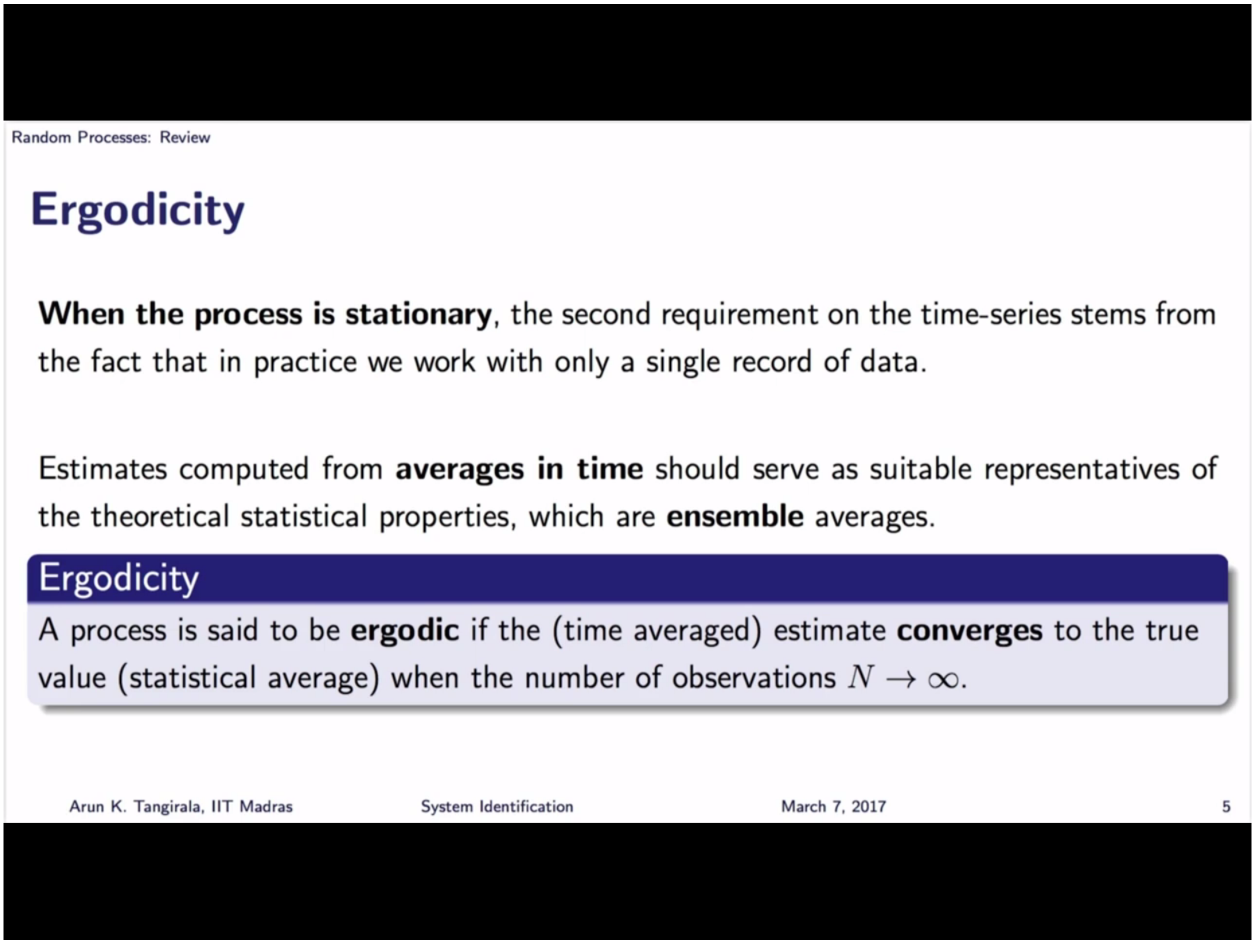
Now very often, not very often, in fact almost always, what we are going to do is, okay assume that the process is second order stationary, given to you, no problem. What are we going to do then. I am going to collect data and I am going to fit a model. But remember when I collect data, I am got to observe… I am going to work with a single realization, whereas this random process is a collection of all realizations, which I never get to see. And my model is suppose to be able to explain all the realizations, not in terms of values. It will not be able to predict accurately, that is the mark of a random process, but it is suppose to capture the characteristics of all realizations, whereas I have only a single realization. So I have to explicitly state another requirement, which is that I am going to work with the single realization, many observations, doesn't matter, may be 10000, million, doesn’t matter, but it's only a single realization. But I have to state this requirement, which basically says that as I collect more and more data and I am going to… that’s what I am going to do, I am going to walk in time and I am going to estimate may be this covariances for example, because I need to fit a model or variance or mean and so on. All these are theoretically, if you take sigma XY, Mu y, sigma square X, or Mu X, they are all not defined in time. Even though axis V at K1, what does it mean of the… what does it imply when I say Mu at K1, Mu of the random process at K1 would mean, I freeze K1, look at all the observations and then evaluate the mean in the outcome space at K1. Then I move to K2, Mu at K2 would be the average, statistical average of all possibilities at K2 and so on. So all this quantities Mu X or Mu Y sigma, XY sigma square X, they are not defined in time, they are defined with respect to the outcome space, but we are going to walk only in time. Ergodicity is this property that we assume to hold for the given process such that whatever averages that I am going to compute in time will serve a suitable representatives of averages that you compute along the outcome space, okay. So it is… I am just say… that is you have to state this, you have to say that this is the framework in which I am going to operate. You may… now the question that comes to mind is, is there a way to verify if a given process is ergodic. In fact you should have also asked this question earlier, is there a way to check if the process is stationary? We will come to that a bit later. Is it possible for a given process to say yes this is ergodic, unfortunately it's very difficult to verify. In… in a lot of situations you can say whether the process is ergodic or not, depending on how you are collecting your data. For example if you have a sensor bias. There are many examples that are given for… you know classic example is the most visited park or most visited theater. I want to figure out what is the most visited park in a given city. Ideally I should freeze time and look at all the parks, but maybe I don’t have access to that, maybe I have, may be with Google images and so on, we can… with satellite images I am able to get it, but let's say I don’t have access to such data. Then what am I suppose to do. I will follow a few… I will have a few individuals who are park going, I mean they love going to parks and I follow basically assuming that they don’t have any bias towards the particular park, I have randomly selected this individuals and I am going to note down, which parks they have visited. Over a period of time, because now I have to… what I am doing is essentially replacing this outcomes with time. And I hope that over a period of time, I have enough data and I look at the data in time and say yeah, this must be the most visited park, but there are two important assumptions, one the preferences of this randomly selected individuals do not change with time. If they change then that means they are non stationary. So this ergodicity property itself is valid when the process is stationary. You don't speak of ergodicity or all, when there is no stationarity. First of all I will look at things in time, only if I am assured that the underlying characteristics do not change with time. If they begin to change with time, I cannot rely on a temporarily. So ergodicity implicitly assumes stationarity. So that is an implicit assumption. The second thing is that I hope I have not selected an individual, who doesn't go to park at all. Then that means if I… suppose I have selected 5 individuals and 4 of them don't go to park, then I have a wrong representative, wrong sample, wrong sensing. The sampling scheme has gone for a toss. So that can lead to completely erroneous conclusions. So as you can see here, most of cases ergodicity is violated, if you have not sampled in a proper manner, if your sampling mechanism is at fault, okay. Of course there are theoretical conditions and so on, but… and we will not go into that, but what you should remember very well at the back of your mind is, when the process is stationary and when you sit to work with time domain data, which is what we are going to do in reality, we are making a big assumption of ergodicity.

