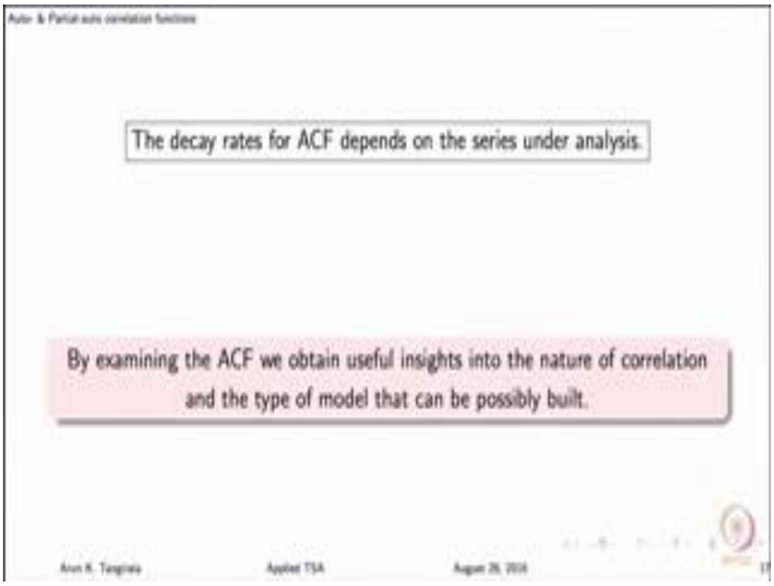
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 28
Lecture 12B - Autocovariance & Autocorrelation Functions-3

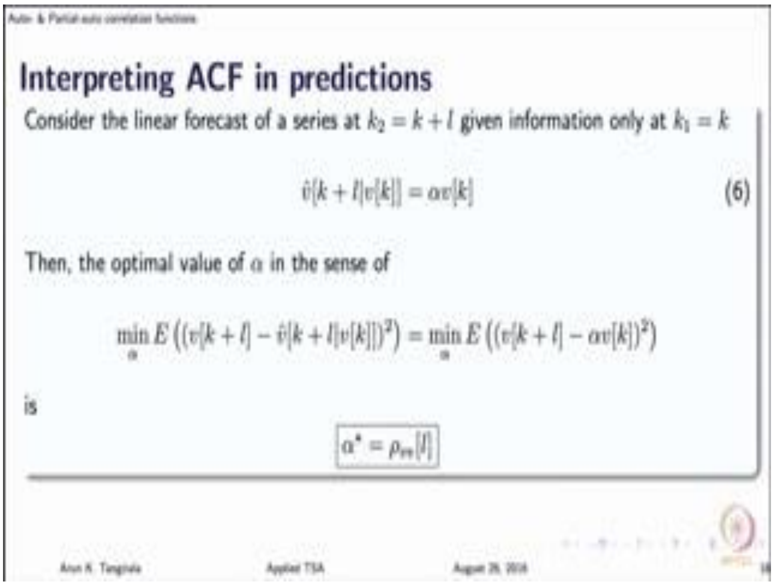
As you have said by examining the ACF, we can actually draw inferences about the series.

(Refer Slide Time: 00:14)



And that is the purpose of this entire discussion itself and some more discussions to come.

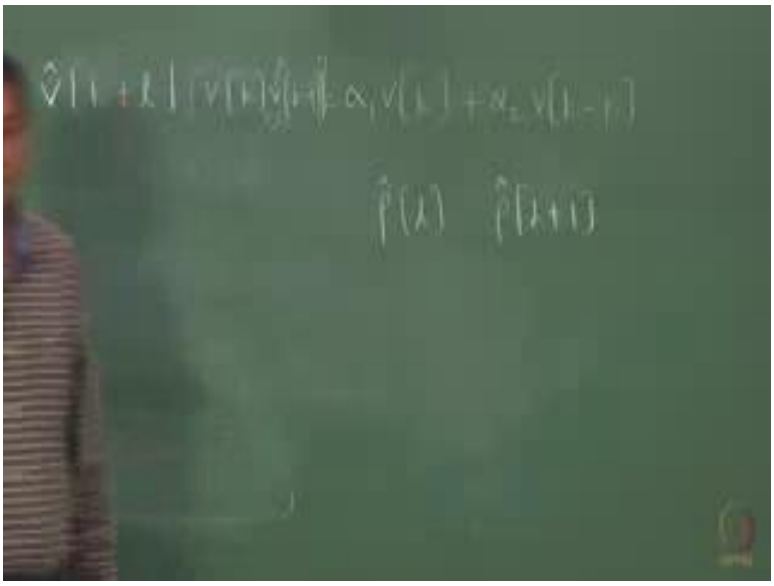
(Refer Slide Time: 00:23)



