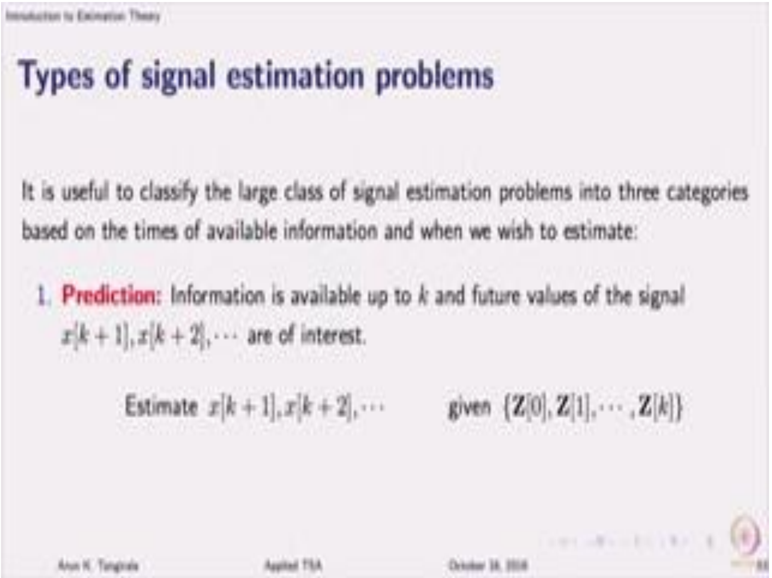
For example, I may have only measurements of the output of a system or both output and input or some other variables and so on. I am given measurements of all the variables that I could measure in the process over a period of time. And I am also given that system is dynamically changing, and there is a model for this dynamical process so that is also given to me. Remember we need data, we need model, and we need an objective function. The objective function is not specified here. The data, what kind of data is given to me? And what kind of model is available is given to me? This model includes both the model for the process that is the process that is generating the response and also the uncertainty description of whatever uncertainties are present in your process.

What we mean by uncertainty is measurement errors for example; your v k could include measurement errors and or stochastic disturbances and may be process noise and so on. So, the complete model is given to me and the measurements are given and I am supposed to estimate the signal from the measurements. And the signal here is x which have do not know. When we go to state estimation we call this underlying signal as a state. One of the early problems that was studied in signal estimation was the classic steady state problem; that is the like the problem that we have discussed yesterday. There is a signal that is at steady state.

Yesterday of problem was quite simple it was a constant signal, but there could be another signal underneath. For example, a sine wave and the sine embedded in noise again that is a very standard problem that has been studied widely. So, these were the problems that were studied early on and then gradually it was expanded to the dynamic case as well.

So, this is the estimation problem.

(Refer Slide Time: 04:29)



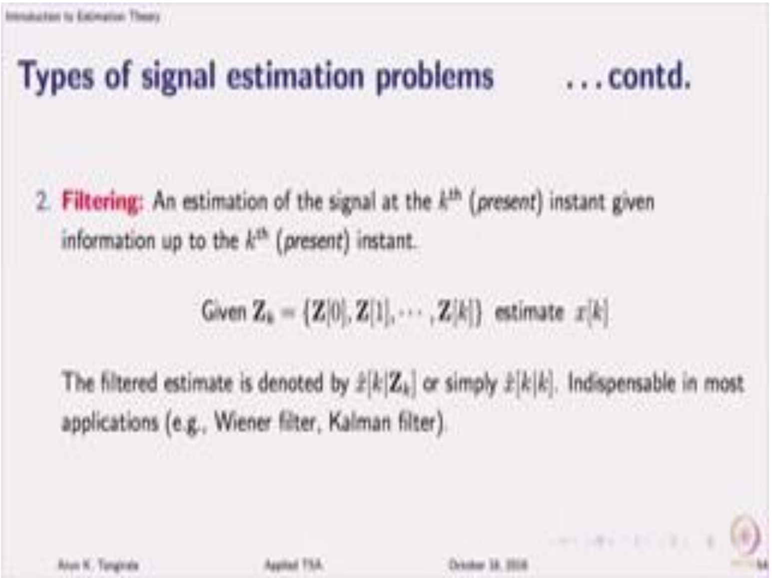

And of course there are (Refer Time: 04:22) of methods that you can think of depending on the objective function that you work with. And within the signal estimation problems the reason why the signal estimation problem is requires it is own attention is because of the three different sub problems that you would run into in any signal estimation problem. And those three sub problems are: prediction, filtering, and smoothing. And the one that distinguishes prediction from filtering and smoothing is essentially what is given to us that is where in time up to what time I am given the information and at what time I want to estimate the signal that is what distinguishes one from the other.