(Refer Slide Time: 06:25)



So if everything fails, you have tried the best model, everything, then you may have to come and revisit this assumption, may be this assumption is violated, okay. So one can go deeper into ergodicity, some formal criteria can be given, but we will not get into that, that's not of interest to us in this course. So there are two pillars, two legs on which our models will stand, one is stationarity, second order stationarity and when that holds, ergodicity. This is... these are the two legs on which or the two pillars you can say, on which our models will stand. It's important to know this, whether we can verify the assumptions or not always in theory, it is... people don't worry whether I can verify it in practice, what is important is to state those assumptions first. Verification comes later on. Alright, so any questions on ergodicity? It's mostly, it's got to do with the way you are observing the process, rather than the process itself. Of course if you include the observation mechanism as a part of the process, then you can say it's the process itself?

(inaudible)

What do you mean by dynamic?

(inaudible) not at steady state and then…

Dynamic process is never going to be at steady state, right, it's values are going to keep changing with time.

But how do (inaudible) then at least for the verification, whether it's stationary or not?

So your concern is whether I can verify stationarity or not? Not about ergodicity. Again it is not easy to verify if a process is strictly stationary, but it is much-much better than our inability to verify the ergodicity property. So suppose I give you the temperature data, atmospheric temperature data over a day, you can look at it and say it is not stationary, because you can see visually that the local means are changing with time, but that is easy visually. There may… there are many other processes, whose means do not change with time, but variability changes with time, the earthquake data for example is a classic example of that or and… and many others signals as well. When it comes to checking for non-stationarities, non-stationarities are of many-many kinds. One has to first postulate that this is the kind of nonstationarity I am searching for, in the given signal and setup a hypothesis test and then you see… and go through the standard procedure to come to the conclusion, whether the null hypothesis is rejected or not. So there are different types of non-stationarities as… theoretically there are millions, but the ones that we are interested in is trend kind of non-stationarity, whether there is a deterministic trend, the mean is changing with time, whether variability is changing with time, that means the spread of outcomes is changing with time? Then there is something called integrating effect, the Brownian motion, random walk process, which also falls under the class of mean changing with time, but not in a deterministic way. So these are the three very popular types of non-stationarities that we normally check for. As I said there are other types of non-stationarities, but these are the three most commonly encountered ones. So when you are… and there are tests… statistical tests available, given a realization, how do you go about testing this hypothesis. So those tests are available, you can do that.

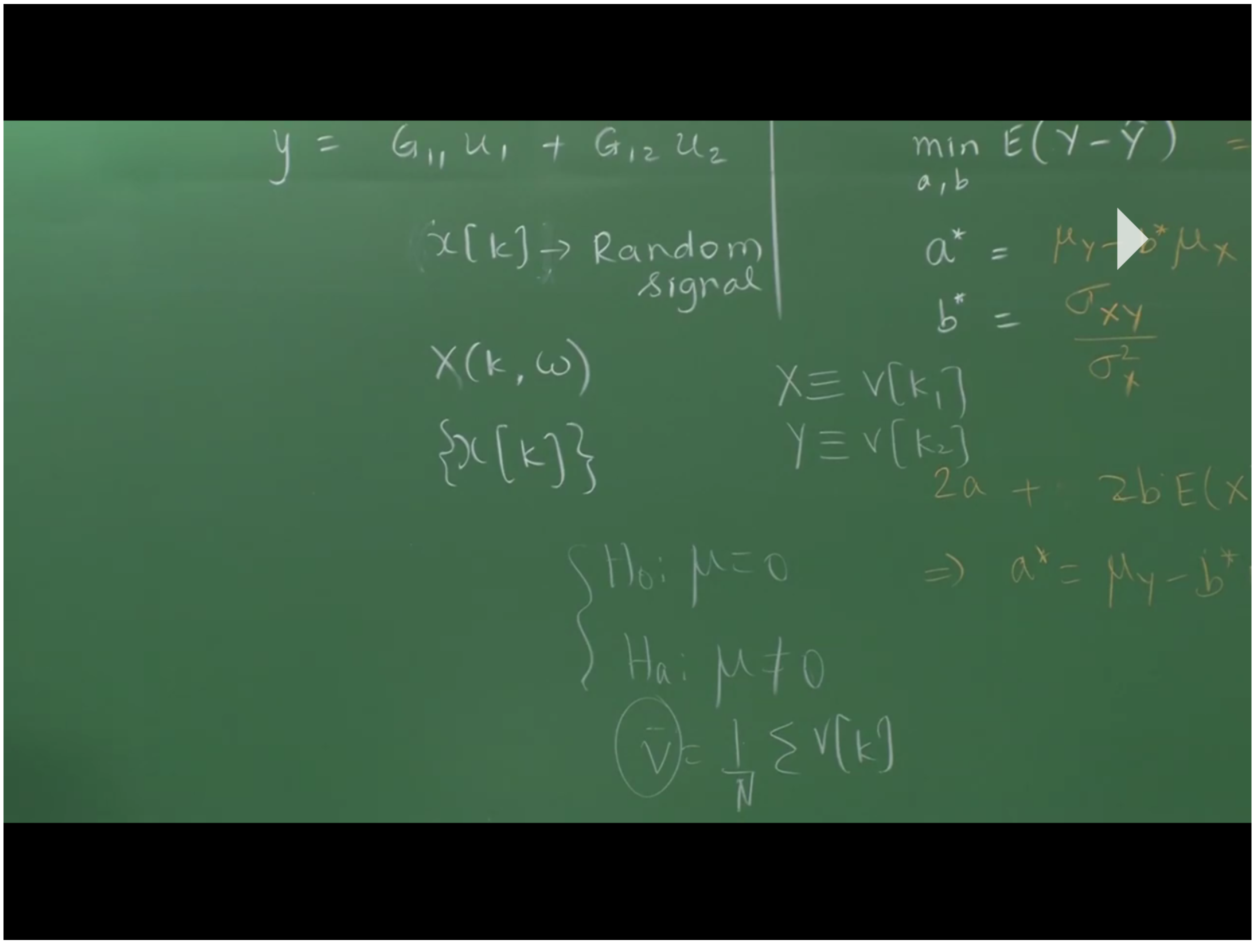
(inaudible) we get the mean by taking the time stamp… time series, we don’t have different realizations…