Before we move on to understanding what are the ACF signatures, it is important to interpret ACF properly and also study an important property of ACF which is a non negative definiteness. So, let us look at a simple interpretation for the ACF which is essentially what we have seen earlier. Remember ACF is a correlation measure and it is hard and we know that computing correlation tends amount to performing linear regression willing a predictive kind of model. Likewise here when I am looking at ACF at a particular lag, what does it mean; it means almost the same, but let us work this out in the context of time series.

Suppose I am given a random signal v k and I am interested in forecasting it in a linear sense, I am standing at k and I am forecasting at k plus l and we denote the forecast by the hats. Any estimate, any predication that we make are all denoted by a hat because after all predication is also an estimate.

(Refer Slide Time: 01:35)



So v hat of k plus l given v k, we have also written a different notation before v hat of k plus l given k. Now are these two notations the same not necessarily, when we write v hat of k plus l given k, it is a succinct way of saying that I have information up to k. In this here clearly we are saying I am given only v k, I am only given information at this instant I do not have any information about the past and I am trying to make a prediction of what happens l time instance into the future.

Suppose I build a linear predictor, I make a linear forecast alpha v k. So, I simply assume that this is how the observations are related and now I want the optimal estimate of alpha. Then again the optimal estimate of alpha is in the sense of minimizing the mean square predication error and that turns out to be not surprisingly the autocorrelations because remember, when we solve this problem for two random variables what did alpha turn out to be, what is it turn out to be; instead of v v k plus l and v k, we had y and x. The optimal estimate alpha was a correlation between y and x, exactly the same result we have here.

So, correlation between v k plus l and vk is nothing, but autocorrelation at lag l. So, this is not some new result that has following form the skies, it has already been there with us, but in the context of random variables clear. So the autocorrelation at any lag l, contains or encodes the predictability inform a predictability or the ability to forecast l steps away, using this sample alone. The other day we solved at slightly different problem, we said if I use two observations in the past then we had this couple of equations which we called as Yule Walker equations; do you remember, when we talked about stationarity. There we had two autocorrelations or autocovariance is coming up at lag 1 and lag 2.

Here we are only using single observations even now you can extend this discussion likewise based on at example. Suppose I am given v k and v k minus 1 then I would like to build a linear predictor, suppose this is a case then what you expect alpha 1 and alpha 2 to depend on, what properties of the series does it depend on or what statistic does it depend on. It is not different from the example that you solved earlier, there instead of l we had v k I mean l was 0; I mean in fact may be v k plus 1; there we had v k plus 1 being predicted in terms of v k and v k minus 1.

Now, you have v k plus being predicted in terms of v k and v k minus 1. So, what do you expect alpha 1 and alpha 2 to depend on?

Student: (Refer Time: 05:10)

Auto correalate autocovariance or autocorrelation at what lags.

Student: l and l plus1.

Perfect l and l plus one that is it, so once again this conforms a fact that; so linear models it is sufficient to know the autocorrelations and to work with autocorrelations as long as there in variant with time my linear model does not change and so the process as well correct; yes.