For example if you take prediction we all know what the prediction problem is. We are given data up to this point; let us I am given up to the k-th instant. And then I am interested; I am given measurements I am mot given signals remember that. So, prediction problem is that of predicting the signal and sometimes even the measurement at obviously in feature k plus 1 k plus 2 and so on. If you are looking at just predicting a k plus 1 it is a one step ahead prediction problem. And if you are looking at predicting k plus 2 given information up to k-then it is a two step I had prediction and so on.

Typically, what we do in a game of chess right, we know the moves until this point and we are trying to predict that the opponent will make in future; it could be one next move or next two moves and so on.

So, this is your prediction problem and this is what we have been talking about lot of times.

(Refer Slide Time: 06:09)




The other problem is filtering. And here the situation is that I am given information up to k. So, the nature of given information does not change, but what makes it different from prediction is the time at which I am interested in the signal. Here, I am not interested in the future at all, I am given measurements up to k-th instant and I want to know the signal at the k-th instant. So you may wonder what is the difference. what is the difference between this problem and the prediction problem?

First thing is we are not looking at future. If am given measurement at k what is that that I am seeking at k; the signal, the signal is a hidden 1, what I have is the measurement.

What is the difference between the measurement and the signal? The noise right, something has corrupted the signal and I have the signal in a corrupted form and I am suppose to go past that look through that and the estimate the signal.

So, what we are doing is you are putting your measurements in a filter and this filter essentially is giving you out the estimate of the signal at this instant. And to produce the estimate at this instant what you are doing is you are using the information from the past up to this instant.

(Refer Slide Time: 07:37)

So, you can now draw a timeline here let say this is your k-th instant here is k minus 1 k plus 1 and so on. Prediction is makes use of the information given until this point; this is the data that is given to me. If you look at the information that is given to us when it comes to prediction or filtering both share the same information; they rely on the same information. It is only what is of interest to me makes the two problems different. In prediction and here I am interested in this horizon here. So, I am looking at future, whereas in filtering and interested only at this instant. That is the instant up to which I am given the data.

And what happens for a pure stochastic signal? I mean do these problems actually apply to pure stochastic cases as well. Not really, when you have a stochastic signal that is just purely stochastic signal then there is no filtering problem per say. Filtering problems come about when you have a scenario like this when the stochastic signal that is given which is y let us say is made up of x and v and I want to estimate x. In a purely stochastic signal case we are only interested in prediction, and that is why we have been always talking about prediction.

Now, when it comes to smoothing the scenario is different. Both in terms of the information that is given to me and the information that I am interested in that is the signal horizon; that I am interested in. So, when it comes to smoothing I am given data up to let us say some point here or let me put it this. So, the information that I am given is up to some point with respect to k. So, reference point is always k with respect to k I am given information the past and in the future.

So, this is the data for smoothing, and I am still interested only in estimated k. So, it is more or less like filtering, but what I am given is the data to the left and to the right. I think I gave this example of when we go to a movie and we have missed out hearing on something. So, there are three different kinds of problems that you see beings solved in a theatre: one the audience trying to predict the dialogue that is going to come up next and the it depends on whether you are watching a Bollywood movie or a Hollywood movie. If it is a Bollywood movie that is quite predictable and you can predict the scripts based on what you have seen until now that is your prediction problem. Filtering is somebody was shouting really, when you are wanted to hear some dialogue and you are trying to guess what words must have been spoken on the screen based on what you have heard until now. Smoothing is more of a correction to the filtering. Smoothing is you guess something at k and you waited for the movie to role to k plus 1 k plus 2 k plus 3 and then your mind is still position that k and you are trying to correct based on what you have heard later on. So that is the difference essentially between prediction, filtering and smoothing.

