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

Lecture – 46
Lecture 20A - Models for Linear Stationary Processes 10

Very good morning, Hope you had a nice holiday. If you look at the weather really sometimes we feel that it is a classic case of intention being there, but implementation being 0, it is intending to rain for last 2 or 3 days.

Student: (Refer Time: 00:34).

But implementation is 0, no rain. So, hopefully it is implementation will be full today. What is your forecast, did you check up the forecast?

Student: Already started (Refer Time: 00:47).

Already started listening some implementation, but did do you look up forecast anytime.

Student: (Refer Time: 00:58).

Sorry.

Student: (Refer Time: 01:00).

No, at least in the morning do you look up the forecast or no; at least for the sake of being in time series analysis you should get use to looking at forecast. And generally you can say if it rains then you reverse the prediction end. That is why you work with a practical forecast. Good, let us gets started. So, today what we are going to look at is exclusively autoregressive processes. We discuss quite at length the moving average process.

Now if you recall the moving average process takes birth from the linear random process when we assume a certain kind of structure for the impulse response coefficients which is that the impulse response dies of after some finite lags. And we introduce that assumption primarily to handle a fact that there are an infinite number of unknowns to be estimated. But obviously, that is not the only way we can take another approach where we assume that the impulse response does not go down to 0 after finite time instance, but decays exponentially. Of course, that assumption alone does not help me address a problem, because still I have infinite number of unknowns to estimate. But before we go further why is it that we are assuming that h of n decays exponentially, where we assume that the impulse response should decay exponentially what is the reason. Where is the question of stationary?

Student: (Refer Time: 02:52).

Ok, but is there anything in the definition of the linear random process that is forcing us to make this assumption?

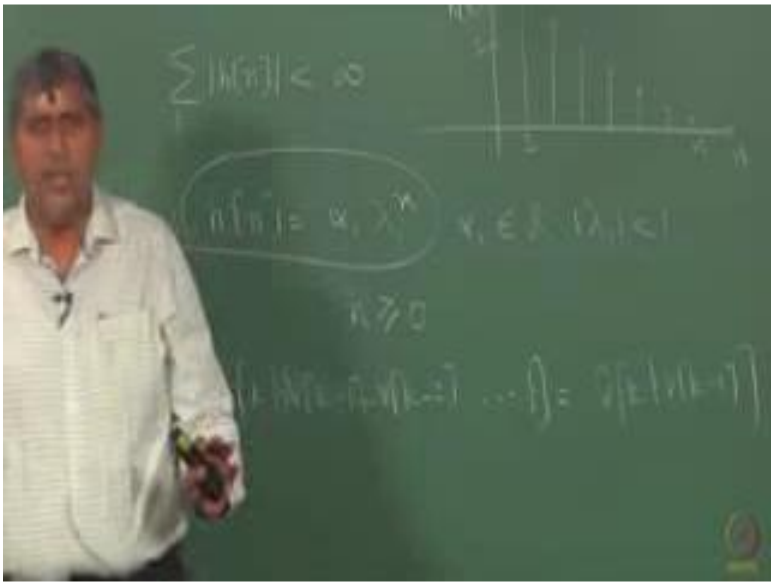
Student: (Refer Time: 03:02).

Convergence of what?

Student: (Refer Time: 03:05).

Go back to the definition of linear random process in your books and see what is it that we have stated clearly, what have we assumed on impulse response coefficients; absolute convergence.

(Refer Slide Time: 03:23)



We have assumed that the impulse response coefficients not only assume in fact we have required this to be holding because we said when only when this is satisfied I can assume that series convergence in a very strong sense almost sure convergence. And secondly, this also implies that the variance is finite and the auto covariance also is absolutely summable. And so there are several nice consequences of this requirement and essentially amounting to convergence and stationarity, and therefore we have to necessarily assumed that the impulse response in decays exponentially, but as I said a few minutes ago this alone does not satisfy or actually solve our problem of this infinite number of unknowns, we have to make another assumption.