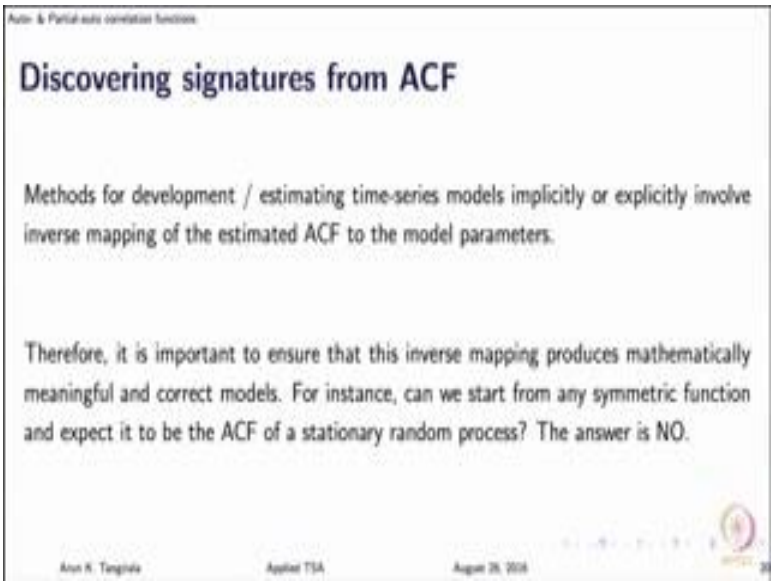
Student: Is that correlation between k and k minus 1 (Refer Time: 05:37).

K and k minus 1 not in this, that is a good question; that will happen in non-linear model. In a linear model it is like a super position how much does v k influence the variable of interest, how much v k minus 1, but together they will participate, but not between v k and v k minus 1, you will have in your equations sigma at l and sigma at l plus 1, but there will not be any sigma 0 per say I mean implicitly the variance does effect, but not alpha 1 and alpha 2 (Refer Time: 06:17). So, if you can see in this example d is the variance actually effecting yes and no you can say because yes because; it is a part of your autocorrelation, but if you look at as a correlation, correlation at lag 0 is any way 1.

So, one affects anything; so in that sense it does, but yes the answer to the question is what is important of the auto what are the important of the autocovariances at lag l and n plus 1only; good question, any other question fine.

So you should remember this interpretation, it helps you in understanding ACF better. Now we move on to this main business of recovering the discovering signature here actual discovering from ACF signatures; the title is slightly wrong.

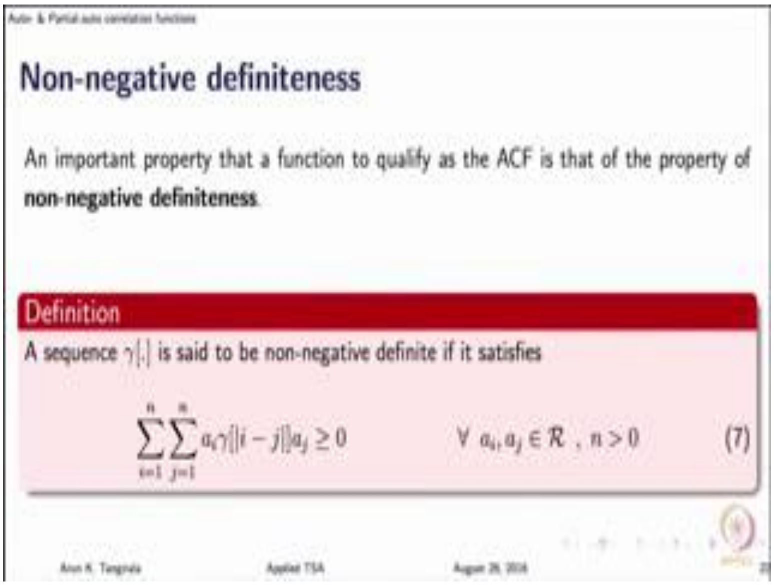
(Refer Slide Time: 07:04)



But what are we trying to discover; the features of the series, the characteristics of the random process from the ACF signatures or the ACF signatures and as I said before we get into a theoretical study on that there is an important property that the ACF satisfies; apart from all the properties that it inherits from correlation symmetricity and all of that there is another property that ACF has because it applies to a time series which is that of non negative definiteness.

Now, non negative definiteness does not mean and it should not imply that the ACF is always non negative value please. We are looking at the entire sequence of ACF, we are not looking; we are not talking about the value of ACF at any lag, we are actually talking about the full sequence from you know minus infinity lag to the plus infinity lag theoretically and you are saying this entire sequence as to be non negative definite, what is a definition of a non negative definiteness of a sequence.