Of the three smoothing is a non causal operation, because you are positioned at k that is the point of interest is k and you are relying on future information. Of course, in addition to the past, whereas prediction and filtering are casual operations in fact, you can say prediction is a strictly causal operation, whereas filtering is causal operation; you cannot say it is strictly causal because you need information up to k. All these three different problems you will encounter in almost all signal analysis problems, wherever you are interested in predictive analytics as they call today. In all predictive analytic problems you will run in to this.

So, if you are familiar with Kalman filter; the celebrated Kalman filter, there is a Kalman there is a Kalman filter and smoother. So, there is a filtering and smoothing embedment sorry, in fact there is a prediction, then there is a filtering, then there is a smoothing. All the three are kind of embedded in the Kalman filtering algorithm, at least definitely the prediction and filtering. Smoothing depends on whether you are interested in going back in time and correcting, what you had filter.

But certainly there is prediction part to Kalman filter and then there is a filtering part. The generic term that is being use for all these three is estimation. So, estimation is like the grand name given to all of this, but you should understand that there are these three sub problems. We in this course are primarily interested in prediction. Remember that there are three different kinds of operations and they have their own set of applications.

Parameter estimation is one of the central problems that we are interested in. As far as signal estimation is concerned we are only interested or primarily interested in prediction, whereas parameter estimation is perhaps the most important problem that we will be talking about. And when we talk of parameters typically in literature you will find two different sub classes: one is the parameters of models like your ARMA models, AR models, or any model that you are fitting does not have to be time series models. So, that is a model parameter estimation problem. But the older problem that has been studied in estimation literature is that of parameters of periods.

So, early on people were only interested in estimating mean, variance or some parameter of the pdf and so on there was no time series analysis before 1930 or 1920s. So, there you would talk about estimating the movements of the pdf and that is a classic problem that has been studied extensively. We will start with that problem and then move on to model parameter estimation. And when it comes to state estimation; state estimation is not so different from signal estimation. The only difference is that now you specifically attach a name to the signal that you have not observed and you call it as a state.

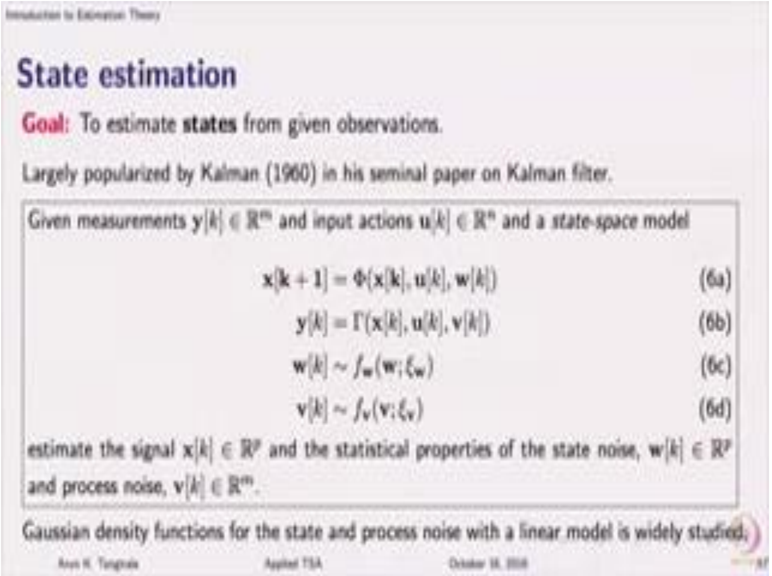
Of course, it makes a big difference because the moment you talk of state estimation you can bring in several problems in to it is gamut. In fact, you can show that the signal estimation problem can be classes as the state estimation problem, because after all signal is an unobserved variable and state is also an unobserved variable. So, you might as well call the signal as the state. You can also cast parameter estimation problems as state estimation problems. Why, because parameter can also be thought of as a hidden quantity.

But there is a small catch there which is that in a state estimation problem typically the state is assumed to be dynamically changing with time. Whereas, in the parameter estimation problem the parameters may change with time, like in your time invariant the descriptions or may not change with time or may change with time; that depends on the process. If they do not change with time then you are looking at invariant time invariant processes. Then you may wonder how you can bring them in to the state estimation frame work well you think of that parameter as a state which is at steady state. We have been used to hearing the steady state term for a long time now you understand what is meaning of steady state. There is a state that is describing the process that has reached the steady value.

