NPTEL

NPTEL ONLINE COURSE

CH5230: SYSTEM IDENTIFICATION

RANDOM PROCESSES: REVIEW 7

ARUN K. TANGIRALA

**DEPARTMENT OF CHEMICAL ENGINEERING,
IIT MADRAS**

Now what we learnt yesterday through that simulation study is how to generate, what do you say how to simulate random process, of course if you want to generate a white noise process, then Gaussian white noise process then you can use random.

If you want to generate a uniform white noise process in MATLAB you can use tan, and one of the things that we learnt yesterday is apart, keeping the specific aside how to approach time series modeling, and in fact even that would be applicable to system identification as well in a systematic manner.

After you have obtained the data you start looking at the series itself, whatever data is been given to you, in our example it was just time series, but in a regular SID as you have seen in the liquid level case study, we spend some time by visually inspecting the input, output data, then of course in the system identification case you start looking at the power spectral density and so on, but here we did not look at the power spectral density, we looked at ACF and PACF, because we want to build a time series model.

And then we went on to guess a suitable type of model for the given time series, of course having ascertain at least visually that the series is stationary, one of the important thing is to remember is after building the model we assessed the model for its goodness and that is a very important and critical step in any modeling exercise, even if you have built a model from first principals you have to ascertain the goodness of the first principals model.

In case of data driven modeling we turn to some statistical analysis, and the one that we looked at particularly is the residuals, we said if the model is good the residual should not leave anything behind that is predictable, and that’s why we looked at again the ACF of the residuals. In example yesterday the ACF of the residuals did not show any predictability, so we were assured that at least under fitting has not occurred, but in case the ACF had shown predictability as a simple you know examples, suppose you had fit an AR1 model, we fit an AR3 model yesterday, had we fit an AR1 model, I’m sure the residuals would have shown some predictability, then the question always arises is how do I go about improving this model, should I add an MA component or should I increase the order of the AR model and so on, and there is no definitive what do you say formula or result that tells you how you go about a exactly refining your model, one has to try out different options that are available, finally we had two models an AR2 and an ARMA 11, question is which model to choose.