(Refer Slide Time: 08:20)



Now any sequence is set to be non negative definite, if it satisfies the condition that you see on the screen. There is something called positive semi definiteness, positive definiteness, non negative definiteness and so on, all of them are related in linear algebra, you do come across these terms as well for matrices you say matrices positive semi definite, positive definite, non negative definite and so on. For matrices what is a understanding if I say matrix is positive definite.

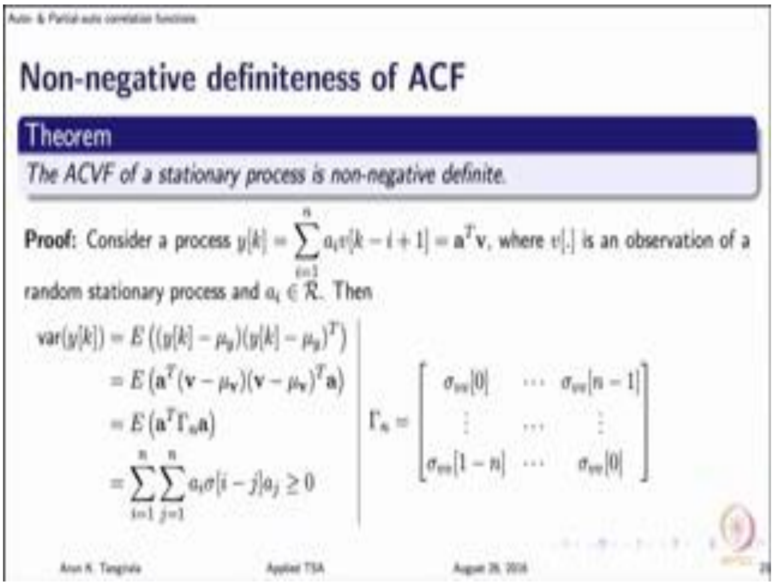
Student: Eigen value.

Eigen values are strictly greater than 0, here we are talking of sequences in fact, you can actually arrange this sequence in the form of a matrix as we will see in the proof and then talk about and then relate the same condition that you see here to the profit, to the condition for the non negative definiteness of the matrix; it is one and the same. This is more of the do not worry about the proof of this; we will go through the proof, but do not worry so much about the proof rather than the fact that ACF is non negative definite. Why is this non negative definiteness important, will become clearer shortly and later on in estimation theory as well.

Now, how do we prove that the ACVF of a stationary process is non negative definite? The proof is fairly straight forward, what you do is given the series v k; you construct another series y which is a linear combination of some n observations of the given series and write it in a vector form as you are seen where a (Refer Time: 10:01) or combination coefficients. We want to prove that the ACVF of this process v k is non negative definite.

What is a requirement, if you go back to the definition it says that, this summation should always be greater than or equal to 0 for all non trivial values of a; that is what it says and of course, the a has been real value.

(Refer Slide Time: 10:33)



Now when you look at the proof is fairly straight forward, you start with variance of y which you have constructed artificially from v k and we know the variance of any random signal as to be non negative; at most it can be 0 rather at least it should be 0. Now we use that to prove that the if the proof is self explanatory, we start with variance of y k substitute for the expression of y k and then perform your expansions and I am not going through it step by step because fairly self explanatory and it is not so much necessary to know the full details, but if you are convinced about the proof, it should be good enough and the last step tells you in the set of equations, it tells you that the ACVF is non negative definite.

We have only assumed a is to be real value that is all; in the construction of y. Therefore, the proof is self evident, but in this we have actually introduced a gamma; big gamma matrix which is a matrix; what is this matrix gamma that you see here, what does it consists.

Student: (Refer Time: 11:49).

It is the variance covariance matrix, but for the series in fact, it is a variance autocovariance matrix sometimes we do not use a variance it, simple the autocovariance matrix. Of course, we have used only n observations here you can now extend this to infinite, the definition says that it should be valid for all n as well, for any n it should be valid, therefore, kept the proof fairly generic we have assumed n to be some natural number and shown the proof, but this autocovariance matrix is something that you should get comfortable with because when we talk about estimation of model parameters for auto regressive models or even in the Yule Walker equations and so on, this autocovariance matrix will make it is appearance.

