Applied Time-Series Analysis

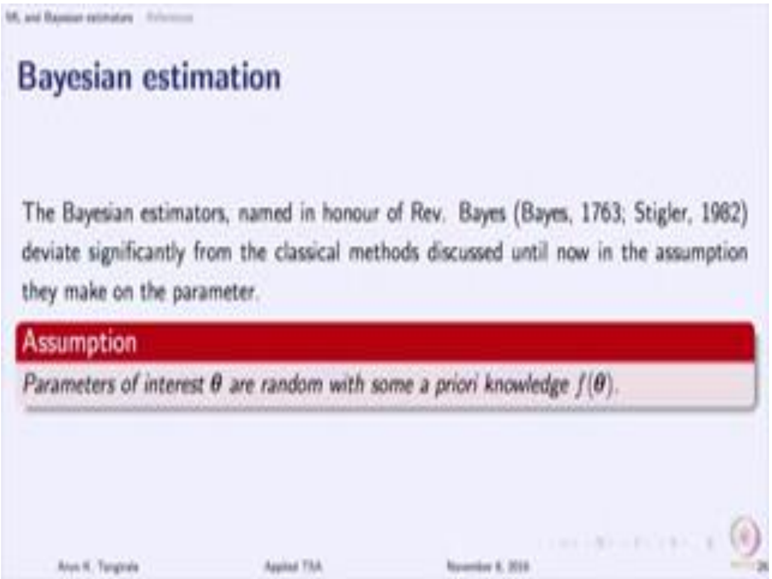
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture - 107
Lecture 46B - MLE and Bayesian Estimation 2

What I will quickly do is just spend about 5 to 7 minutes and then we will adjourn just on Bayesian estimation. I wanted to introduce to you, the idea is extremely straight forward and always straight forward ideas will have very complicated implementations; generally speaking.

I remember my brother used to say if the question is a 1 liner then you are doomed; that means you have to really supply the information and write a long answer if the question is very long then you are happy because there is a lot of information. So, likewise here the idea in Bayesian is extremely straight forward, there is only a single equation and that equation stands from Bayes' rule, but the implementation and the philosophy is not simple.

(Refer Slide Time: 01:00)



The Bayesian estimators depend on a very very important and different philosophy that is that is in fact, different from the only 3 estimators that we have discussed method of moments least squares and MLE.

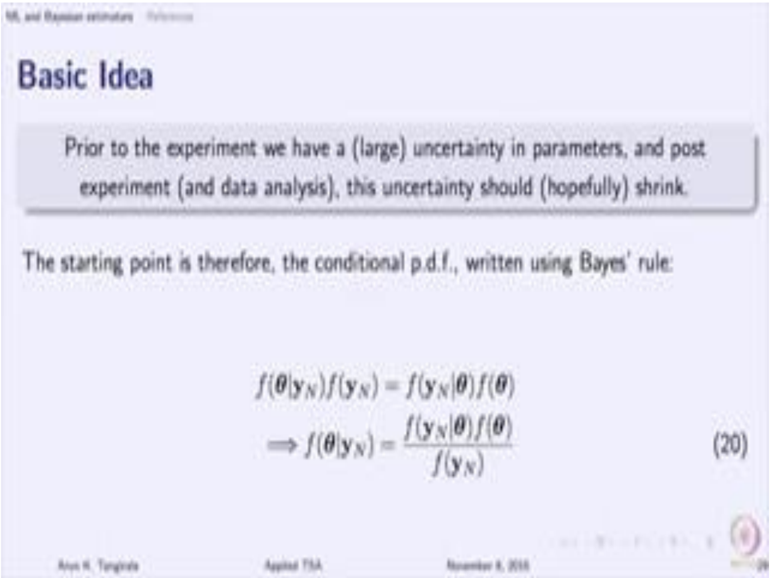
The assumptions here in the Bayesian philosophy which many people do not like or subscribe to is the parameters that you are estimating are random, they are not deterministic. So, imagine if I were to say the estimate; the mean of a Gaussian white noise process and I tell you the mean is also a random variable, you may not like it so much. How can the mean be random variable? It is a statistical property, I have been thinking all along that it is a fixed quantity and now suddenly somebody comes along and say it is random. So, that kind of puts off many people when it comes to Bayesian, but you just have to relax, you just have to hold on, do not get so hesitated about this assumption, what it says essentially is it is not necessarily that the true parameter is random, it is; it appears to be random because of our uncertainty that is how you look at Bayesian and then you are at ease with this philosophy otherwise you will constantly be hesitated and be engaged in endless debates.

All you have to tell yourself is truth may not be random, but the mask that I am wearing which I am unable to throw off makes the truth uncertain; that means, I always have some uncertainty shrouding the truth. Bayesian philosophy says why do not you consider the truth along with the uncertainties that you have put together and that is why it says you begin with an uncertainty about the truth, what we mean by begin is before data; b d, before data and after data, before data you have uncertainty in the truth your knowledge and after data also you will remain certain. Your data is only going to help you shrink the uncertainty, but it will never take it to 0 because you can never estimate the truth with certainty that is the basic idea and that is why there is this concept of prior and posterior in Bayesian.

