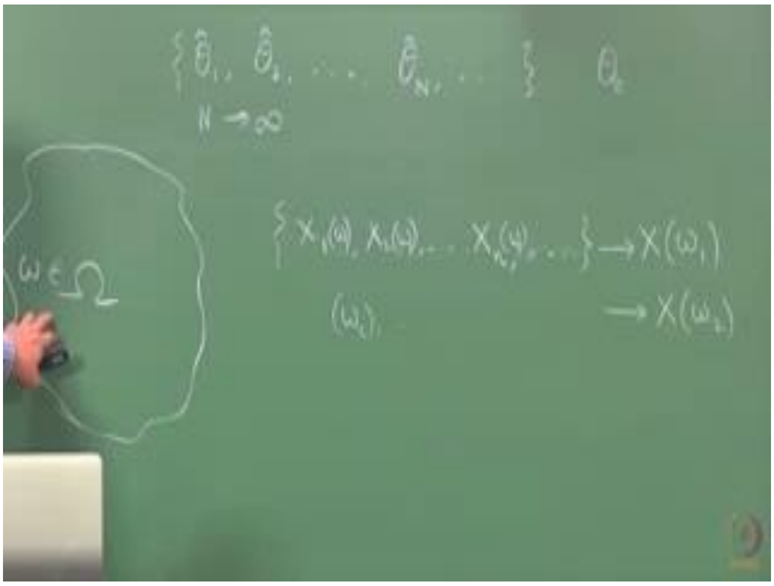
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 89
Lecture 39A - Goodness of Estimators 2 -4

Good morning, what we will do today is we will continue with our discussion on consistency and also hopefully wind up the discussion on properties of estimators. So, just to give you a quick recap of what we were discussing yesterday is we are looking at these properties of estimators and towards the end we started talking about consistency: which is a very important property.

(Refer Slide Time: 00:31)



Essentially this consistency is all about how this sequence of estimates that we construct for different sample sizes, how they behave as N goes to infinity, where do they go? Ideally we want this sequence to converge to the truth and the truth is denoted by theta naught.

Without knowing the value of truth we are able to establish and we should be able to establish the convergence of this sequence and knowing very well that this theta hat is actually a random variable, we had to turn to the theory of a convergence of sequences of random variables and essentially I said that there are 3 forms of convergence of course, there is a 4th one, which I discussed the first which is a point wise convergence. So, in total you have at least 4 forms of convergence and then a 5th form of convergence is convergence and distribution which we will talk about today as well.

The first one that we talked about is the point wise convergence and point wise convergence to understand point wise convergence or any form of convergence all you have to ask is now any form of convergence essentially is asking how close or whether this sequence converges to theta naught. If it reaches hits exactly theta naught it is great, but if it is within the vicinity of theta naught also that is which is a weaker form of convergence.

Regardless of the form of convergence what we are actually looking at is for different possible sequences. What we mean by different possible sequences? Imagine that this is your experimental space and denoted by big omega that is every point in omega corresponds to an experiment let us say. In general we say that this is a sample space of all possibilities that is generating some sequence of random variables X 1, X 2 and so on of course, we have used lower case n in the theory on convergence, but you should not get confused both of both mean the same to us here. So, each point in this sample space generates one sequence. So, you can therefore, index this by some omega, unfortunately this omega notation coincides with frequency, but then you have to observe a distinction here I am just using the conventional notation.

So, for every point in the sample space a sequence is generated and convergence is all about where the sequences are going that is where this rivers are heading are they all heading to an ocean or some rivers go some other ocean and some other rivers go to another ocean or sea and that is essentially the convergence. The point wise convergence demands that for every point for every possible realization, whatever sequence that you have all of them should each remember for a fixed value of omega, this sequence is only a bunch of numbers, any sequence that we have is a bunch of numbers.

But the only difference is that there is a controlling factor omega which will determine what sequence of numbers you are working with. So, every sequence of numbers is going to converge to some number hopefully, if it diverges then that is it you can dismiss all forms of convergence; provided that each sequence converges to a number what point wise convergence demands is that all each of the sequence when they converge to a number, the collection of those converged values should be the possible values for a random variable that is all.