So, you can think of the parameter as some kind of state of a process that has not been observed, and if you are looking at time invariant systems then you think of the parameter as a steady state. Whereas if you are looking at time varying systems where the parameter changes with time then you can think of it as a state in a classical sense, and you can give it a model; that means, you can then assume that the parameter changes in a particular fashion and so on. And that becomes a state equation.

So, the x here is a state as against what it was before in the signal estimation it was the signal.

(Refer Slide Time: 16:45)




The notation of course is different in between the parameter estimation and state estimation. In parameters are specifically denoted as theta. Now, in practice you may have a combination of state and parameter estimation problems; that means you may be simultaneously estimating states and the parameters of the model. You may want to observe the distinction. So, that problem is called state space identification which we teach in system identification. That problem has been only recently solved then compare to this state estimation problem. The state estimation problem was solved by Kalman initially and then of course a lot of works followed. Somewhere in 1960s Kalman's seminal paper came about and he showed how to estimate states given the model. Then somewhere in 1990s people started looking at both state and model identification process.

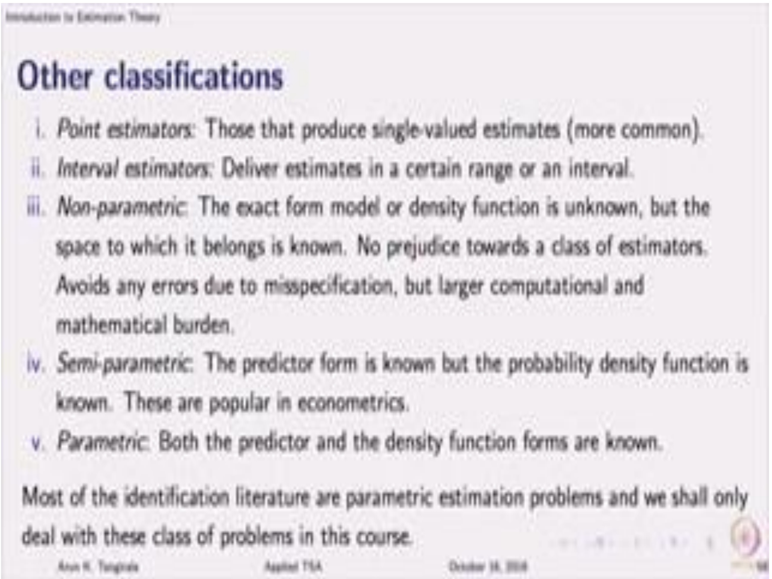
So, model is also not given to me, state is also not given to me, what is given to me the data. So, that is a lot of unknowns that one has to estimate. And the problem is therefore more complicated than a simple state estimation problem. That is essentially your state space identification. And generally people dread to read the topic at least beginners in system identification. But, you know over the years things have become simpler to understand and so on. Still people are working on state space identification. So, these are the different types of estimation problems and you should remember that all though there are differences in the formulations, in the statement, problem statements they are all equivalent at some level. And this equivalence is exploited extensively in literature, when you cannot solve a problem in it is classical form in it is standard form you cast it into another form where solutions are available.

Remember, the parameter estimation problem as I said originally was of interest to this statisticians and even today, because they were interest in estimating the parameters of pdf's right of the distribution functions. Therefore, lot of contributions have come from statisticians, whereas you look at state estimation contributions have come engineers predominantly. And if you look at signal estimation it is a mix of both even signal estimation has come contributions have come a lot from engineers by enlarge.

But at some point you do not make any distinction between engineers and statisticians you say they are all in the same framework. They are equally responsible for making our subjects very complicated and as a result assignments increases and so on. So, we hold all of them equally responsibly.

Fine, so any questions on this; these are the three different problems.