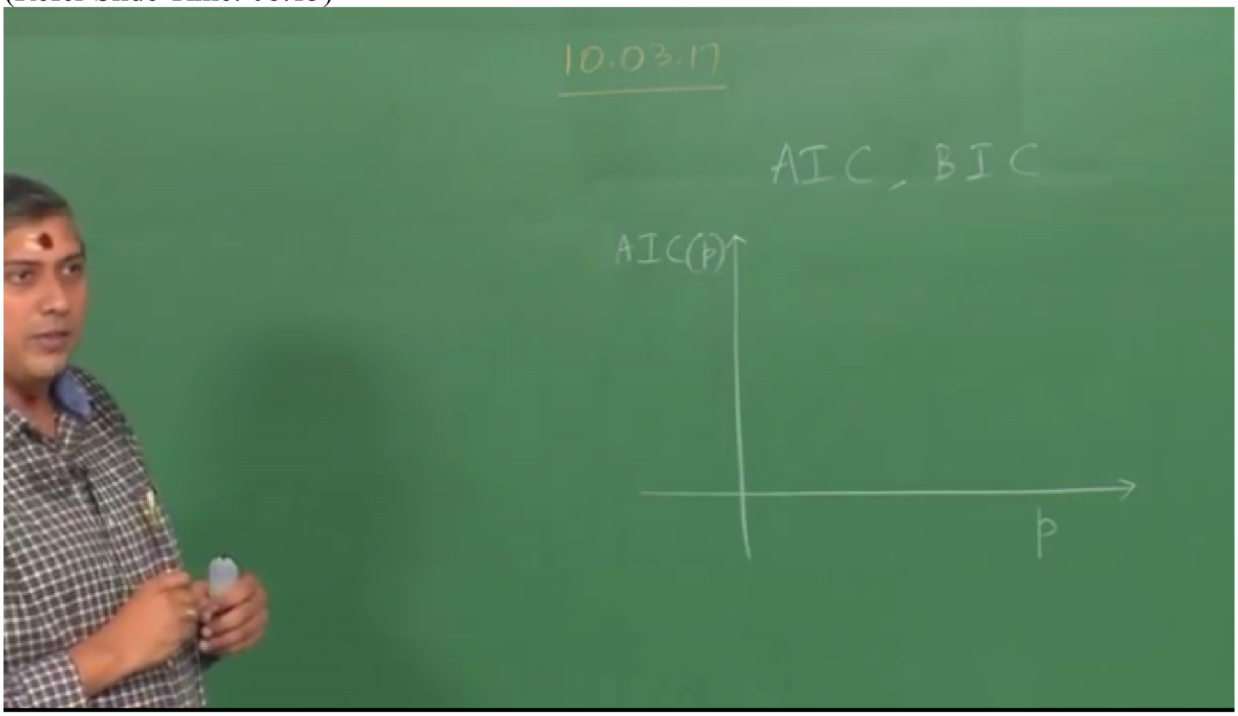
Now at this point if you look at the traditional approaches in a literature, people have come up with some criteria to select a model, and model selection is again very active branch of system identification, when I have multiple models, how do I go about dealing with it? Should I select one model or can I use all the models that are estimated and make prediction from each of this models and use some optimal combination of those predictions, that is also possibility, there are number of options that are available, we'll keep things simple, as is followed generally in the identification literature. The model, each model what it is doing is it is fitting the data that has been given to it at the time of training, and it achieves a certain degree of it that is goodness of it, and then when you present a fresh data it achieves some goodness of prediction, that is during the cross validation stage, so what people said is that look eventually it's a trade-off between how it performs on the training data and the fresh data, if it generally even over fitting occurs a model performs very well on the training data, but performs poorly on the fresh data that we have already realized at least for couple of examples.

When under fitting occurs of course it performs poorly on both, now there are what are known as information criteria methods which quantify the tradeoff between so called bias and prediction versus variability in parameter estimates, what we mean by bias and prediction is? How far it is away from fitting the data, given the training data, but when you want to achieve goodness of fits on the training data, that general tendency is to include more and more parameters in the model, like we did AR3 for example. Whereas AR2 also was sufficient, now when that happens as we have discussed in example, in the early examples that I had presented, when you over parameterized then you have included more number of parameters then necessary then in fact not only then necessary, but also more parameters then what you can actually estimate reliably from the given data, it should be viewed that way, data has some information and that information content is limited, it's not infinite.

And that information is used in estimating the parameters, when you over parameterize it is as if your, as I gave the example long ago, is as if that the food is limited and you have invited too many guesses, so lot of them are going to go out hungry, so here the parameter estimates are going to go out with lot of error, therefore the idea is actually not to really offend the predictions, make sure that when we say predictions let me say fit, let me not confuse you, we want good fits, but not at the cost of over parameterization, and so called information criteria measures some popular ones being Akaike Information Criteria called, shortly called AIC or Bayesian Information Criteria measures and so on which came about in mid-70's, early 80's, they look at these tradeoff between complexity of the model and the variability of parameter estimates or you can say complexity of the model and its ability to predict well on the fresh data, if you have parameter estimates in your model that have high error in them, then you will have poor predictions on a fresh data set because it's a sign of over parameterization, so there are measures like AIC Akaike information criteria or BIC.