And that assumption should allow us to go from infinite unknowns problem to a finite unknowns problem that is the idea. Earlier in the moving average case we resolved that issue by simply truncating the series, and we justified that by saying yes there are many processes that satisfy that. In fact, if you take many real life processes the ACF does support that assumption, so it was not very unrealistic one. Nevertheless, there are many other class of processes that do not really lend themselves to that assumption of finite impulse response coefficients. As you see in many of the real series the ACF kind of decays exponentially, which means that the impulse response is also not really find deign down after finite lags. And that is a reason we are pursuing this second approach.

So, now the problem is of converting these infinite unknowns problems to a finite unknown problem by what an approach known as parameterization. That is now we assumed for example h of n as you see on the slide is for example, alpha 1 lambda 1 rise to k. Where alpha 1 is some real number and lambda 1 is such that it is less than 1 in magnitude. Why we are requiring that lambda 1 will less than 1 in magnitude, because the impulse response questions have to decay. Otherwise there is no guaranty, remember now we have already said that IR coefficient should decay.

Now, why are we doing this? Why have we assumed that h to be of this form? Can it be of some other form? As long as it decays it is that is fine, but are there multiple ways of decaying yes it is possible, this is not the only way that h of n can be decay it can decay in complicate manner to. This is one of these simplest forms of decay known as the exponential decay, exponential does not mean always e to the power of something some kind of power law you can say.

Now, did some angel come and tell me that this is how it should decay? No, I am just assuming it for mathematical convenience. Why is it mathematically convenience? We will see shortly. But the most important thing that we should observe is by assuming this power law this kind of an exponential decay we have done something, we have actually converted a problem that was unsolvable to a solvable one, how?

Student: So, only (Refer Time: 07:15).

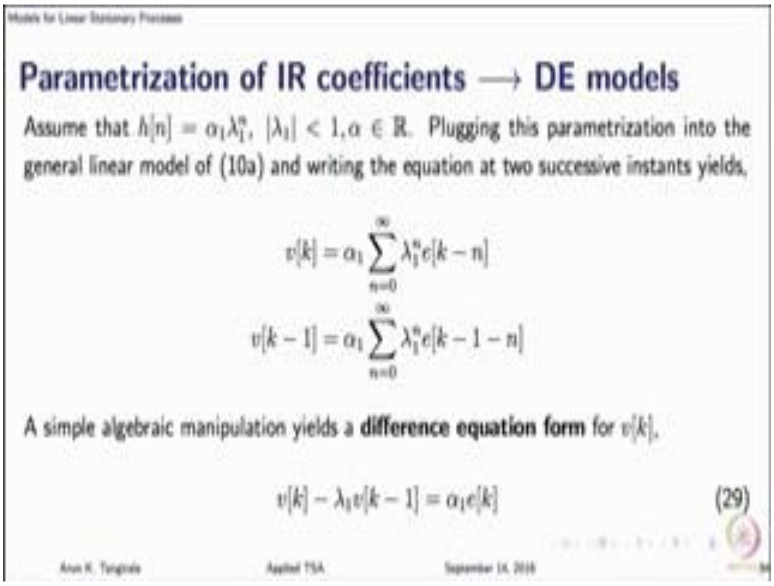
So only two unknowns I have to estimate alpha 1 and lambda 1 that it is. But this is a standard technique it is not something new we have seen this in high school even in our plus 1 plus 2 syllabus we have seen Maths; parameterization of curves. See you have to look at this way I have some sequence h of n that decays. So, let us n equals 0, anyway we have fixed it to 1. And then it could increase for example, it could be greater than 1 at h at 1 and may be even greater than 1 subsequently, but gradually it should decay; that is a requirement. And this is required for stationarity. In the MA model we have assumed that here after some m it is 0 thereafter.

But now we are saying; no, it need not be 0 it can go to 0 only asymptotically, but then there are infinite unknowns. However, now I solve that problem I assuming that there is a thread that is connecting all those dots, all those points. And instead of trying to estimate every point on the curve or the thread I am going to assume a certain equation which we actually call as a parameterization. And why should it be this form? What is the meaning of is assuming this is what we are going to study now.