It is just back you know arranging your autocovariance is in a particular way like we do in the variance covariance matrix; that is all it is, the diagonals contain the variance and half diagonals contain the autocovariances. So, with this hopefully you are convinced that ACVF of a series is non-negative definite. Why are we talking about this, the point is we have said ACVF is symmetric or ACF is symmetric; does any symmetric function that I give you qualified to be the ACF of a stationary process and this property tells us no, they not any symmetric sequence qualifies to be the autocorrelation function of a random process; why are we raising this question because in practice I am going to estimate ACF, when I estimate ACF, why I am estimating ACF of course, to figure out whether the series is predictable, but beyond that we also use ACF in estimating the parameters.

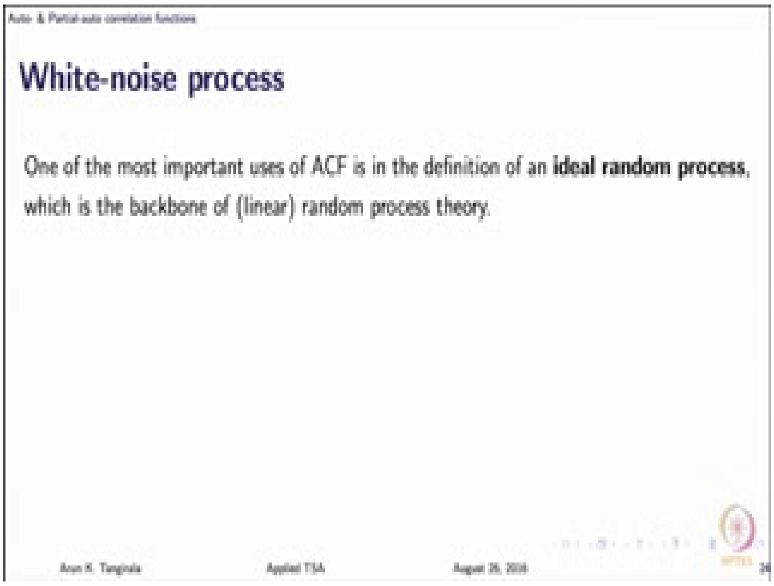
For example here we know, if I want to estimate alpha 1 and alpha 2, I need estimates of the autocorrelations; I cannot use any estimator that will get me a symmetric ACF, that estimator has to produce ACFs that are non negative definite because I know that the underline process is random and I am building a model for that random process and I know from theory now, that the ACF is non negative definite. So, when I sit down to estimate alpha 1 and alpha 2; I am going to work with the estimates at lag l and lag l plus 1 and so on for other types of models.

This estimates the entire sequence when whatever formula that I am using for estimation should produce ACF estimates that are non negative definite. Otherwise what can happen, I will still be able to estimate alpha 1 and alpha 2, I will get some values, but the problem is the model does not have an important feature that the random process has, which is if I were to compute the ACF s from that model, I cannot guarantee that it will produce the non negative definite ACF. In other words, I cannot give you any symmetric function and claim that it is ACF of a random process, it has to be non negative definite as well.

We will look at an example shortly, but before we do that let us actually pay or obey sense and respect this very fundamental process in the entire world of time series which is the white noise process. We will come back to non negative definiteness and submit later we keep this at the back of our mind; it is time to move on. This white noise process as already made it is appearance many a times in our lectures, now it is time to define it formally. Before we do that, let me just quickly tell you why we are defining a white noise process, we said that given a series we would like to test the series for predictability.

Suppose the series is unpredictable and how may I testing it through autocorrelation. Suppose the series is unpredictable, I need to know how the ACF looks like theoretically to begin with that is in other words I am bench marking all predictable process against an unpredictable process, so, this white noise process serves as a bench mark.

(Refer Slide Time: 16:32)

