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

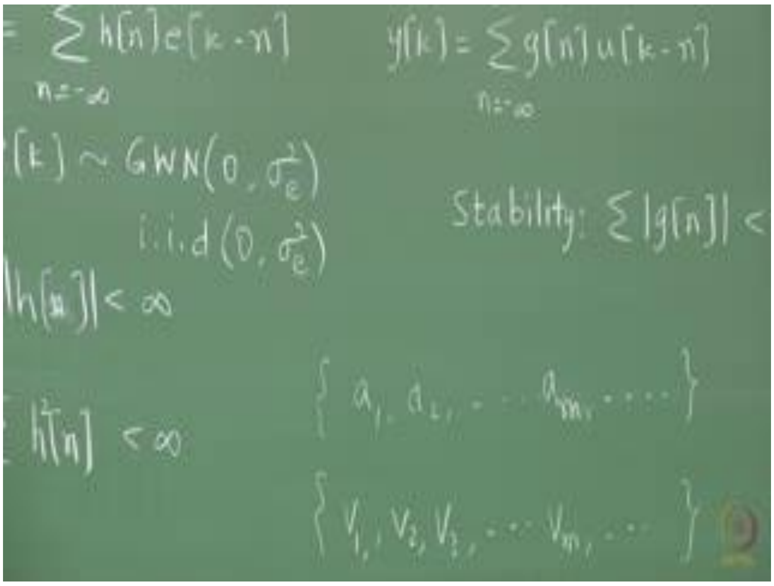
Lecture – 41
Lecture 18B - Models for Linear Stationary Processes-5

Now although I said we will talk about the theory later on, it is good to know up front what are the different forms of convergence that we are talking about. What we mean by convergence here is, imagine here instead of using infinite summation; remember we learned this theory of convergence of sequences of real numbers, how do we, what do we speak of there when we talk of convergence of sequence of real numbers what is it that we think of or how do we approach the convergence problem when I have a sequence of real numbers.

Student: (Refer Time: 00:50).

Is that the really the way that we do it? If you go back to the definition of convergence there are two ways of looking at it.

(Refer Slide Time: 01:06)



If I have for example a sequence let say now a 1, a 2 and so on, let us say I have bunch of real numbers here. This is infinitely long sequence; we would like to know if the sequence converges to something. As this index goes to infinity let me not confuse n here let me use m. So, as m goes to infinity what we are looking for is if this sequence converges to some fix number, let us say convergence of real numbers that we are looking at. The other way of talking of convergence is we take summations; for example, we sum up for example, that is when we suppose this a m is being generated by some numbers and then sorry; so you have this bunch of numbers here when I move on to some other number here the difference should be arbitrarily small.

That is another way of looking at it. That is if I consider up to m numbers and then move to another number let say I up to m 1 and then another bunch of numbers in the sequence up to m 2 at some point I should be able to show that the difference between those two numbers is arbitrarily small. So, a m 2 minus a m 1 you should be able to shrink-that distance between those two numbers to an arbitrarily small value. What is the situation here? Where do we see a sequence here? The way to look at the situation here is; imagine that we are not summing up the infinite terms here; we are actually summing up finite terms.

Let us say I start with 1 term and then 2 terms 3 terms in the summation and proceed, when I do this I am actually generating let say I denote here v 1 I am going to drop the k I have v 1 which is generated by only a single term in the summation, and then I have v 2. That is how I study whether the summation actually yields random variable or not. Slowly I keep adding numbers, and I want to be assured that the inclusion of more and more terms in the summation will not blow up v k. In fact, more importantly here should result in a random variable that is very important. Here, the only requirement in deterministic world is as I keep on adding more and more terms how in whichever way I am generating a 1, a 1 is being generated by some m equals m equals 1; a 2 is being generated let us say by including 2 terms by 3 terms and so on.