What tells me that h of n should have this form. It could be for example, the second one that you see on the slide, it could be sum of two exponentials alpha 1 times lambda 1 rise to k plus alpha 2 times lambda 2 to k; nothing is stops you from assuming that. Can I know up front that such power law is binding the impulse response coefficients for the given series? Yes I can, but first let us understand now what is the consequence of this parameterization on the ACF, with ultimately we will always turn to ACF to tell us something about the series that holds a key.

So, now what we shall do is we shall take this parameterization for the impulse response, we will solve it for the simple case that is we will begin by assuming that h only varies this way. And see what form the linear random process equation takes. You must have guest what form it takes, but let us see what happens. So, let us now look at actually this case and write the equations for the linear random process at two different instances: at k and at k minus 1. And the equations are shown on the screen fairly straight forward there is nothing magical about it.

(Refer Slide Time: 10:16)



All I have done is I have substituted in place of h of n I have substituted alpha 1 times lambda 1 rise to n.

Now, the reason for writing v that is this equation itself at two successive instants in time is what? You must have; I mean we have we look at this kind of situation early on even in geometry series when you are looking at a geometric series, sum of geometric series we use this trick right to calculate the sum of the geometric series more or less it is the same thing. So, when we algebraically manipulate these two equations and write we end up with this nice first order difference equation.

So, that is the main point here. Do you see that how the different equation is born? It is very straight forward it is the simple algebra there. So, v k minus lambda 1 v k minus 1 is alpha 1 e k. So, look at how this simple parameterization as allowed us to convert first of all an unknown infinite number of an unknown problem to finite unknown problem and we have change the shape of the equation itself, we have converted that convolution equation to a difference equation form.

What is the advantage of a difference equation form plenty, there are number of advantages. One of course is it is easy to interpret. Now one of the interpretations are immediately falls out as a consequence of this assumption is that by assuming this kind of a parameterization what I am implying is that the process is dependent on it is immediate past; it evolves as a function of it is immediate past. So, I may make this mathematical statement- yes, behind every assumption there is an implication and this is the first implication of this assumption that we are making on h.

Whenever I parameterize h this way whether I stated explicitly or not I am implying that the process is evolving according to a first order difference equation; whether you stated or not that is a fact. And this mapping is unique, in the sense that instead of assuming now this parameterization I might as well postulate a first order difference equation model. If I have to solve the first order difference equation for the impulse response what would I get as a salutation for the impulse response?

Again I will recover the same thing alpha 1 times lambda 1 rise to; and of course assuming 0 initial conditions. Asymptotically at least I will end up with that kind of an impulse response. So that is the point here, whether you assume parameterization of that form or you assume a first order difference equation the meaning is the same. In the literature of the time series analysis you will actually see the journey starting from the postulation of different equation form, but unfortunately it is not so enlightening, why? Somebody would define linear random process in a convolution form and suddenly think of a difference equation form is not clear.

Of course, you know history tells us a few more things that people did not come up with a linear random process in the first place. The Weiner school of approach assumed an autoregressive form like in the difference equation form as a starting point, whereas Kolmogorov school of thought started off with moving average thing and then they both converged when they asked the question when are these representations possible then they collided at one point which is the point of the definition of linear random process. So, we have taken that as a starting point. In certain text books, in fact in a lot of text book that are widely available the linear random process is not necessarily the starting point, but I would like to present it that way so that you see the unification right up front.

And then see the moving average and autoregressive as children of those forms. And that how they evolve under some special circumstances. We need not always present any subject, and this is true not only for teaching but for research also that the chronology of presentation need not coincide with the chronology of discovery. Suppose I am cooking nice dish and I have discovered how to make this very nice dish, I might have gone back and forth in terms of iterating you know I might have forgotten coriander chilies or whatever ran to the shop that do not form the part of the recipe; run to the shop get coriander have you see anywhere in the recipe forget deliberately some spices, it is nothing like that. Discover accidently after burning your tongue nothing like this is a part of the recipe. How is the recipe you are represented? As if it was discovered in the mother's womb itself.