So, if you think that omega 1 is one realization, let say this converges to some converged value here right and likewise if I were to generate the sequence for another possible realization then assumes that this converges to x of omega 2 and so on. This way I have a set of converged values for every possible trigger here there is a switch here and point wise convergence demands that every point for every trigger, that the sequence generated by every trigger should converge to first of all some finite number and the collection of those converged values should be the possible values for a random variable.

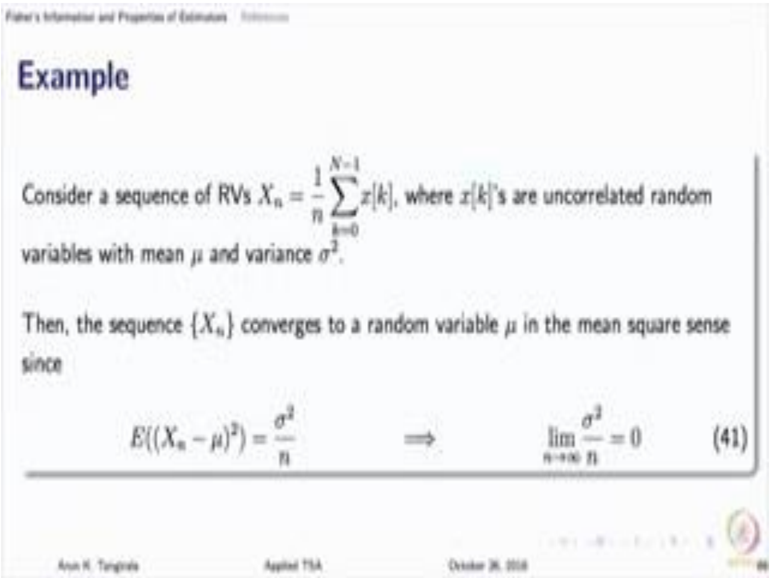
So, that is a very very strong requirement, convergence in probability is a much weaker requirement. What it says is essentially that it looks at the definition in the probabilistic sense, it says that all this converged values yes they do converge, but they are within the vicinity of a random variable, within an epsilon disc of a random variable and the probability of finding what the convergence and probability says is there is a possibility that a few or more of this converged values do not fall within the possible values for a random variable but that probability keeps shrinking as the sample size keeps increasing.

So, the probability of finding the converged values outside an epsilon radius of a random variable keeps going down, eventually it will converge. So, we do not know when it will converge; that is what the probabilistic statement says. The mean square convergence looks at in a different metric it does not look at probability, it says that distance between these converged values and the random variable, those that distance that is in a Euclidian squared distance sense, that keeps shrinking to 0 as the sample size goes to infinity.

Now, that is a stronger statement because as the distance goes to 0 you are confirming that it will hit the some random variable that is what we want. So, mean square convergence is a more strong statement then convergence and probability. Typically you will see that in many estimation exercises, we are looking at mean square convergence, but we may actually demand a more stronger form of convergence which is called almost sure convergence and we will talk about that, but before we do that let me give you an example of mean square convergence, after all of this I will show you how to visualize this concepts in R, I will show you some and then the rest you can do on your own, it turns out that there is a beautiful package called in R convergence concepts you just have to install those and there is a nice tutorial document surrounding that the purpose of the package is only to help you understand this different forms of convergence with nice simulations.

I will show you how you can do manually then you can install the package and check how you can visualize those concepts using the package. So, let us talk about this example here which is concerned with mean square convergence, again we turn to the example of a sample mean assume that your computing sample mean.

(Refer Slide Time: 08:21)