Here the only concern is that it should reach some fixed value, but in the convergence of random variables there is an additional concern which is that this should not converge to fix number; if it can that is ok, even a fix number is a random variable with variance 0. But the more important requirement is that it should converge to a random variable defined on some sample space. That is the important requirement. At this stage it is hard to think of this because we must have not encounted this convergence of random variables in any other course, but since we want to move on with the modelling I am not discussing the theory here.

So, coming back to the point here v 1, v 2, v 3 and so on; so I keep adding more and more terms in the summation. And what I want to be assured is that as I add more and more and more terms as m goes to infinity; that means, this summation comes up I should be assured that this v converges to a random variable. It is very difficult to imagine because how do we talk of converging to a random variables because random variables is not fixed. So the notion of convergence that we have being used to really prevents us from thinking further, but now you have to expand your thinking and say well at in the end this summation should generate random variable on a sample space where I can have probability is defined.

In other words this need not actually generate fix number; what we mean by fix number is suppose I have one realization of e, e generates one value of v k; I feed in another realization of we remember this e k is a random process. I feed one realization to the summation I get one value of v at k-th instant, I feed another realization of v I get another value of v at k and so on. For all realizations, now I repeat this for all realizations what would I get at the k-th instant I would get a bunch of numbers, but all at the k-the instant these bunch of numbers that we have should be the outcome of some random phenomena. That is what we mean by convergence of this sequence to a random variable. Which is quite different from the convergence of deterministic sequences, you understand. That is what we mean by convergence in I mean that is another way of looking at, it if helps you understand.

So let me tell you again here, as I feed in different realizations of white noise I will generate different values of v. For all possible realizations of e, I will generate all possible values of v they should constitute a random variable space. I mean in the sense on a sample space with probability measure defined on it properly. Now, whether we talked of convergence of random variables or real numbers are deterministic variables numbers. We are talking of some closeness, some distance between target variable whether it is a deterministic or random variable and the number in the sequence. So, we are essentially talking about the distance and are we getting closer and closer and closer to this fixed value eventually there should be some closeness. For the deterministic case the closeness should be closed, I mean it should be 0 as m goes to infinity.

In the case of random variables we are asking if this v m here; let me erase this sequence here converges to some random variable v. Now, let me give you a simple example of a result that we have already seen. We recall a central limit theorem right, what do we discuss in the central limit theorem right, convergences of what. So, there also we are talking of convergence it is not that this is the first time you are encountering.

There we asked a question when I add a bunch of random variables in a linear sense, I do get or another random variable and we want to know as I add more and more random variables what happens to the distribution of this random variable. There is an implicit assumption that I am generating random variable there, it is already understood. So, the question that we are asking is a weaker one, not a very strong one. We have already assumed that as I add a bunch of random variables I will generate a random variable.

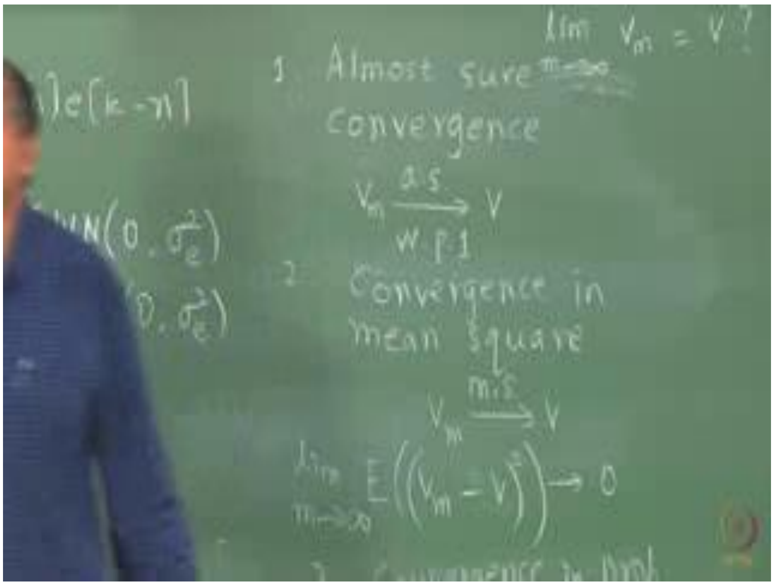
Fine, given that I will generate a bunch of random variables the central limit theorem is trying to tell you what happens to the distribution of the random variable. Here we are not even talking of the distribution. Of course, we know if e k is Gaussian then what happens to the distribution of v. That is also Gaussian so we do not worry about that. We in fact central limit theorem now tells us if e k is not even if it is not Gaussian I can be guaranteed that if the summation is fairly long involves large number of terms and v k is going to be Gaussian.