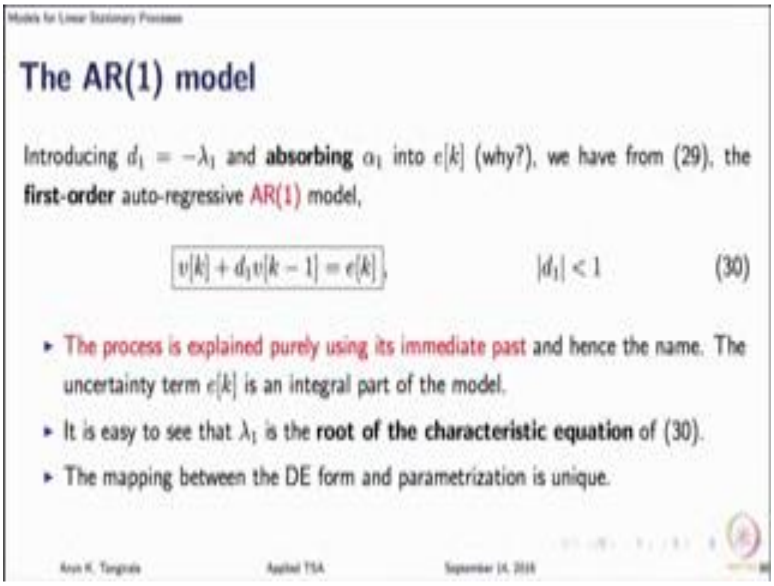
Take this many these are the grains, this much salt, this much pepper and so on. No, as if Kekule discovered benzene structure in his dream. Kekule apparently had discovered the structure of benzene in his dream, now you cannot go in to his dreams now and verify, but let us assumes it is true. So, these things did not come out of dreams, therefore do not get confused when a presentation in your course is different from what you see in the literature in terms of the chorology of discovery. The most important thing is after discovering all the results it is very useful and also important to put things in perspective and then put in order and say- I know I wish this was the starting point.

So, that feature generations do not have to go through the same ordeal that the person would discover this had to go through; you spare people of those ordeals. Likewise here, if you look at the time series literature you will see you know people independently proposing autoregressive models, moving average models, and then some where linear random process is born the definitions is born and so on. We will keep all that aside, we will learn as if the linear random process was the starting point and see moving average and autoregressive as special cases.

So, this kind of perspective need not exists in all books, it is just a matter of how much time you have spent and how you have reflected on the topic. Anyway, so coming back to the point the autoregressive model whenever you assumed now you should understand at the back of your mind it behind the since what you are actually implying is that the impulse response coefficients decay exponentially. And we will discover something more as we go along; any questions until this point.

So, now we have already assumed that lambda 1 is less than 1 for stationarity, so we know that this different equation form is guaranteed to give me; I am not made in a additional assumptions in deriving these difference equational form it is just an algebraic manipulation. So, I am guaranteed that these difference equation forms will give me a stationary v k so long as lambda 1 is less than 1 in magnitude some that we have seen earlier also. So, we replace now to bring everything into the notation that we have been using, we replace lambda 1 by minus d 1 or you can say minus lambda 1 by d 1.

(Refer Slide Time: 18:06)



And absorb the alpha 1 into e k; I do not have to assume alpha 1 as a separate thing. In fact, if you see this kind of d k here I have not mentioned, but we are assuming this kind of a d k for sorry, I should have corrected n is a dummy variable. So, n greater than or equal to 0, anyway we are assumed h of 0 to be 1. So, what will happen at n equals 0 alpha 1 as to be 1. So, by observing alpha 1 into e k we are essentially saying set alpha 1 equals 1. Then we have already gone through the explanation and arguments as to why we chose h of 0 to be 1.

In other words now instead of two parameters we have only one parameter; that is how it appears to be. However, there are two unknowns. What are the two unknowns?

Student: Sigma square e.