Correct

So how we can say?

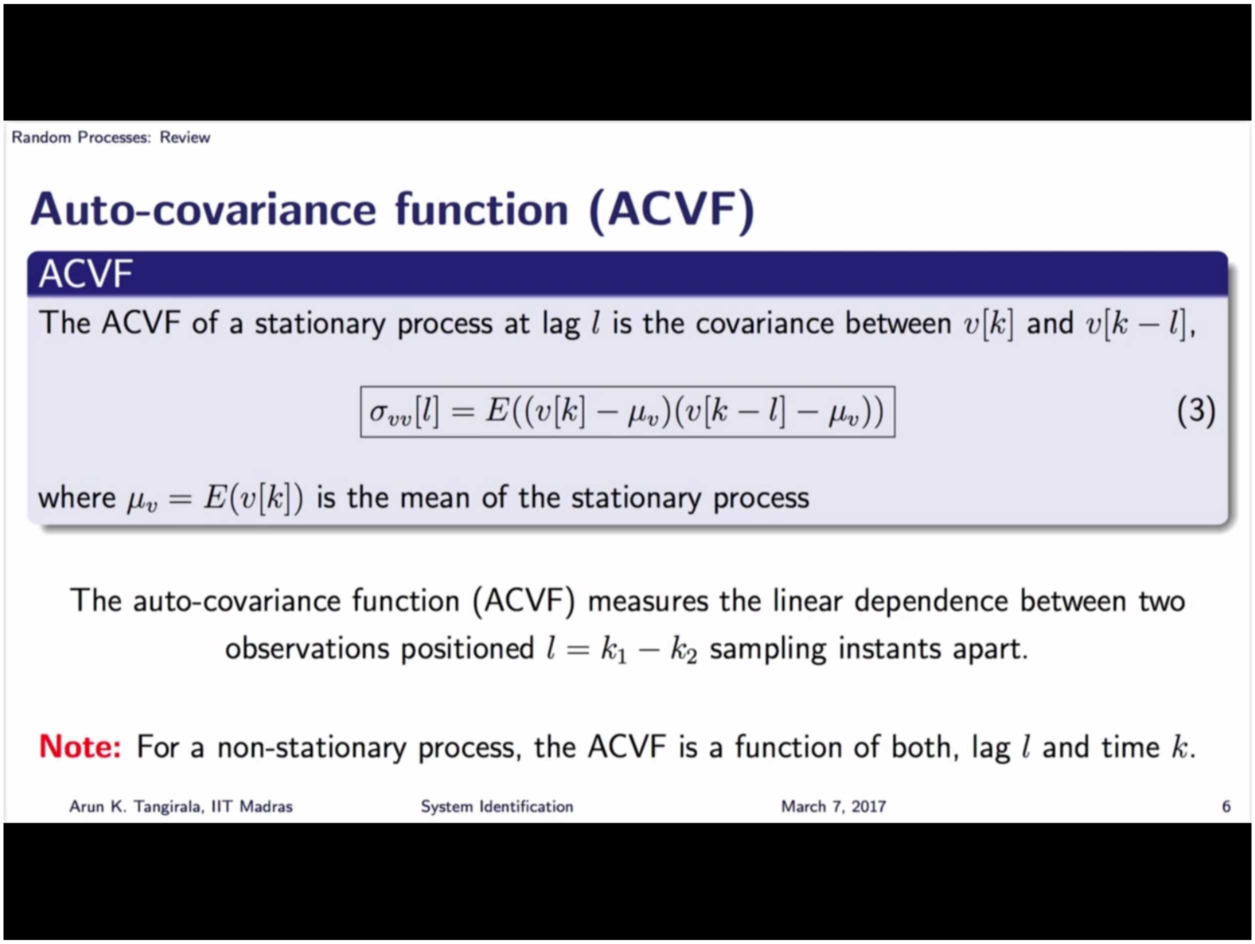
That is where ergodicity we assume, of course that’s where estimation theory and inference comes in, what can you say about the… first of… what can you say about the true mean, which is not a function of time, I mean, which is not evaluated in time, I should say, based on the averages that you may compute in time. And that's what statistical inference is all about. But underneath that is this ergodicity property, that you can make inferences. How you make inferences is the… is what estimation theory is going to teach and that we will learn, right. And the course on hypothesis testing also tells you how you do that. But that you are allowed to make inferences is what ergodicity property tells you, gives you that license, so to speak. But it's a very valid questions and that's a question that comes to the mind of everyone, and say well I am going to look at things in time, how am I able to comment on what is happening in the outcome space, right. But that's where ergodicity property comes and says, okay you are allowed to, but how you do that, you still have to strike a relation between the estimate that you compute in time and the true one. The classic example is sample mean. If I take… if I… let us say I am looking at the mean inferencing problem. I compute as usual the mean in time, average in time and that is what we call a sample mean, from this what can I say about the true mean, which is a theoretical one. So suppose I want to say the true mean is 0 and I have an estimate that I have computed in time. How do I conduct this kind of a hypothesis test that the true mean is 0 versus it isn't, using the statistic V bar, which would be the sample mean? What I have with me is this, what I want to test is this. That is what is the subject of hypothesis testing and broadly speaking statistical inference.