So coming back to the point, even in the central limit theorem we are talking of convergence, but that convergence that we are talking about is convergence in distribution. There are different forms of convergence that one can talk of when it comes to random variables; the weakest being convergence in distribution. All we are saying is in the central limit theorem as I add more and more terms what happens to the distribution of v m, how does it look like; we are not asking anything more. In fact, it turns out that is the weakest question one of them there are different forms of convergence, but among the four different types of convergence that one talks about in random variables.

The convergence in distribution is one of the weakest things that you can study. Just because distributions of two random variables are identical it does not mean that they are their outcomes are also necessarily identical. We are just talking of probabilities right. Probability of some event when you have a sample space and you are talking of probability of events on that sample space if that is identical to some other sample space at the probability of events in some other samples if we it does not mean that the outcomes are identical. So, that is why it is considered actually weakest form of convergence.

Then there are three other forms of convergence, and I will just name them and I will talk about what this which one does this guarantee and which one does this condition guarantee and then we will move.

(Refer Slide Time: 11:36)



So, the first form of convergence which is considered the strongest among all is the almost sure convergence. I will erase this here, suppose do they exceed that line. So, this is the first form of convergence that one and discusses in the theory of convergences of random variables, it is also known as convergence with probability 1. You may wonder if it is convergence with probability 1 then why cannot I say yes it surely converges; any idea. This is also sometimes; so the way it as written as is v m convergence to v with a dot s dot on the arrow there indicating almost sure convergence or sometimes it is written as with probability 1.

If it is with probability 1 then why do not you say it surely convergence? Why do not we say that, why are we saying almost sure. That is the beauty of math they keep dissecting to the core and they imaging many many different kinds of situations. Now although I am not going to go in great detail at this point the clue to that answer is there are events that can occur with 0 probability. Generally, what is the thinking that we are tuned to; if probability of some event is 0 you do not occur. Unfortunately that is not necessarily true you can have events with 0 probability.

Think about this, if your sample space is continuum if I ask-the question what is the probability; suppose you take this board here and I ask what is the probability that they chock particle is exactly at some location, what is the probability? 0. But it can be right the chock particle can be at a particular location exactly. My the probability measure is such that it gives me a 0, but that does not preclude the event from happening, the chock exactly being at that location it is got to do with them measures that we have used for defining probabilities. We may say that the probability that the chock is a exactly at some location is 0 that is fine, it only reflects my in ability to predict exactly that the chock location, chock particle can be at that location.

But does it mean that a chock particle cannot be exactly at one location? It does not; it can be it is only that if you ask me precisely whether the chock particle, what is the probability that the chock particle is going to be exactly at this location, I cannot tell you; I will say no that I have no idea it in fact we say no that cannot happen because not because the event cannot occur it is simply. The probability measure is such that which means that intuitively you can have, I have just given you a very intuitive explanation you can formally prove that there are events that occur, but with 0 probability. So, this is one of the paradoxes in probability theory.