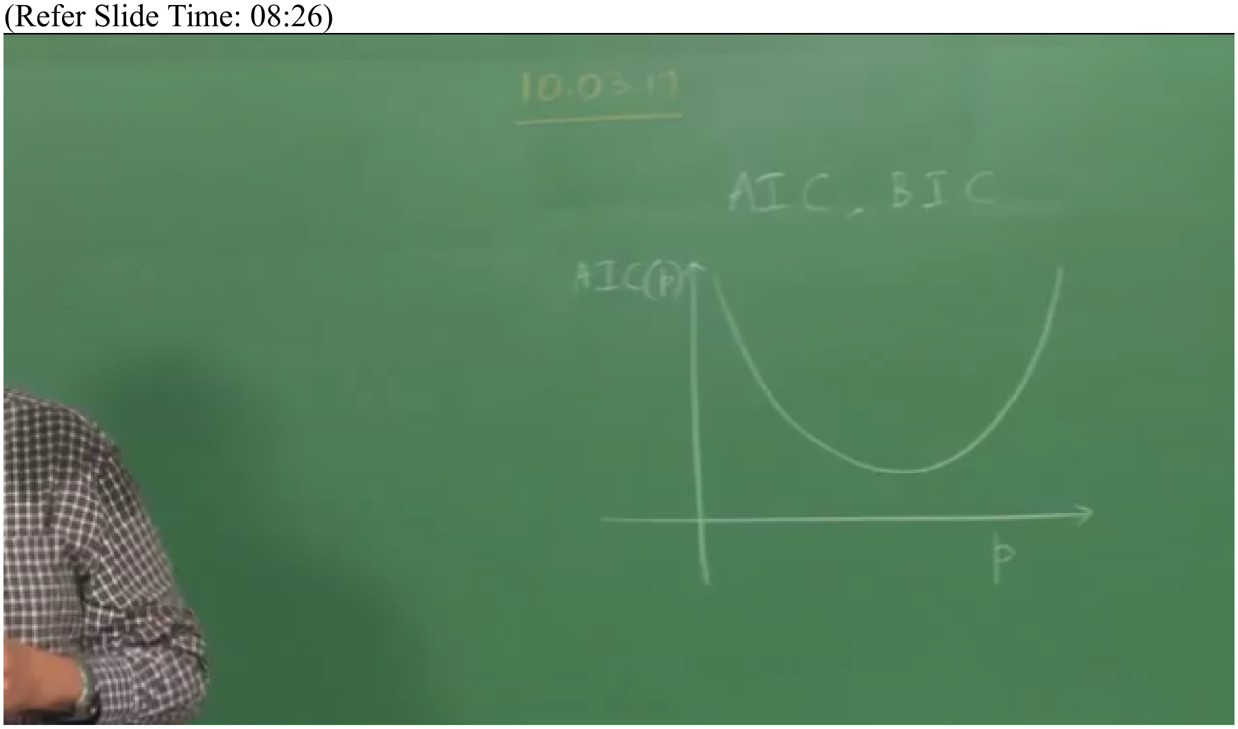
And the premise in using these information measures I'll talk about the, I'll give you expressions later on, but just let me quickly explain the philosophy, if you were to plot for example AIC, it's a function of the number of parameters in the model, so on X axis you have the number of parameters in your model,

(Refer Slide Time: 08:13)