(Refer Slide Time: 12:31)



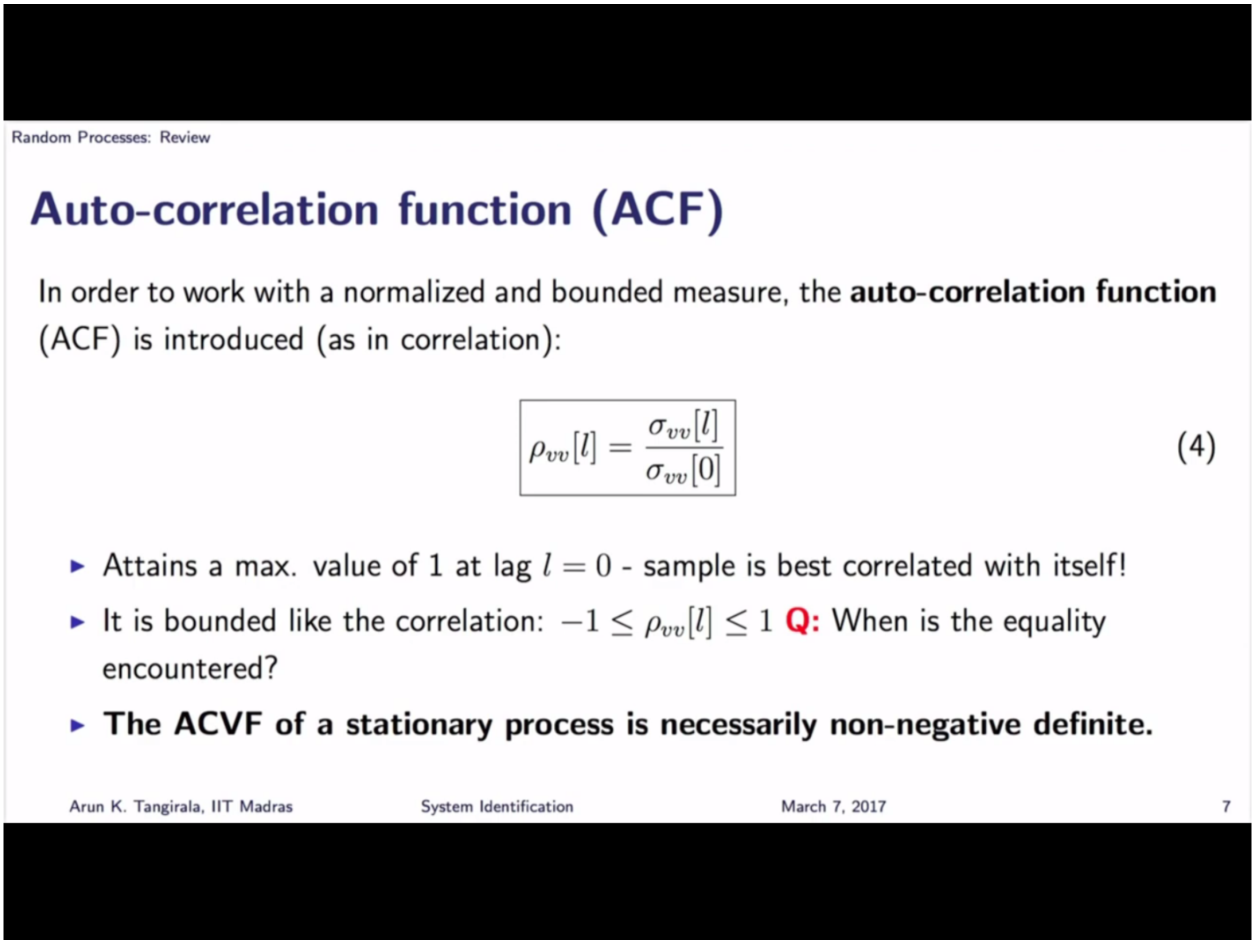
For that you have to understand now the relation between V bar and Mu and that's where we get into the estimation theory. Okay so let me just quickly talk about auto-covariance and auto-correlation function. We have already spoken about it. You should not think of auto-covariance function as another concept, in fact all the concepts that we need to understand random signals, we have already learnt, when we were reviewing random variables. All we are doing now is applying those concepts to the signal, that's all. So what is auto-covariance function, it is a covariance between two observations, you can say of the same random signal. You can say it's internal covariance, because you are looking at within the signal it is called auto-covariance to distinguish it from cross covariance, which we will review as well. Cross covariance by a stretch of imagination, you should be able to guess, would look at covariance between observations of two different signals at two different instants, correct. We are studying all of this, you should not forget to build a model, don't forget that fact. Everything that we learn, you should in your mind keep asking, how does it lead it's way to the development of a model. Why... why do you think we are studying auto-covariance function, because covariance is a linear measure and this auto-covariance is going to tell me whether there is linear dependencies within a signal. If there is, then I can fit a linear model to make a forecast. If there is none, then there is no hope, at least in the linear world, there is no hope. So this auto-covariance function has many roles to play. First it serves as a measure of linear dependence. So why does it serves as a measure of linear dependence, because at its heart is covariance. And we know covariance is a measure of linear dependence. So don't think that auto-covariance is actually a very alien concept. Now I have given the definition for a stationary process and I have already said for a second order stationary process, auto-covariance is only a function of the lag L. What is this lag? It's the distance between the observations in time and obviously it will keep changing with the lag L. How two successive observations influence each other is going to be different from how two observations positioned 100 samples apart are, 100 instants apart are going to influence each other. So sigma is going to be only a function of L, not the times at which you are going to look at this auto-covariance. The other thing is that this auto-covariance is a symmetric function… symmetric function of what, function of lag L. That means whether… whether I look at how V4 influences V6 or V4 is influenced by V2, it's the same. That means if lag is positive or negative, doesn't matter, the auto-covariance only depends on the magnitude of the lag, clear. It's a symmetric function. Why is it a symmetric function by the way? Can you reason it out in your mind? Why… why shouldn't it be asymmetric? Because covariance is a symmetric measure and you are applying it to the same signal, okay. We know covariance is symmetric, that alone is not the reason, but that's the beginning and you are applying it to the same signal. That means, whether I switch the order VK or VK-L it doesn't matter. Whereas cross covariance, it will matter, because I am looking at two different signals, alright, so the auto-covariance is symmetric function even from that view point.