And because of that we say almost sure; that means there may be a chance that that v m does not converge to v. Anyway, so we will talk more about it later on. The next convergence is the convergence in the mean square sense. In this case we write here v m converges to v in the mean square sense, so instead of a s we write m s. What does this guarantee? All of this, sorry each of this convergence criteria is talking about in what way v m gets close to v. The first one is strongest it becomes a random variable. In the second case it is same that the it is talking about the distance between v m and v in an Euclidian sense, but in a statistical sense this goes to 0 as m goes to infinity. That means, v m converges to v in mean square sense the mean square vanishes.

And the third form of convergence is the convergence in probability. Do not confuse the third one with the first one. This is actually the simplest to understand, in fact this weakest once are the easier to follow. For example when we talked of converges in distribution fairly easy you know no body worries about convergence when we talk of central limit theorem- yeah I can understand what happens to the distribution of y I can feel, I have a good feel of that kind of convergence.

The fourth one; we have not listed but as I have said fourth one is convergence in distribution which we will use extensively in estimation theory. In fact, we will use all forms of convergence in estimation theory. So, may be some reading up front will help you. So, what is convergence in probability? It is not the same as the; do not get confuse with this the convergence in probability says that the probability of finding that; that is v m let me put it this way. The v m which is your some variable in this sequence random variable in the sequence converges to a random variable within an epsilon radius in the probability sense. So, let me write this here what this means maybe I can write- I will write here itself. The probability of v m being within an epsilon distance or an epsilon radius of v or outside you can say even outside you can state it different in two different ways; the probability of finding v m outside an epsilon radius of v goes to 0 as m goes to infinity.

So, it is only a very probabilistic thing and epsilon is arbitrary small, but not 0. So, what this means is v m does not have to exactly hit v. Think of the simple case, think of this v that we are talking of which is a random variable as a constant; even a constant can be thought of as a random variable that is a degenerate case and which is what we will discuss in estimation theory where in estimation theory we apply this to parameter estimates as we for example, if you take sample mean how do you estimate sample mean you keep you take a bunch of observations take the simple average. And the question that we ask-there for example is as I add more and more observations will the sample mean converge to the true mean. There the convergence that we are talking of is not to a random variable but to a fixed value, but fixed value can also be thought of as random variable with radian 0; that is not an issue.

So, think of v for now as a constant as a fixed value, so what we are asking is as I include more and more terms in the generation of v m where does it hit the true value. And the third condition says well it does not necessarily hit we do not know it may hit, but what is guaranteed is that it will be the probability of finding that v m outside and epsilon radius goes to 0, which means in the end as you in increase the number of terms in the summation for v m there is no guarantee that it will hit v at a finite for finite terms. May be you need to include finite infinite terms, may be with finite terms it will happen. So, it there is guaranteed there, that it will hit v exactly after finite terms; yes.

Student: (Refer Time: 20:36) when we are talking about convergence of probability the (Refer Time: 20:39).

Why?

Student: Because when you are talking about the probabilities of v m taking on particular value that cannot be done continuously very (Refer Time: 20:50) if I have a real from 0 to 1 then I cannot talk of that the number we get exactly, but if I have some or say integers.