Prior is before you perform your experiment so suppose I were to ask you, what is average temperature in this room? You may give me some value, but; obviously, you are not going to be certain about it, you say there is some uncertainty around it, then you perform an experiment, if there is a sensor somewhere, you would or you bring a temperature sensor, measure, collect data and now the purpose of collecting data, what is the purpose? To get more precise estimate, to come up with the more precise estimate, what Bayesian philosophy does is very beautifully, very nicely allows you to fuse 2 different pieces of information - one information is your own prior knowledge, prior guess, whatever you may want to call and the other piece of information is data. So, you have prior information, you have data, you have to really marry them very well so that what comes out is a more precise estimate of the average.

Until now, we have not talked about prior at all, whether it is maximum likelihood or least squares or method of moments, we have pretended, I do not know anything, I have no idea, what the average temperature is that is the case, really in many applications, I will have some idea, you really do not know anything, it is not true, you will have some idea about the estimates sometimes you may not, but which is in Bayesian method that is also possible. If you say that I do not know much about this thing, about I do not have a good knowledge, a initial guess of the theta that is also you can actually incorporate that information.

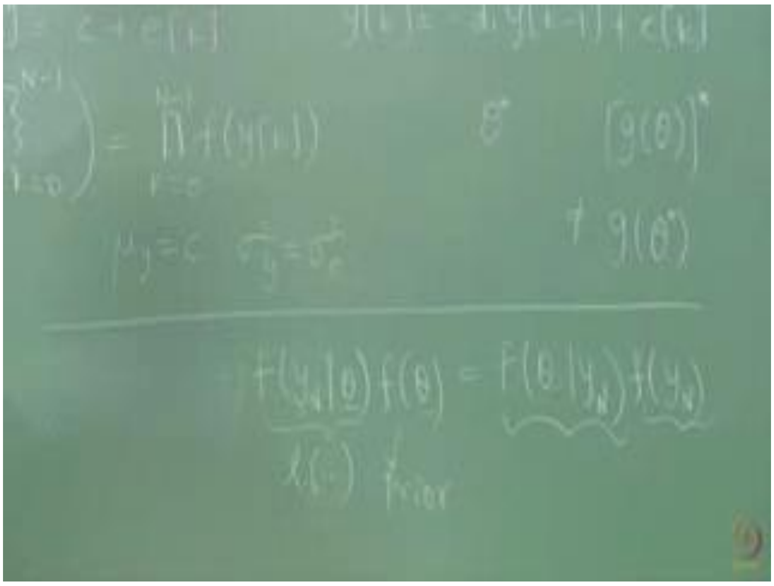
(Refer Slide Time: 05:03)



Prior to the experiment, we have a large uncertainty typically in parameters and post experiments and through proper data analysis, hopefully that uncertainty had shrunked and you should conduct this also tells you, what kind of experiments you want to perform, you should conduct experiments that will reduce the uncertainty in your priority; that means, do not get some data which is no information, we have already talked about Fisher's information. So, everything comes about very nicely. So, the only equation that governs the Bayesian estimation philosophy is this single equation; the single equation stands from Bayes' rule, now that we have assumed theta to be random we already know measurements are random therefore, we can say the joint p d f of the observations and theta, now theta is also random. So, I have to now consider the p d f, the joint p d f of the observations and the parameters.

This joint pdf we know from Bayes rule f of x comma y, what is it? It is f of x given y times f of y of course, marginal.

(Refer Slide Time: 06:24)



I can write this as f of y given theta times f of theta, it is understood that this is marginal, this is 1 way of writing it, am I right? What is the other way of writing this joint p d f, f of theta given y? So, this is also equal to f of theta given y times f of y, again this is the marginal. So, this is, these identities are the once that are being written, this identity is on that is that has been written there for you.

Now, we can throw away this, I am what am I interested in, what do I know? Right, first of all what is f of theta? f of theta is the p d f of your parameters prior is that it is the prior, before you begin your experiments; that means, no data is given to you absolutely no data is regardless of the experiment that you are going to perform you have some knowledge of theta that is your f of theta, it is an there is knowledge with some uncertainty say therefore, this is a prior. What is f of y that is the.