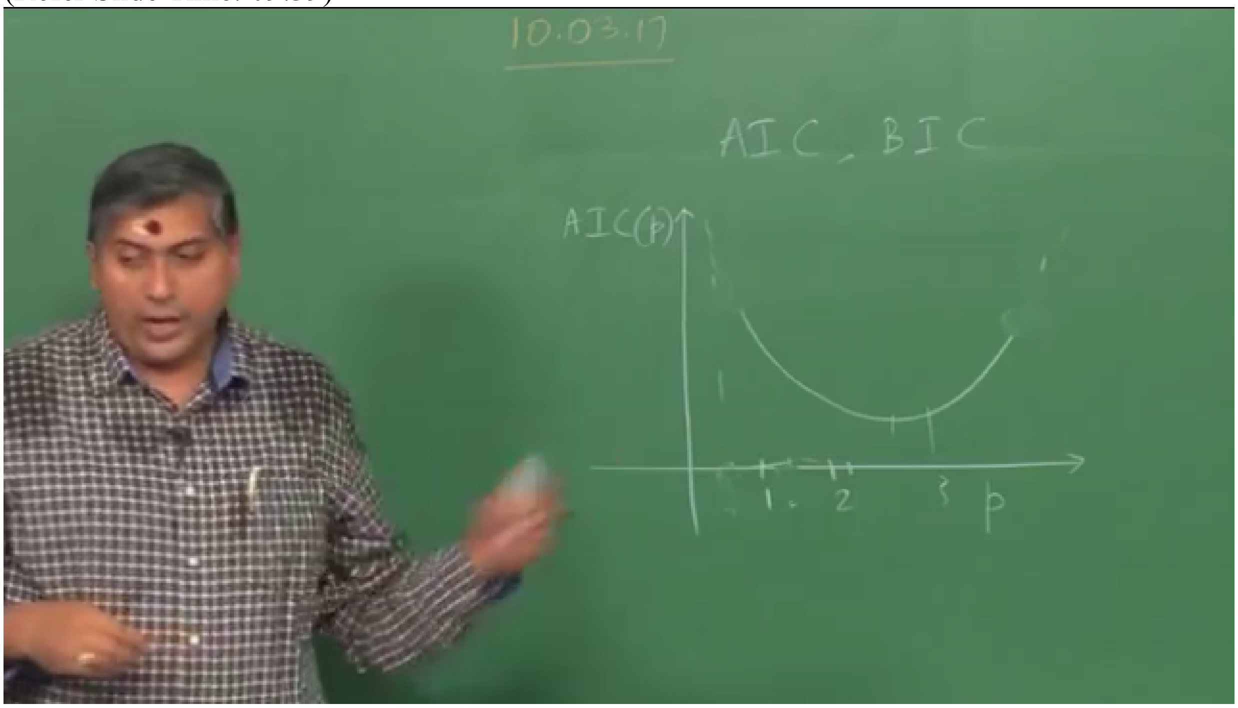
typically that is you look at ideal scenario, then you should expect to see for a given modeling exercise as you calculate this AIC and plot it versus parameters you should see U curve like this,

(Refer Slide Time: 08:26)