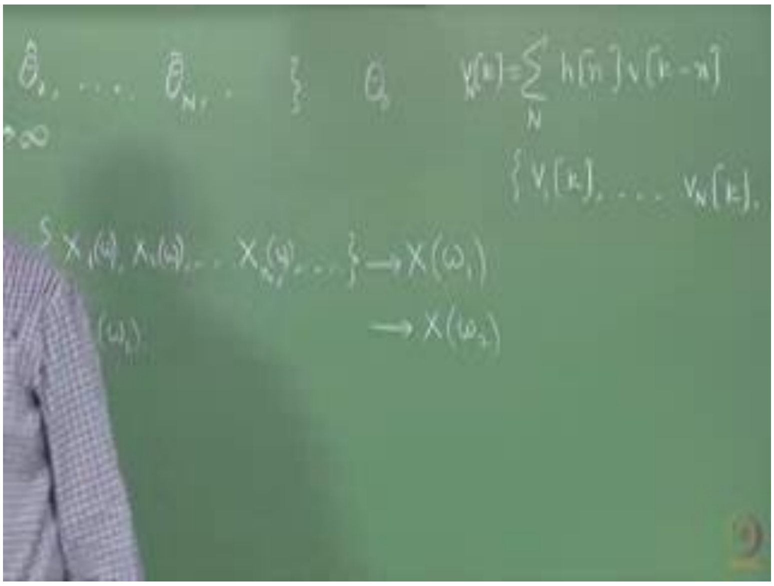


From observations that are uncorrelated and we know that the variance of this sample mean is sigma square over n right and what mean square convergence is essentially looking at is whether in fact, for this problem the question in hand is whether the sample mean converges to the true mean mu?

Here we have talked about convergence to random variables, but as I mentioned yesterday a constant is also a random variable and the reason for talking about this convergence of random sequences to random variables is to also help you understand the other part where convergence is required, which is where we are looking at linear random processes, where we have required that the summation should converge to a random variable. So, this theory that we are talking about helps you understand concepts, the concept of convergence of a linear random process the model that we have and the concept of convergence of parameter estimates.

In the case of parameter estimates we want the sequence of these parameter estimates to converge to a fixed value whereas, in the case of linear random process we want this summation that we have for the linear random process if you recall we had the summation.

(Refer Slide Time: 09:40)



This summation should actually converge to a random variable as I keep increasing the number of terms right? Imagine that here you are summing up only some N terms and you can now construct. So, this is your v k denote this subscript N denotes the number of terms you have used in the summation, imagine generating a sequence of a random variables there we demanded that this sequence should converge to a random variable.

So, the he here also you have a sequence here also you have a sequence, but the only require difference is we demand that this sequence converge to a random variable whereas, we demand that this sequence converge to a fixed value and this problem pertains to parameter estimation, we want the sample mean sequence that I am constructing from different sample sizes to converge to the truth mu and mean square convergence requires that the mean square error, expectation of x n minus mu to the whole square is nothing but the x n bar sorry it should not be x n it should be x n bar, they or if I have defined x n this big x n itself as what I have done then it is correct.

So, whatever random variable which we have denoted we have used to denote the sample mean, the mean square error of that should go to 0; we know from our prior derivations that this mean square error is sigma square over n and as n goes to infinity the mean square error goes to 0. What this means is that the sample mean does converge to the truth in a mean square error sense? If you are still confused all we are talking about a why we is different ways in which the sequence of parameter estimates go and reach the truth, some are strong some are weak; the converge and probability is a weak weaker requirement, convergence in mean square sense is a stronger requirement and convergence in the sense of almost sure convergence is the strongest among the three.

We have earlier discussed point wise convergence right, almost sure convergence is no different from point wise convergence except in one aspect and that it says it recall first point wise convergence. Point wise convergence demands that for every possible trigger here, whatever sequence that you have all sequences should converge and they should converge in such a way that you should be able to define a random variable with those outcomes.