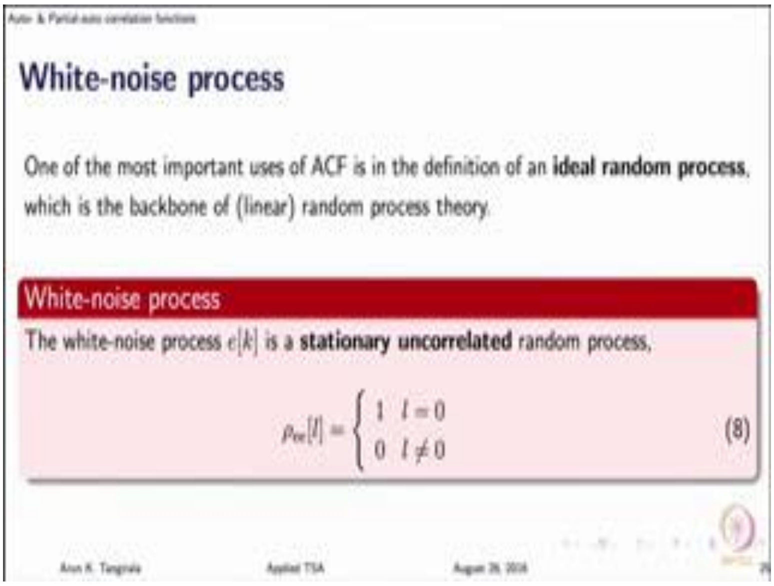


If bench mark in the sense of unpredictability, if the given series as the characteristics of a white noise process then I give up in terms of building linear models, why because the white noise process is defined in terms of ACF correlations only this is different from an IID process.

An IID process demands, what is it demand; independence which means what there is absolutely no relation between any two observations whereas the white noise process by definition has no correlation only; 0 correlation which means there is no linear influence between any two observations; that means, I can have a white noise process and I can still predict, but using non-linear models. A linear models cannot predict what you mean by cannot predict is, you cannot improve upon the prediction of white noise process beyond its mean.

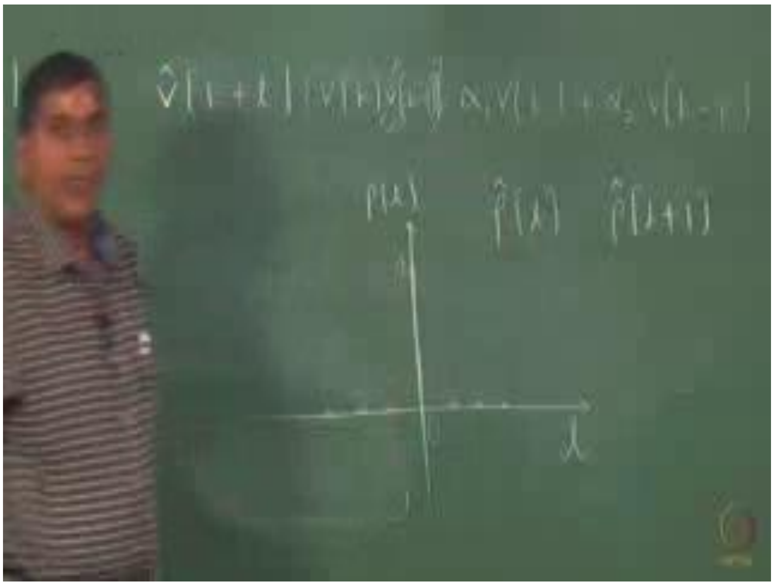
The mean is a best prediction; in time what is a difference between prediction of a time series or of a random process and general prediction of a random variable. In the prediction of a random variable, we simply use expectation where as prediction in time series works with conditional expectations that is given the past, I am making a prediction and we are not working with conditional expectations per say, we are working with linear functions that is what we are doing and correlation is measure and that is how all the theory comes in.

(Refer Slide Time: 18:09)



So white noise process is a stationary; uncorrelated process only, it does not tell you what the variance is; it only talks about the correlation, that is why it is useful to talk about correlation rather than covariance's.

(Refer Slide Time: 18:22)



How does the ACF look like, it is an impulse function for a white noise process at lag 0 it is 1 by definition and elsewhere it 0 and so on. So, the ACF is an impulse sequence, so remember one of the reasons why we are defining white noise process is; it is equivalent to bench marking.

Now, there are two other important purposes to white noise process and that is why you should really be comfortable with this concept of white noise process, it is all friction. In every class you will be learning some fictitious thing, but all is friction helps us in making a forecast that is all you should remember and for some it helps you in getting placements in good companies, so that you can make white noise and interview.

Now, the second purpose of a white noise process is in modeling. Every stage of modeling the white noise process appears, the first stage is for bench marking the predictability, the second stage is in accessing the goodness of a model, what we will understand by asking this question, what do we aspect of a good prediction. When I say, I have made the best prediction of a series or of a signal random signal, what are we implying; are we implying a predicted accurately.