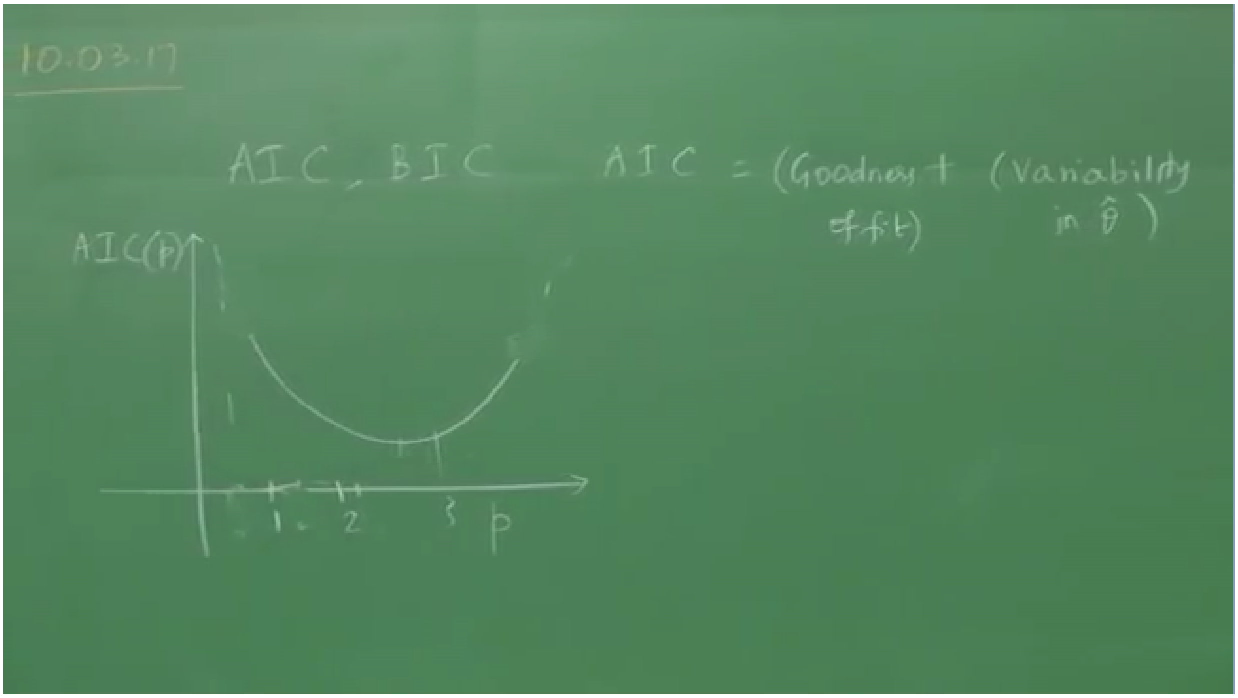


in fact the ideal situation is where it has a sharp turning point that is only one value of the parameter at which this you know minimum exists.

So what is the idea here? As you increase the number of parameters the complexity of the model increases, your ability to explain the training data also improves, but beyond the certain point, right, so for example for the series yesterday you could start with AR1 that's a single parameter, we know that AR1 would be in sufficient for a given series, we looked at the PSF, clearly it said at least a second order is required, so you would be someway here, let's say this is P = 1, right, and then we had two parameter model and then in fact we started off with the three parameter model, I'm just drawing here, so let's say 3 is here, and 2 is here and 1 is here, okay, for the example yesterday.
(Refer Slide Time: 09:39)