Sigma square e; so there is no escape it just manifests in a different form. That is something that you will see routinely in data analysis, there is no escape to things. Whatever were the numbers of unknowns had the beginning unless you have pumped in some prior information the number of unknowns does not change. So, this is your first order autoregressive model. The name for regression is auto regression is fairly straight forward. The series regresses on to itself. And regression is a term that is more widely used in hypnotherapy and so on, where people are regressed into the past to resolve some problems that have been troubling them and so on. So, it is a more of a term borrowed from sciences.

So, the auto regressive model has to be understood carefully. It says that the process evolves as a function of it is immediate past, but that does not mean that there is no connection with it is further past, it is but that is more of an indirect or a propagated effect. Directly this AR 1 model is dependent only on it is past. What is the prediction view point and that will set the interpretation of an autoregressive model very clearly. From the prediction view point the AR 1 models says- so long as you give me v k minus 1 I can make the best forecast, I do not need v k minus 2 or v k minus 3 and so on. I just need the immediate past.

In other words any prediction for an AR 1 process, any prediction of v k of v at k given k minus 1 k; let me explicitly write here v k minus 1 v k minus 2 and so on given the infinite past is same as just relying on the immediate past. Just v k minus 1 alone is sufficient; the rest is not needed at all. This is completely contrasts into the MA model. If you recall when we wrote the prediction equation for MA model we had to recover e k and we said- boy theoretically recovering e k requires infinite past.

So, in the sense an MA 1 model is an autoregressive model that actually request infinite past. And we have talked about this equivalence also and we will talk about it bit a later also. So, the AR 1 model is such that the conditional expectation of v k given the infinite past is the same as the conditional expectation given k minus 1. In fact, many a times this is called a Markov process, but I do not want to scare people I myself scared of that term. So, I will not use Markov process here, but there is a huge literature on Markov processes. And AR 1 process is more or less a Markov process. That is how the Markov process is defined as.

Now, you can extended this definition an AR 1 process to an AR p process; straight away you can extend this definition, but let us look at this from a parameterization view point. What is the parameter, what would be the parameterization of the impulse response for an AR p process? How would you modify this parameterization? What kind of parameterization will lead to AR p?

Student: Sum (Refer Time: 22:45).

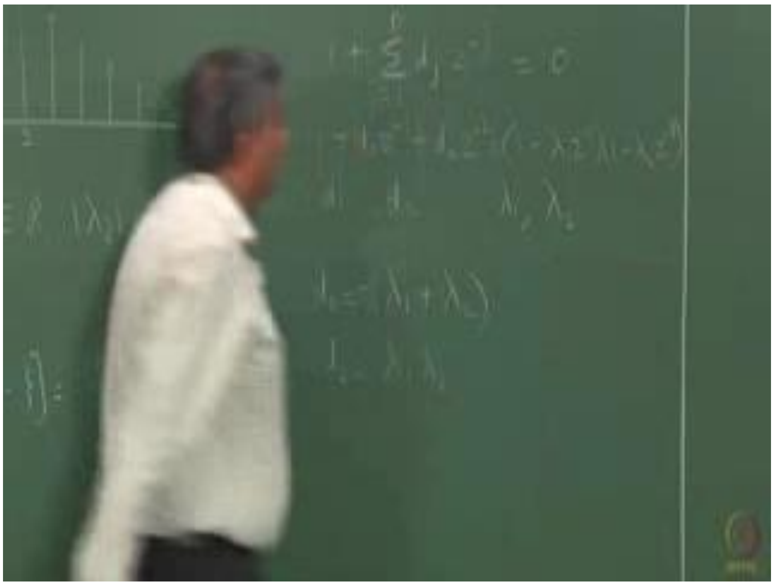
Sum of exponentials very straight forward; so you can say here see instead of alpha 1 lambda 1 rise to n sorry, you say alpha i lambda i rise to n. And, still here of course how many if you are looking at a AR p model you are looking at p terms here. So, the now instead of estimating h we are estimating alphas and lambdas that is the point. And here we assume that each lambda i is less than 1 in magnitude, in order to meet this requirement here the absolute convergence of the impulse response coefficients; clear, cool.