(Refer Slide Time: 19:43)



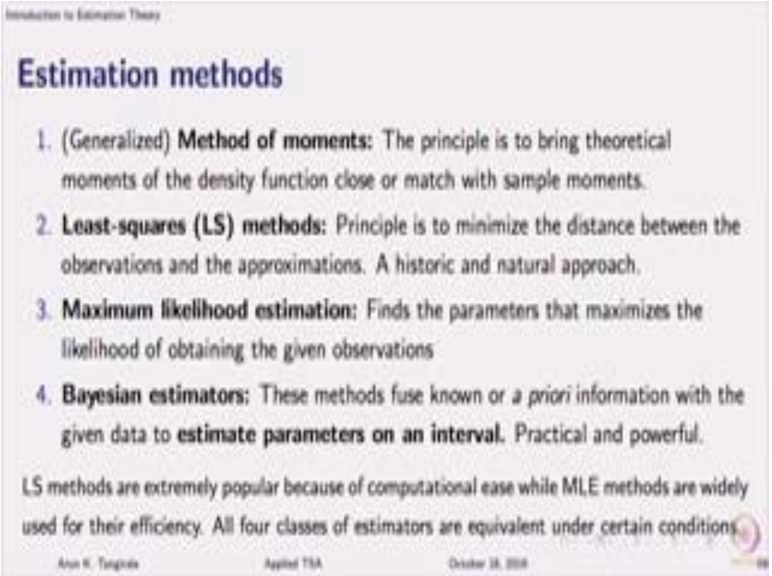

There are the other classifications when it comes to estimators. We have talked of estimation problems, sometimes estimation problems are classified based on what kind of estimator you are looking at. Until now we have talked about the kind of estimation problem that we are solving, but there can be classifications based on what estimators what kind of estimators you are interested in: non-linear, linear, point estimators, interval estimators and parametric estimators, non parametric estimators, and so on. I do not want to go in to the details, but the top two which is the point estimators and interval estimators you should be aware of the distinction between them.

Point estimators are those which will give you a single value for the parameter, like the sample mean it just gives you a single value. Now that is an estimate, ultimately as I said yesterday we would be interested in constructing intervals for the truth. After all what is the purpose of estimation? Guessing the truth, coming up with a good systematic optimal guess of the truth; that is what essentially is purpose of estimation. So, if I give you the single value, you and I know very well that that single value is only consequence of this data record had I given you different data record you would have obtain the different value.

So, you do not want to place full faith in that single value which is what we call as a point estimate, rather than that you would say give me an interval I know very well that I cannot estimate the truth pin point I cannot pin point the truth give me an interval and there are estimators which give you interval estimates, they do not give you a single value these interval estimators are still constructed from a signal record only they are not constructed from multiple records.

So, the approaches to construct interval estimates are direct and indirect. What we mean by direct and indirect? Is if you look at Bayesian estimators they give you straight away interval estimates, whereas you look at the classical least squares MLE, method of moments and so on they give you point estimates first from where you will construct the interval; that is an indirect kind of approach. But the underlying philosophies are worlds apart, and we will talk about that when we talk of different methods of estimation.

(Refer Slide Time: 22:15)




And those four different methods of estimation that we will learn are the method of moments which I have been talking about Yule Walker’s method is one such method. For today we have what is known as generalized method of moments. We will discuss least squares methods which we are familiar with. Then we will talk of maximum likelihood to which we will give introduction today. And then we will end with Bayesian estimator.

So, if you look at ultimately when we reach Bayesian estimate estimators we will realize that the Bayesian estimator is like the great grandfather of all of this, because it encompasses maximum likelihood methods which in turn encompass least squares methods and least squares methods can by shown to encompass method of moments. So, you can say it is like the super dada of all of this. But people would disagree, because the philosophy mathematically you can show that one is contained in the other, but the philosophy that is underlying the first three methods is completely different from the philosophy that underscores Bayesian estimations. And it is completely different.

So therefore, as long as Fisher was alive he argued vehemently; Fisher was the proponent of maximum likelihood and he all constantly argued that all though maximum likelihood estimation can be shown to be a special case of Bayesian he just maintained that no way they can be same because the philosophies are different. But then you see everybody has their time and people just waited for him to pass away. And then now he is not around to argue, there are others you are arguing but still mathematically Bayesian estimation is the equal; MLE is the equivalent to Bayesian estimation you can show.

Of course, you should also keep in mind the fact that the philosophies are different. Good, so I am not going to summarize anything now. We have summarized too many times.