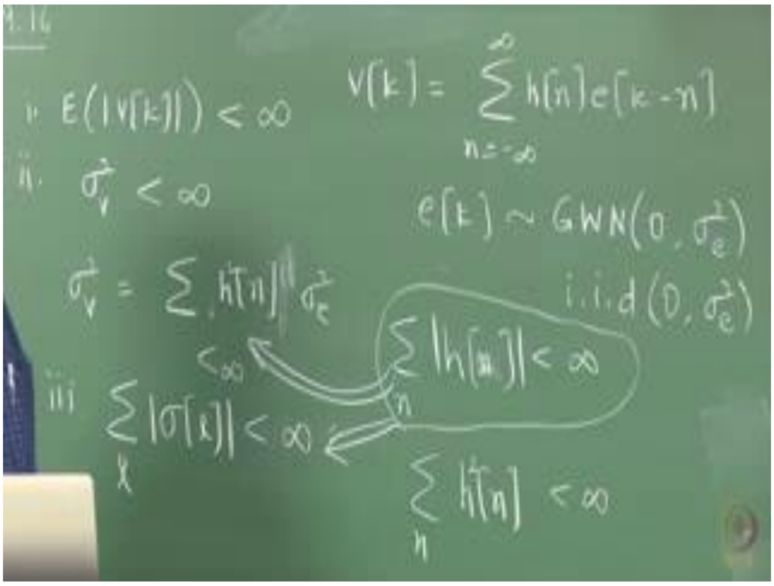
So we are not talking of that at all, we are not talking of probability of v m taking on some value. So, do not get confused with between that and this. We are talking of where does v m go and sit. And here we are saying v m goes and sits anywhere within the epsilon radius of e, so there is some randomness there. But your question is also good one, you are asking how valued is a question of v m going and sitting at some value right? Well, it need not be discrete value, all we are saying is as we add more and more and more numbers then as I average them the average can go and sit not the individual; that is in that case in the case of sample mean v m is a sample mean it is a random variable and if you are talking of whether the sample mean converges to the truth yes you are asking whether it loses it is randomness after some m.

That is what, in fact we are we mean by convergence essentially, but there are different ways of asking the same question that is the point here. We are asking in general when we apply this to estimation theory where we are concerned with parameter estimates reaching fixed known true not known, but some true parameter value which is fixed. We are asking if it loses it is randomness it becomes degenerate at some point as I add more and more terms; which is fair it is to ask if random variable in a sequence cease us to be random, well in only in this sense therefore we are not asking. For example it would be in correct to ask if limit m goes to in the limit as m goes to infinity if this is possible we are not asking in this question at all.

This question would be invalid. That question that we are asking would be invalid because then you are assuming that v m ceases to be completely run. That is why the definitions are quite delicate, we are still living in the probabilistic sense all we are saying here in the third one is v m goes and sits, what I can guarantees v m goes and sits somewhere within the epsilon radius of v. That epsilon is not I can guarantee that if it. So, happens that if hits v so v, but I cannot guarantee with that. Here also you see we are very careful, here also we are saying of course I am not the limit here, but the moment we say almost sure it is understand that there is a limit there.

There also we are being a bit careful, so no where we are guaranteeing that v m will exactly go and sit at v there is a possibility that it cannot, but this is a much stronger guarantee then this one. In fact, the almost sure convergence implies this and then the mean square convergence implies conveyance convergence in probability. We will talk more about this later on when we again revisit this definitions in the context of estimation theory, but even if you do not follow the full details of this what you should understand here is that this summation that you have here in the definition of linear random process is not domestic animal. It is a wild animal, we have to learn to understand it and that there are requirements it just does not necessarily guarantee that for example, it will generate a random variable all the time and that it will converge to stationary sequence all the time and so on; and then of course that this is guaranteed and so on.

(Refer Slide Time: 25:00)



It turns that with this restriction everything is taken care off. If this restriction you guarantee all the three including almost sure convergence. And with this restriction you can guarantee that the variance is finite and the summation converges in the mean square sense which is not bad, many a times we try to prove the weakest run if not the weak- then the next stronger and then the strongest.

So, just to summarize in minute; here this summation that we have for that appearing in the definition of linear random process converges to a stationary random process if this is guaranteed. And this is a slightly weaker requirement, but we will stick to this one because this is a much stronger requirement this guarantees all the things that we require.

Now again you should keep in mind that we are working with stationary processes; what we mean by stationary here is second order stationarity. And secondly, that we are dealing with you are assuming that the process as a spectral density and that the spectral density satisfy some conditions and for optimality that the process is a Gaussian one. When we say Gaussian process the joint distribution of any set of observations all was a multivariate Gaussian distribution. So, under these assumptions in this frame work we are building our time series model.