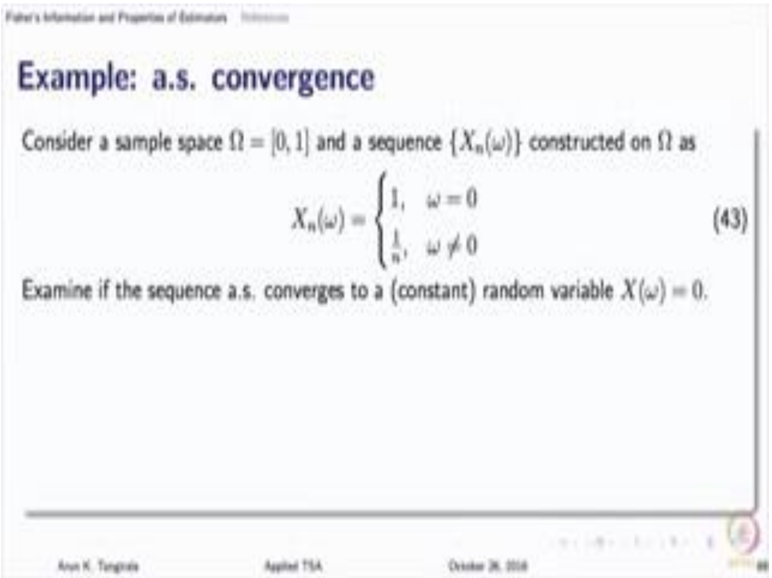
But that is a very strong requirement almost sure convergence relaxes that requirement a bit and says it is for some sequences not to converge to a random variable, but it they should be for those triggers here which are called 0 probability events. What are 0 probability events? So, if I ask you if imagine this to be a continuum, then if I ask you what is a probability that some random variable here corresponds to some possible trigger omega naught that is 0.

But from a measure theory because probability is a measure from a measure theory view point, the probability of the trigger being exactly equal to omega naught is zero; however, it does not mean it cannot occur; it can occur. So, this is the big irony and anomaly in probability, that is because we are working with measures and measures associated with the point are always zero. So, that is why we say 0 probability events which means there are events that can occur, but which have a probability of 0; un for fortunately this only occurs for a continuum continuous valued random variable, if you say the if you turn to discrete valued random variables and it turns out that the probability is zero; that means, the event cannot occur.

So, is this notion of 0 probability evens exists only when the random variable is continuous valued and for our parameter estimation sake this trigger here is nothing, but your experiment, that trigger that we are talking that is the sample space that is triggering your sequence of parameter estimates is simply an experiment. It says, all possible sequences the point wise convergence says, all possible sequences for corresponding to all possible experiments should converge and they should converge in such a way that they constitute the outcomes of a random variable and you can think maybe in the sense of if they are corresponding to experimental index. Then there is no issue of almost sure convergence, but if it is something else then we have to distinguish between almost sure convergence and point wise convergence.

Let me give you an example.

(Refer Slide Time: 14:58)



Suppose I my sample space is continuous valued, with 0 and bounded by 0 and 1 and both inclusive 0 and 1 are also included as you see from the square brackets. I suppose you know the distinction between using the square brackets and using the parenthesis, these are closed sets with 0 and 1 included and let us assume that a sequence is constructed as I we have given in 43 equation, equation 43 where the sequence assumes a value of 1 if the trigger takes a value of 0; if the trigger takes on any other value the sequence will take on a value of 1over n right.

Now, the question is if this sequence converges to a constant valued random variable, the constant being zero? So, I define I can define a random variable that is fixed also, only difference is that the variance of it is 0 that is. So, what do you think, how do you approach this problem? The first step is to see where do all possible sequences converge.

So, I start with omega 0, where to what value does a sequence converge? 1 very good for omega not equal to 0 to what value does it converge? 0.

Now, the question is whether this sequence now converges in an almost sure sense to this random variable that we have defined, what do you think yes or no why? Because there is one value one of the converged values for the sequence is 1, but it corresponds to what kind of triggers 0 probability events, because it corresponds to a specific value for that trigger therefore, we say that this sequence converges to a this random given random variable in an almost sure sense, almost sure meaning there are a few trigger points, there are a few realizations in the sample space, which do not generate the sequences that do not converge, but otherwise more or less it converge otherwise it converges.

So, all you have to do is in a given problem, if you have to determine almost sure convergence, you just have to see where those sequences are converging and you have to fig if you are given the random variable then you ask if it is converging to the random variable fine and you also have to see for what values of omega the sequence is not converging to the random variable and if those values of omega corresponds to 0 probability events then you can say that this sequence converges yes.

Student: (Refer Time: 17:54).