So obviously, you can see that in the AR 1 case lambda 1 is nothing but the root of the characteristic equation. So now you can expect for AR p the lambdas to be the roots of the p roots of the characteristic equation. And we know what the characteristic equation. And again I would like to say this that the mapping between the parameterization and the difference equation form is unique. If you assumed a particular parameterization you are implying a difference equation form of that order and vice versa. So, as we just mentioned here on the board, as we wrote on a board; the parameterization of the sum of exponentials form amounts to writing an AR p model.

Again as I said earlier you will not find this in fact I am not found this in any of the time series text as a connection between the AR form and the parameterization of impulse response from. It is purely presented as an autoregressive form, but this parameterization allows us to view the autoregressive model as a child of the convolution model that we have written. And of course, the consequence is that in on the stationarity requirements are now different. With MA model it was pretty obvious in the sense the stationarity was not an issues at all just apart from the boundedness of the impulse response coefficients, but invertibility was an issue because we had to guaranty stable forecast.

On the other hand in autoregressive models stationarity as to be guaranteed; that means, not all difference equation forms of the form that you can see on the screen are necessarily going to give rise to stationary v k. We are deviated from the linear random process model, so now we have to give a fresh set of constraints restrictions on the coefficients and those constraints are that the roots of the characteristic equation here; if you imagine in terms of the z inverse they have to be outside the unit circle or in terms of the z they are inside the unit circle.

(Refer Slide Time: 25:58)



So, for example for an AR p process the roots of the now characteristic equation in order to write would be one plus sigma d j z to the minus j; I am just using this as a different dummy variable here instead of i, I am using j, j running from 1 to p. So, this is our characteristic equation. Lambdas are the root of the characteristic equations in terms of z. They have to be the inside the unit circle, but if you think of roots in terms of z inverse they have to be outside the unit circle, in any case you will have p roots.

So, one has to be careful when on reads the conditions of stationarity in the different text books. And there is a proof given for stationarity; that is when this stationarity is guaranteed if and only if that the characteristic equation, the roots of the characteristic equation is outside or inside the unit circle accordingly. In for example, the book by Shumway Stoffer, but I feel that the proof is now obviated it is not needed, I have already stated right from the we have already stated from the beginning that h is of this form and then lambda such that they are less than 1 in magnitude and I will chose these such that lambdas that condition is satisfied.

The only issue is it is not as easy as it was to derive the difference equation form for AR 2, AR 3 and so on. For AR 1 it was pretty straight forward, you just had to write v k equations for v at k and k minus 1 and do a simple algebraic jugglery. But now for AR 2 you would have to write at k k minus 1 k minus 2, think of some kind of jugglery and then come up with your difference equation form. In fact, for the AR 1 for example, what would be the relation between d 1 and d 2 and lambda 1 and lambda 2? In AR 1 case it was straight forward d 1 minus lambda 1, what would be the relation? You are already given that lambdas are the roots of the characteristic equation.

Student: (Refer Time: 28:25).

Sorry.

Student: (Refer Time: 28:30).

D 1 is lambda 1 plus lambda 2 there or minus ok.

Student: (Refer Time: 28:39).

D 2 would be; is it correct? All you have to do is imagine the what is the characteristic equation in terms of d 1 and d 2 it would be 1 plus d 1 z inverse plus d 2 z inverse square and this you will have to write as 1 minus lambda 1 z inverse times 1 minus lambda 2 z inverse; sorry. Now you clear. In fact, you should quickly verify that it specializes to AR 1 when you throw away lambda 2; when lambda 2 is 0 it should simplify it to AR 1. Does it? It does. So, that is the simple check. Any questions, fine. A simple quadratic stuff.

So, that is the connection; that is the mapping between your lambdas and di's. And we have converted the condition on lambdas to a condition on d. You are seeing that d should be such that the characteristic equation has the roots outside or inside the unit circle and so on.