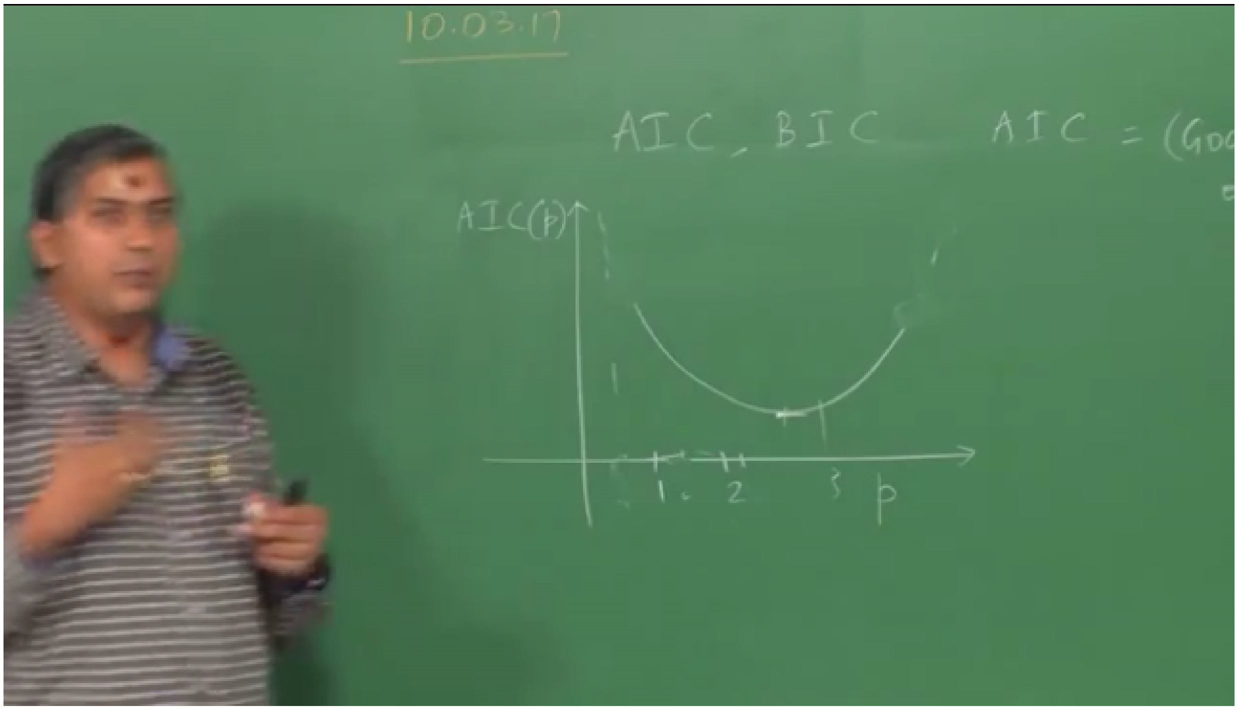


Now what happens is, what would have happen is I had fit an AR1 model it would be insufficient, and as a increased order of the AR model the fit improves, so AS improves, these term quantity is AIC and BIC have two terms in them, alright, the AIC for example is made up of two terms, one that accounts for goodness of it, I'm not giving you the expression here at the moment, because we need to learn something small to understand this, and then here you can say variability in theta herd.
(Refer Slide Time: 10:30)



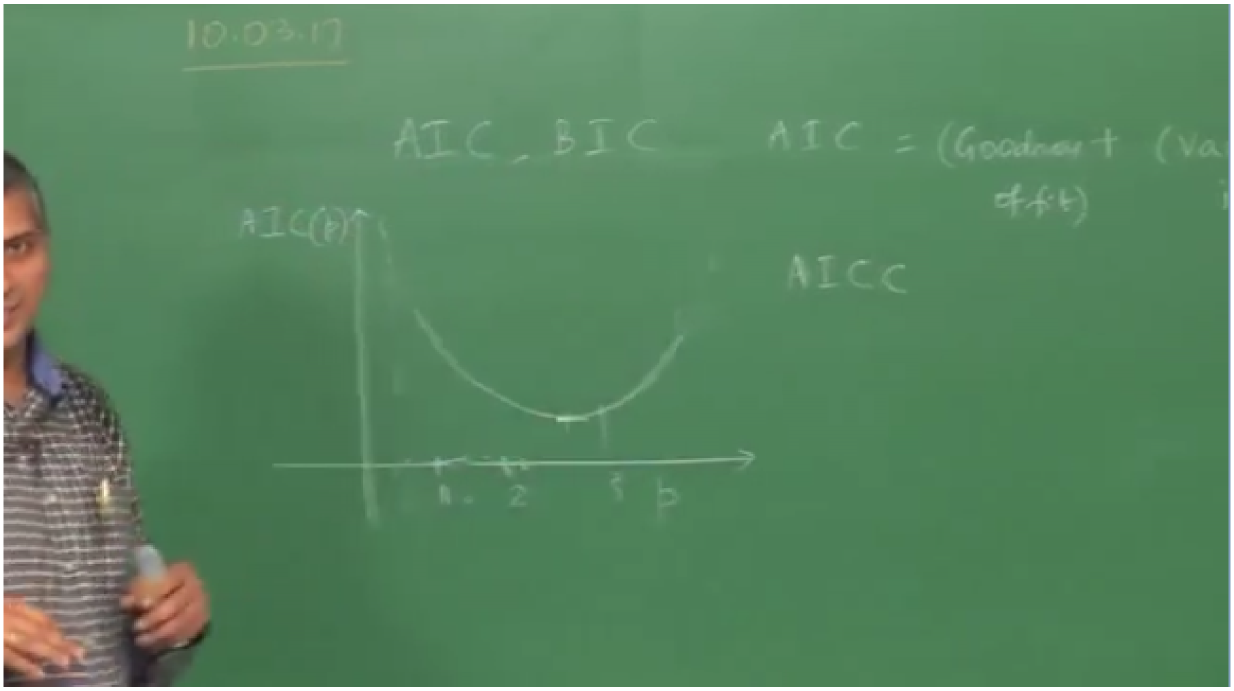
So as it fit improves of course at the cost of more and more parameters that is with the inclusion of more parameters in the model, the variability in theta herd also slowly starts to rise, because estimation theory says simply the variability in theta herd is proportional to number of parameters that you include in the model. As a result at some point when you have over short, so what happens is if you go from AR1 to AR2 in the example yesterday, it actually improvement or increase in the goodness of fit is much more than the variability, so that starts, that dominates for a while.