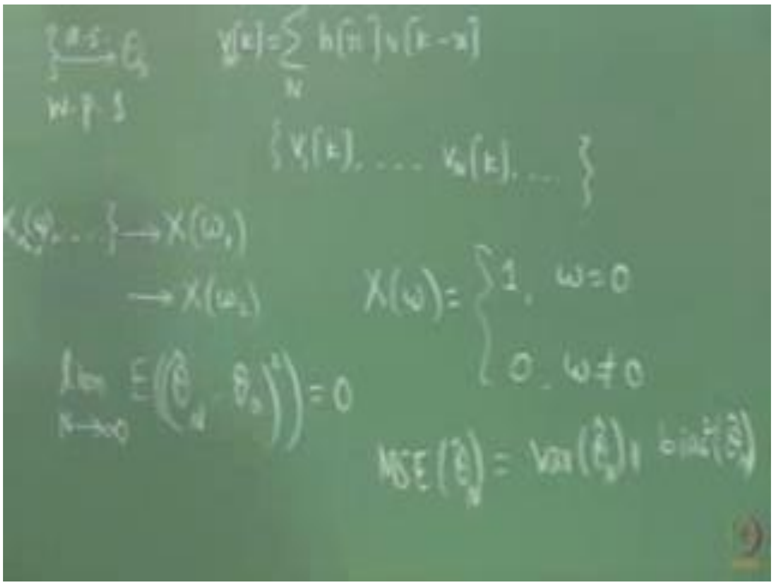
Yeah then for every point it has to converge, that is why I said when samples.

Student: (Refer Time: 18:01).

Sorry yeah that is why. So, in that case we do not talk of almost sure convergence you take in just assure point wise convergence. This almost sure convergence is introduced to exclude to consider these possibilities, to take care of this anomaly of 0 probability otherwise no, I mean this sequence is not point wise convergent right it is very clear, because there exists a point in omega space which generates the sequence that does not converge to x of omega.

On the other hand I can actually redefine a random variable, suppose I define a new random variable such that.

(Refer Slide Time: 18:42)



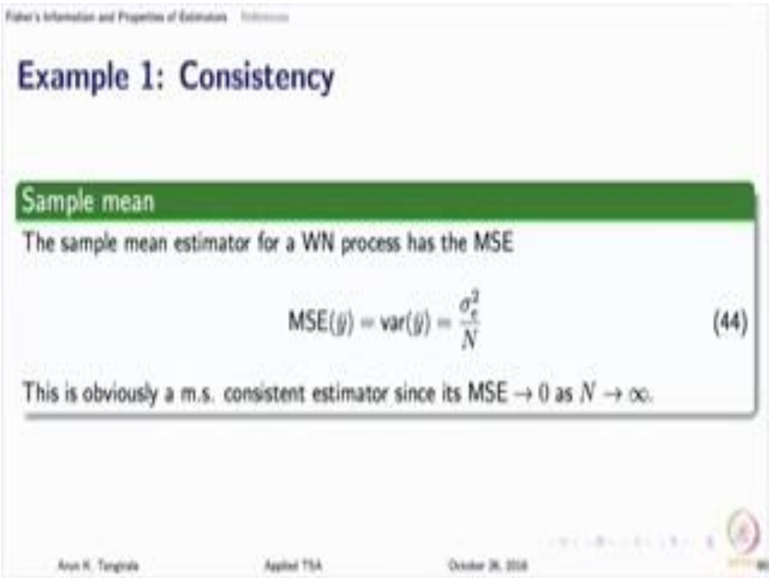
I say x of omega converges to 1 if omega is 0 and 0 now does the given sequence converge to this random variable in a point wise sense, it does. Almost sure convergence requires that if there exists omegas over which the sequence does not converge, those omegas should be 0 probability events, it does not demand that they should always exist omegas for which the sequence should not converge. So, for discrete valued omega for discrete space here the point wise convergence and the almost sure convergence exactly match, there it means that there should exist no omega; obviously, for which the sequence should not converge. So, it is all about the given random variable.

So, coming back to parameter estimates here, when we say that the sequence of parameters converge to theta naught almost surely sometimes this is written as with probability 1. So obviously, the Layman's question is if it is occurring with probability of one why do not you say definitely it converges that is because of this anomaly of 0 probability events, there are events that occur even if the probability is 0.

So, mathematicians and statisticians have thought very rigorously for our benefit so that we do not have to break our heads on it. So, when we say that a sequence of parameter estimates or an estimator produces parameter estimates such that it converges to the truth in an almost sure sense, then that is the strongest statement that you are making about the convergence. If this is guaranteed then you can show that mean square convergence is carry guaranteed which in turn guarantees convergence in probability. So, that is how the convergence are hierarchically addressed.