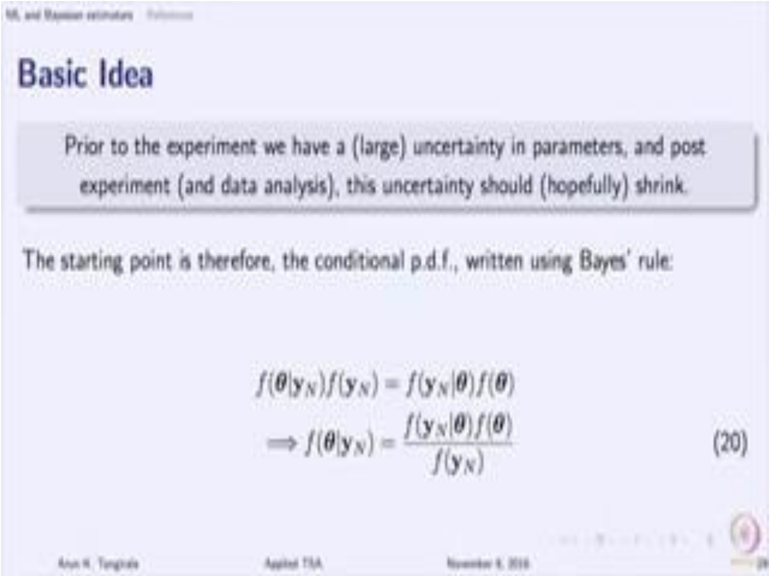
Student: Joint pdf.

Joint p d f of your observations, this is coming from your experiment now that in fact, this is just f of y independent of theta, you can see that because it is marginal regardless of the theta that is f of y, what is this here? You have seen this before that is a likelihood that is the measurement that you are assuming, you are now parameterizing that is now the difference between f of y and f of y given theta is this f of y is the marginal, it is it has got nothing to do with theta it is some fixed value as for as the moment I freeze y, the moment I freeze the data f of y n is fixed.

Whatever maybe that; whatever maybe the theta f of y is fixed whereas, this here is a p d f of y conditioned on theta it depends on value of theta that you assume it is nothing but your observation p d f that you have assumed that is coming from the user. So, you have to supply this, we have we are anyway doing this in MLE for, in the previous example this was a joint Gaussian. So, this is your likelihood. In fact, you can say l of theta given by, but I am just going to write l. So, this is likelihood then you have prior and then this is what you are looking for.

When before you begin the experiment you have f of theta after you have conducted that experiment you want to ask, what is the uncertainty in theta given this data? So, look at this carefully before you begin the experiment, you have some guess of theta like you say the average temperature here is 20 degree Celsius with some let us say Gaussian distribution, the Gaussian kind of uncertainty, whatever uncertainty it may have, now you bring it a sensor, you get the readings and you say now given this readings, what is your estimate? What is the uncertainty of your estimate that is f of theta given y which is what we want to calculate, this is therefore, the posterior after the experiment has been conducted, this is before the experiment and that is why you have the equation f of theta given y is f of y given theta v times f of theta divided by f of y. But we have already said that denominator is independent of theta in that equation that you see on the screen in equation twenty there the denominator is independent of theta. So, I can assume it to be constant f is a scalar f of y n is a scalar value.

(Refer Slide Time: 10:23)



It is some constant that is independent of theta, now what Bayesian estimation does is it says this is your posterior; it gives you everything whatever you want wanted about theta, after the experiment, what do you mean by everything? What we mean by everything is whatever we wanted out of an estimation exercise, what did we do with MLE or least squares and so on? We first constructed point estimates and then we went through the pain of deriving the distribution of that point estimates, all of that is captured in this single f of theta given y, it is straight away giving you the interval of possibilities along with the distribution.

Now, you can derive whatever you want to do with this f of theta given y, you can say f of theta given y is too much for me to work with, I need to provide a single estimate to the user, I need to provide a point estimate. So, situation is reverse here, there we first derived point estimate and then derived interval estimates, here straight away I give the interval estimates, these are called set estimators you derive point estimators.

What we mean by you derive point estimators is suppose I give you the p d f of a random variable, what are the things of interests to us when I look at the p d f of a random variable, mean variance or mode or median and so on. So, if I were to ask you; give you the p d f of a random variable and I ask you to find out a single number that represents the outcomes, we have talked about this earlier on, mean, median mode, Bayesian estimation essentially works on the same idea.

First I give you f of theta given way, I give you the main piece of information from there, it is up to you, you can derive the mean of this and say that is one point estimate, what would be the mean of what do you mean by expectation of f of theta given y? What would be that? That is nothing but your conditional expectation of theta given y, what is conditional expectation? It is a best prediction.

(Refer Slide Time: 13:03)

You see the Bayesian estimation is giving you the minimum mean square error prediction that is given f of theta, if I have f of theta given y, I can calculate expectation of theta given y; how do I calculate this, integral? What is the expression at least, theoretical expression?

Student: (Refer Time: 13:21).

That is it. So, theta f of theta given y d theta, so given f of theta, given y that is once I compute f of theta given y, I can straight away calculate the conditional expectation, the conditional expectation is the best estimate of theta, best in what sense? Minimum mean square error sense, this will get you minimum mean square error estimate, earlier that something we could not achieve, we said I cannot solve MMSE problem because I do not know the truth, Bayesian estimation does that for you.

What we will do tomorrow is we will quickly go through an example on Bayesian estimation and then scan through the estimation of signal properties namely ACF and spectral densities and then day after we will look at estimating time series models.