Student: No.

No because is a random signal then what is it mean?

Student: If there any (Refer Time: 19:53).

Sorry.

Student: Reducing the error

Using the.

Student: Reducing the error.

Reducing the error that is a (Refer Time: 20:00) statement.

Student: performance of the model on some kind of (Refer Time: 20:03).

Good, but that is not what we are looking for, that is a at a later stage. You will only apply it on a test, see I will give you an exam of course, the instructor ask me to give an exam, but given the freedom I will conduct an exam only after I am convinced that you have learnt correct, but institute is forcing me; the calendar is kind of forcing us to conduct exam despite the fact that you probably have not learnt enough. In the case of modeling, I have that luxury; there is no calendar asking me to conduct an exam for that model; validation that you are speaking of is like conducting an exam. How do we know the model has learnt enough; think any answer from the other hall; yes.

Student: If we subtract the additional from the model from per data and if we get white noise.

Why?

Student: Then (Refer Time: 21:00) if a model is good enough to predict a data will

Student: (Refer Time: 21:04) then what we should get by interpret should be white noise so that.

Student: That means, we have taken into account everything that is predictable

So let us see if there is something else from the other hall quickly then we will wind up and continue our discussion to the next class. So, I do not here me answer from there, it is a good point in fact that is a key point; when I am building a model and you should really get this ingredient in your minds because every time you build a model, you have to subject the model to what is known as a residual test. You have built a model, using the model on the given data you make a prediction on the so called training data, like I keep asking you know there is so much analogy between time series modeling or any modeling; not only time series modeling and the learning process.

You are in a course, you are preparing for an exam and you are also trying to understand the subject, you are given assignments. What is the purpose of giving you assignments is to train you to; test your learning also. At the time of exam, you should take up the assignment problems that you have already been, already looked at and see if you have understood everything; that means, whatever is left over; should be kind of random. There should nothing to learn systematically, what we mean by random is something very specific; the numbers in the problem or something like that where as for as concepts are concerned, you should have learnt it.

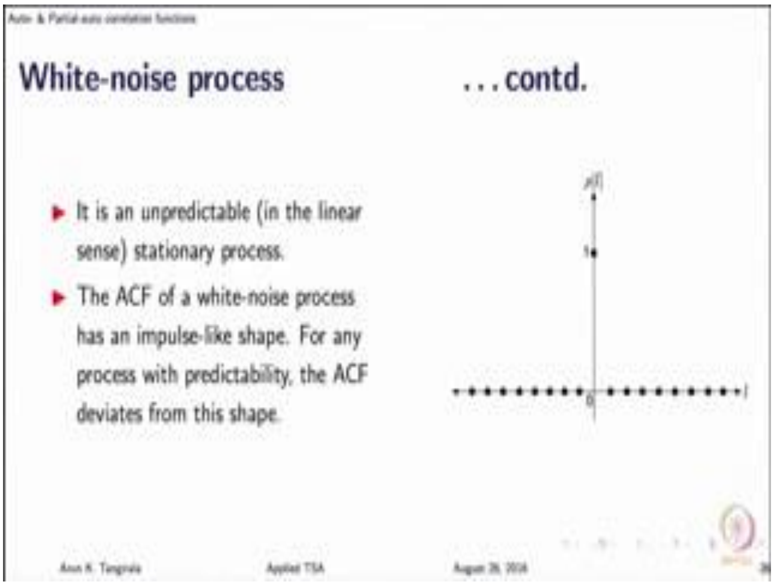
Likewise when I present the training data to a model with the help of the estimation algorithm, the estimation algorithm is a teacher there. It is helping the model to learn the process characteristics in such a way that whatever is left over and given that there will be always be something left over because I am dealing with the random signal.

Whatever is left over should not have any predictability in a linear sense, we are not going to repeat that it is a; for us predictability means always a linear sense unless otherwise specified. So, the left over which are technically called as a residuals should not have any predictability in them which means it should have a characteristics of a white noise process. We are not worried about the values remember; a white noise process is not defined in terms of values; it is defined in terms of a statistical property.