Generally we look at mean square error convergence, if that is not possible I mean; obviously we go in the order, if you cannot guarantee almost sure convergence you turn to see if mean square convergence can be guaranteed, consistency can be guaranteed, if mean square consistency cannot be guaranteed, at least there should be consistency in the probability sense. The sample mean converges in the mean square error sense.

(Refer Slide Time: 21:10)

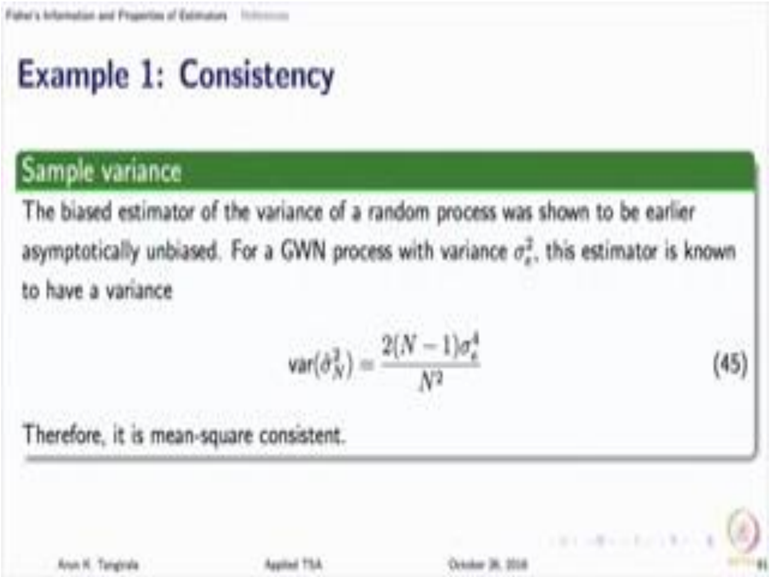


So, again the sample mean is an example that we have seen earlier, consistency as I have defined as I have said earlier depends on the definition of convergence, if you are looking at mean square error consistency then what you are asking is essentially if in the limit as a sample size becomes very large or goes to infinity, whether the mean square error goes to 0 and the sample mean does and we are able to prove this without knowing the value of theta naught.

How do you do it in practice for complicated estimators unfortunately theory cannot help you there, that is where you have to turn to simulations and that is why I am going to show you a sample simulation, so that tomorrow if you are working with a complicated estimator, a non-linear kind of estimator, it may not be possible for you to prove consistency by hand, because they are taking expectations of the resulting expressions are going to be very tough.

So, the second example that we are looking at for in the context of consistency is the variance estimator, the bias estimator of the variance sigma square hat and you can show that for a Gaussian white noise process with variance sigma square e.

(Refer Slide Time: 22:32)



The estimator this bias estimator as this variance given in the expression on the screen, 2 times N minus 1 times sigma square sorry sigma power 4 by n square. It is obviously, mean square consistent because as n goes to infinity this mean square error although I say here variance I am sorry one has to be careful this is not the mean square error.

What is the mean square error? Mean square error is given by the sum of variance plus the bias square right, this is mean square error we have learnt this yesterday. So, from the given expression how do you prove that it is mean square consistent? I had given a expression for the variance, but the statement that I am making is that it is mean square consistent; the mean square error should go to 0, it is not sufficient if the variance alone goes to 0 why is this is zero? So, this estimator itself bias is not zero.

What happens to the bias as n goes to infinity? The bias also vanishes since variance and bias both vanish therefore, the mean square error will also vanish and hence this estimator is mean square consistent. So, simply do not use the variance expression to prove mean square error consistency, because variance and mean square error are different when there is a bias. In a sample mean case the sample mean is an unbiased estimator therefore, we did not have to worry about the bias right that is the case with the sample mean whereas, there the variance estimator that we are looking at is a biased estimator, but this bias vanishes as n goes to infinity and therefore, the term 1 and term 2 both vanish as n go to infinity and therefore, you are guaranteed mean square error consistency.