At some point when you have over parameterized, then you have the improvement that you get in the fit is so marginal compared to the increase in the variability in theta herd, so the second term starts to dominate, and that’s when you start to see an increase, and you want to avoid that because you don’t want marginal improvements in fits at the cost of large errors in parameter estimates, you want significant improvement of the fit at small increase at the expense of small increase in variability that is the mantra that generally we want to follow.
(Refer Slide Time: 12:00)



And what Akaike argued is that this is the point that you should watch out for, because it provides an optimal tradeoff between the goodness of it and the variability in theta herd, will you see this kind of a curve for every morning exercise, not necessarily but it's okay, so typically the model that has lower AIC is the one that is selected, and in MATLAB there is a command called AIC that comes ship, that ships with the system identification tool box, so as a simple exercise what you can do is when I post the script you put the script, you estimate AR2 and ARMA11 like we did yesterday, calculate the AIC value and see which one fairs better, we know the data generating process is ARMA11, so if AIC is doing a good job it should pick ARMA11, but that doesn't always happen.

However in the simple case it should work out to be, the ARMA11 should work out to be the winner, alright, so to summarize when you have multiple models, that is competing for a given data set you pick the one that offers the best tradeoff between model complexity and variability, and measures the information theoretic measures like Akaike information criterion or Bayesian information criterion quantify the tradeoff for you, you can use of course you know with researches right they are almost like moviemakers as I always say, in the film industry when as I say when you have a movie called thirudan let us say, then all kinds of thirudan movies will appear later on, (unsupported language 00:13:40 to 00:13:44) and so on, right, (unsupported language 00:13:45 to 00:13:46) something like that, so all this sequels follow, so when AIC was proposed all the variance of AIC came about, corrected Akaike information criterion, extensions of Akaike information criterion and so on, so you will see several improvements have been suggested, so generally in literature you will see AICC you know, that is corrected information Akaike information criterion.
(Refer Slide Time: 14:14)



Corrections essentially, the basic philosophy doesn't change, they still has two terms, how you quantify those two terms, what factors you include, why you want to include, to cater to which kind of situations and so on is what gives birth to a variant or a you know mutated value of AIC.

BIC is known to be very good, in fact AIC is conservative that means it kind of overestimates typically, and then there is issue whether this information criterion is suited for small samples or large sample sizes, whether it suited for univariate or multivariate processes and so on, so there are recommendations in the literature, we'll not go over that, go over them right now, at an appropriate point we'll probably touch base on those things, but you should be aware of this information criteria measures, of course you know theory keeps evolving, today you have other modern methods of selecting the model order or an appropriate model, but we will not discuss them, because this is an introductory course, alright.

So let's quickly move on, so what we have learnt until now is what is ACF, VACF, how to theoretically compute them and how to also estimate them, I've given you the at least the idea of how to estimate ACF and VACF routinely, we went through an example.

Online Editing and Post Production

Karthik
Ravichandran
Mohanarangan
Wilhelm Benjamin Anand
Sribalaji
Komathi
Vignesh
Mahesh Kumar

&

Web-Studio Team

NPTEL Co-ordinators

Prof. Andrew Thangaraj
Prof. Prathap Haridoss

IIT Madras Production

Funded by
Department of Higher Education
Ministry of Human Resource Development
Government of India

www.nptel.ac.in