(Refer Slide Time: 17:01)



Now there are many uses of auto-covariance and we will review those shortly. As with covariance, we have issues with auto-covariance, that is it is not bounded, it is sensitive to the choice of units for the signal and so on, therefore we work with auto correlation.

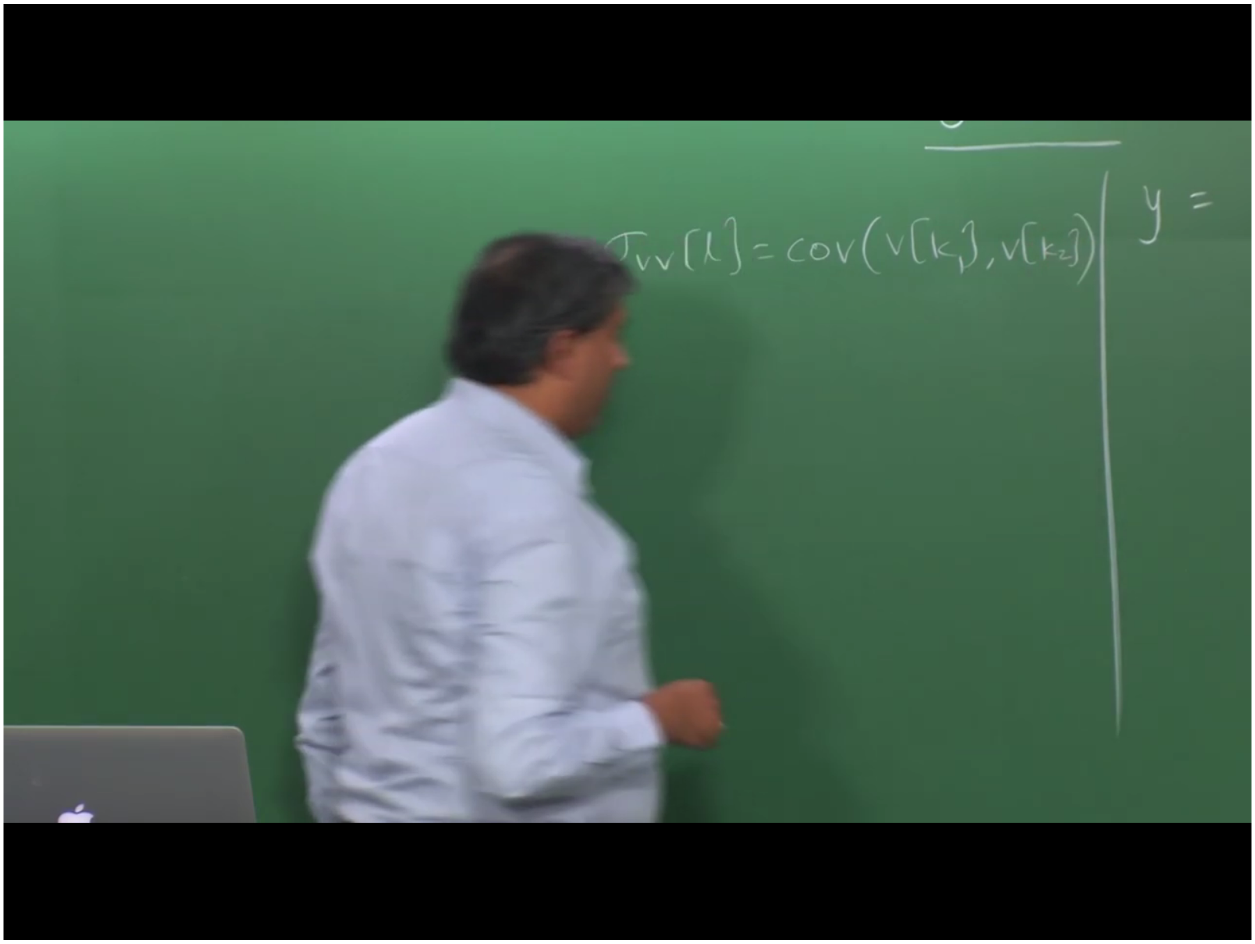
(Refer Slide Time: 17:30)



You just have to remember that auto-covariance is sigma VV at lag L is simply covariance between V at K1 and V at K2, that’s all.