So, the residuals should have the same statistical property as that of a white noise process which means when I plot the ACF of the residuals of course, in practice I am going to look at estimates, it should look like an impulse and that is called residual test for the model, for any model it does not matter whether you willing ARIMA model, non-linear models it does not matter the leftover of the residuals should not have any predictability in them then your model is adequate.

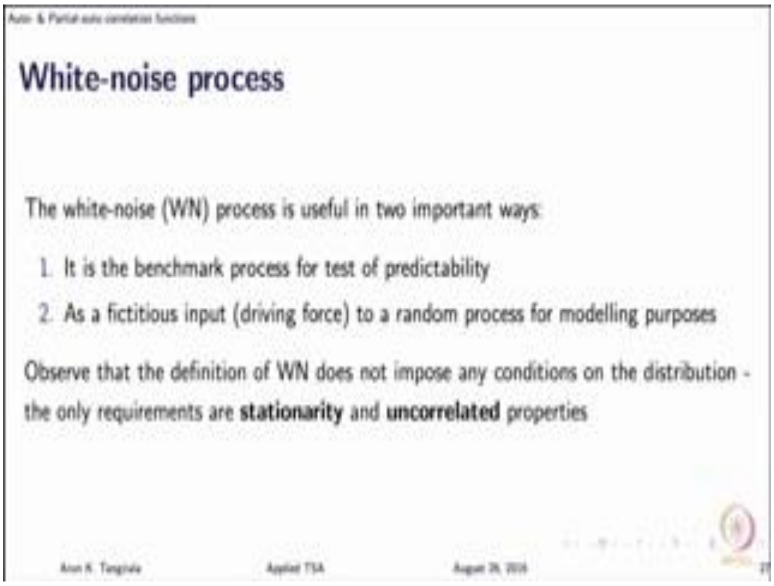
So, what is that test, it is a test of under fitting whether you have under fit or not. Whether you have fed the guest that has come with enough food or not sometimes you do not bother the guest can over heat also, but my concern is a guest should not under heat, the guest can over heat, roll on the floor I does not matter it; I will give a bed water with, but should not hungry because then I cannot sleep, but in modeling unfortunately you have to worry about over fitting also. So your model is a very special guest; over fitting test will worry about later on, but under fitting is what is the most important.

(Refer Slide Time: 24:43)

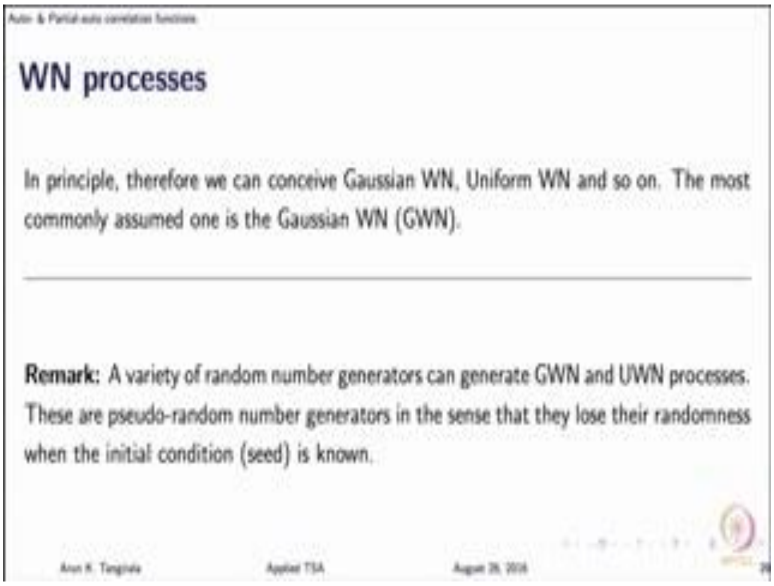


So to conclude this class, we have this white noise process which has an impulse like ACVF and this is different from the IID process and remember there is no imposition on the variance of the white noise process, we are not worried about that, we are only worried about the correlation and by definition it is a stationary process and we will use E k to denote do not ask me why I choose that notation, may be it is the common last letter to both white and noise and that is why I picked I do not know.

(Refer Slide Time: 25:11)



(Refer Slide Time: 25:14).



But that is what it is and as I said, it has important roles to play that is yet another important role to white noise process, which we will discuss in the next class.