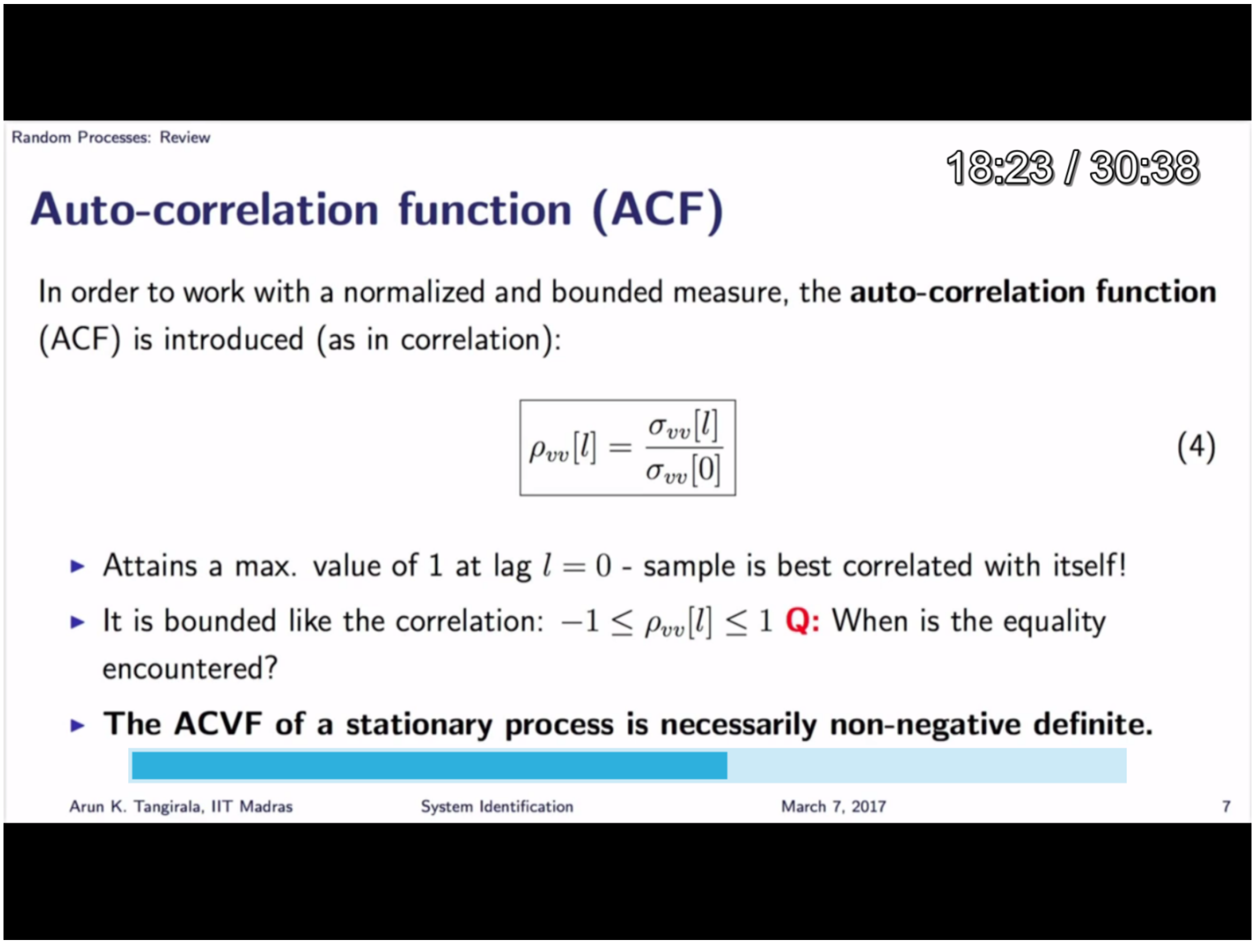
(Refer Slide Time:

(Refer Slide Time: 17:36)



Since this is a covariance measure I have issues, those two issues of unboundedness and sensitivity to the choice of units, therefore I work with auto correlation. And like auto correlation… like correlation, auto correlation is also a bounded measure and it attains a value of 1 at… at lag 0, what is the maximum value of correlation unity. That maximum value by definition here is achieved at lag 0, because of the way we are normalizing it. you have to argue in your mind. Why we have divided by the auto-covariance, why we have normalized it with auto-covariance at lag 0. How did this come about?

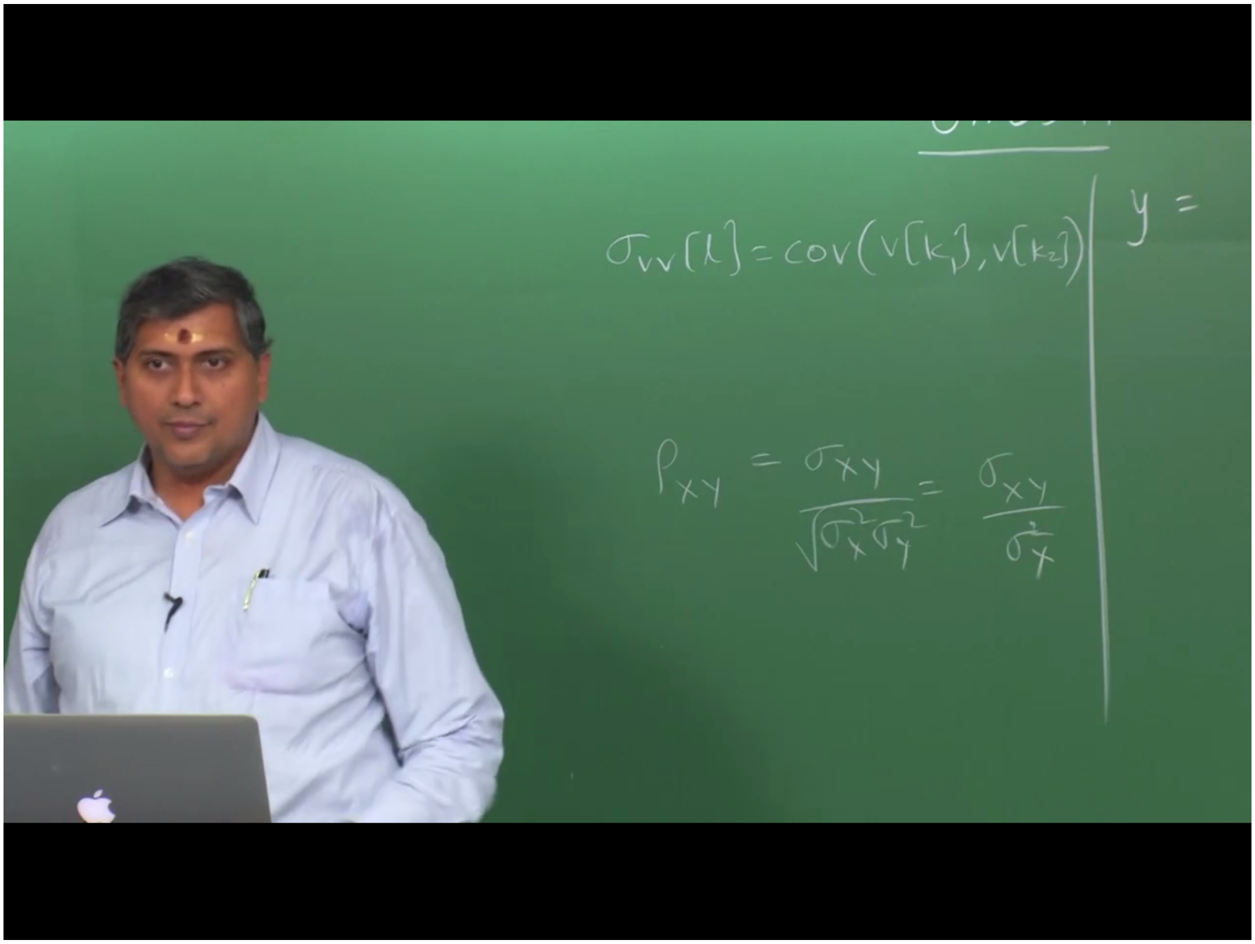
(Refer Slide Time: 18:24)



Why this normalization we have done? How do we normalize covariance, what is the definition of correlation? It's covariance divided by... Now how is correlation defined between two random variables. Sorry... no-no go ahead, how is correlation defined between two random variables?

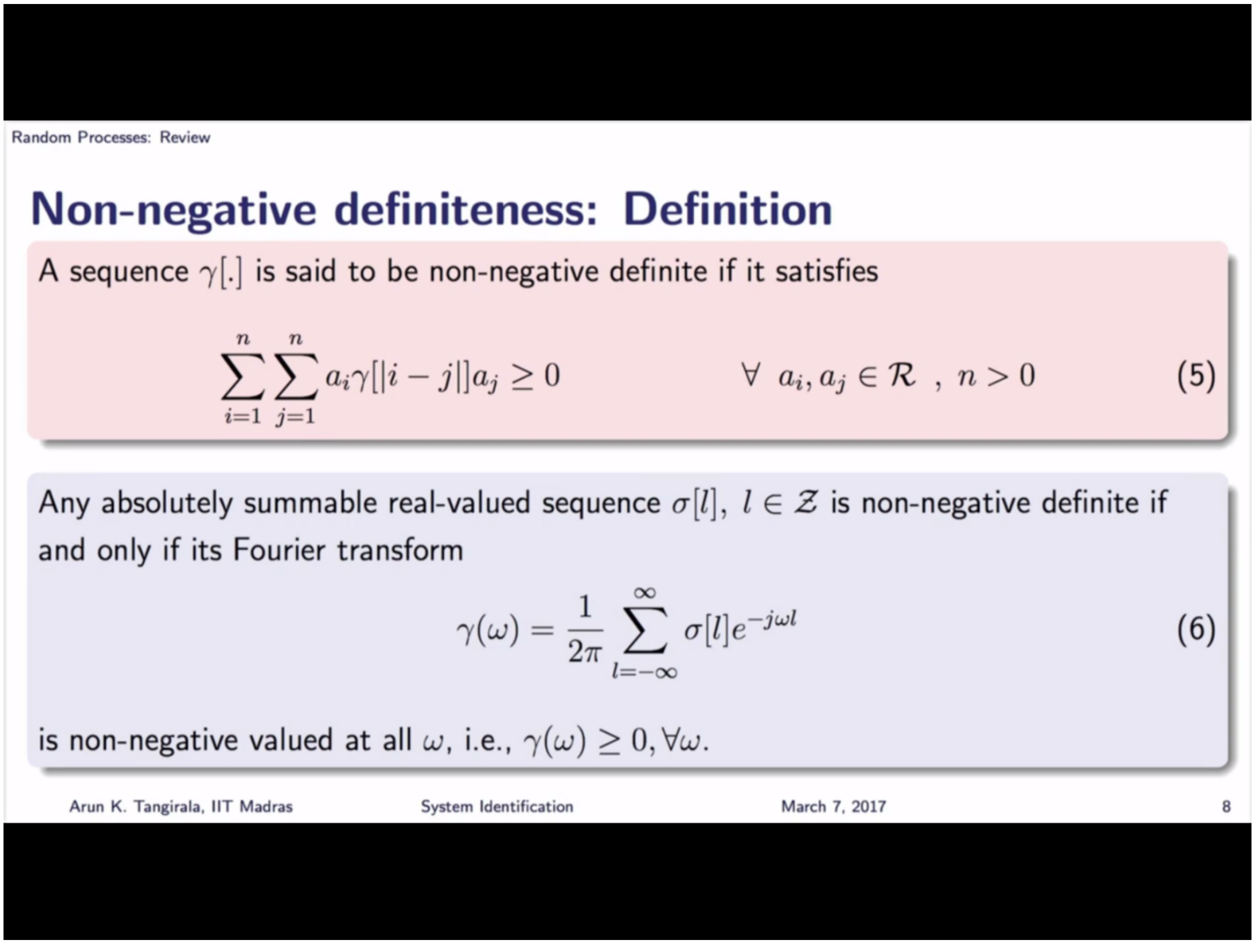
Sigma XY

Sigma XY by sigma X times sigma Y. What are X and Y here for us? V at K1 and V at K2, right. So if I look at the denominator I am looking at standard deviation of the or you can say, square... square root of the product of variances. So sigma square X would be the variance of the signal at K1, sigma square Y would be variance of signal at K2, but I have assumed it to be stationary, what does that mean, they are the same, right. And by definition variance is nothing but auto-covariance at lag 0, right. So ultimately here when I take the square root, I would get sigma XY, because sigma X = sigma Y, I get this, and that’s why you are normalizing with that.
(Refer Slide Time: 19:52)



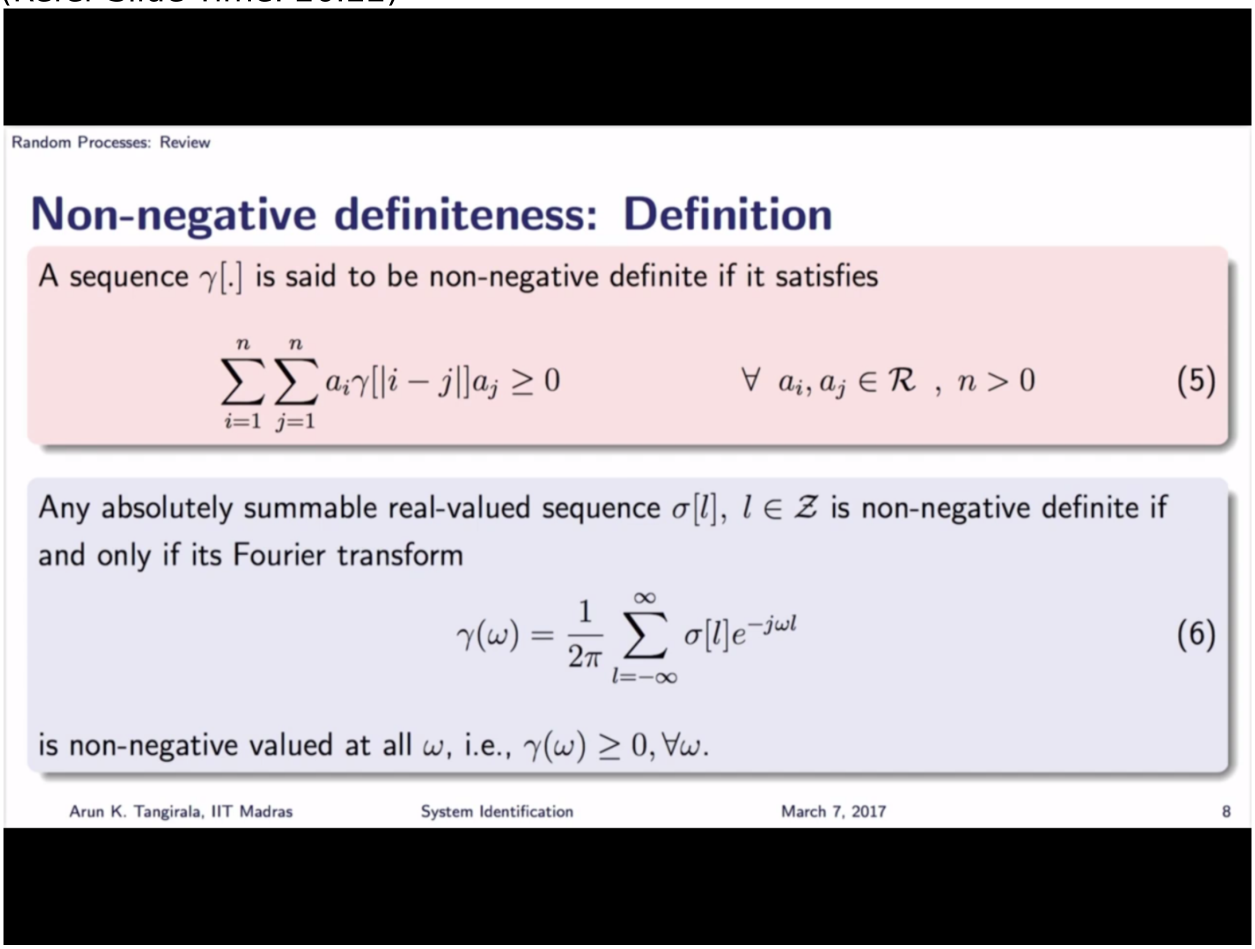
Now the correlation, I said… as I said, sorry the auto correlation attains a value of 1 at lag 0, that is obvious by definition. What does it mean? It means that every sample or every… at every observation, sorry… if you look at every observation… any observation, what is it's best correlated with, itself. What… who is the person who looks most identical to you or perfectly identical to you, yourself. The correlation between you and yourself is going to be 1, similarity. Correlation is the measure of similarity as well. Any other person, the correlation may be anywhere between 0 and 1, although as I say and as we hear, there are seven people who are suppose to be looking identical for any individual. So you… I don't know if you have found the remaining six, if it's Bollywood, it's easy to find. Even there I bet the correlation won't be 1, okay. So when you look at this auto correlation being 1 at lag 0, all it is saying is that… you know, the… what is the observation that is best dependent on, itself. Fine, the most important property of an auto-covariance function is… is that it is non-negative definite. Please do not think that it means the values are non-negative definite. Auto-correlation, like correlation can be negative valued also. Auto correlation is a sequence now. So we have graduated from a single number to a sequence. This sequence is said to be non-negative definite and there are two ways of defining non-negative definiteness. One way is this standard definition that you see, any sequence is said to be non-negative definite, sorry I used gamma here, please I will change the notation on the top, I should have used sigma there. So any sequence sigma is said to be non-negative definite, if it satisfies this condition, okay. But it's a difficult definition to use in place. A much easier definition to use in practice, to check if a given sequence is non-negative definite is this condition comes from functional analysis, Bochner's theorem, which says that any absolutely summable real valued sequence, sigma, it places some requirements, in fact the top, the earlier… the first definition doesn't place any requirement on absolute summability, the second condition says, any sequence that is absolutely summable, that means it should be absolutely convergent, is said to be non-negative definite, if it furrier transform exists. You recognize the right hand side to be the furrier transform, it’s a discrete time furrier transform. Now you realize why absolute summability is being given as a condition, because it has to be absolutely convergent. So it turns out that this furrier transform of the auto-co… I mean if you think of the sigma as auto-covariance function, it turns out that 5this furrier transform of the auto-covariance function is called the spectral density function. So slowly we are getting introduces to spectral densities. That omega is there is actually truly frequency. But if you don’t care about that, given any sequence, I want to figure out whether it is non-negative definite. And if I am given that it is absolutely summable, I can use the second result, and straight away arrive at the conditions for non-negative definiteness.

(Refer Slide Time: 23:52)



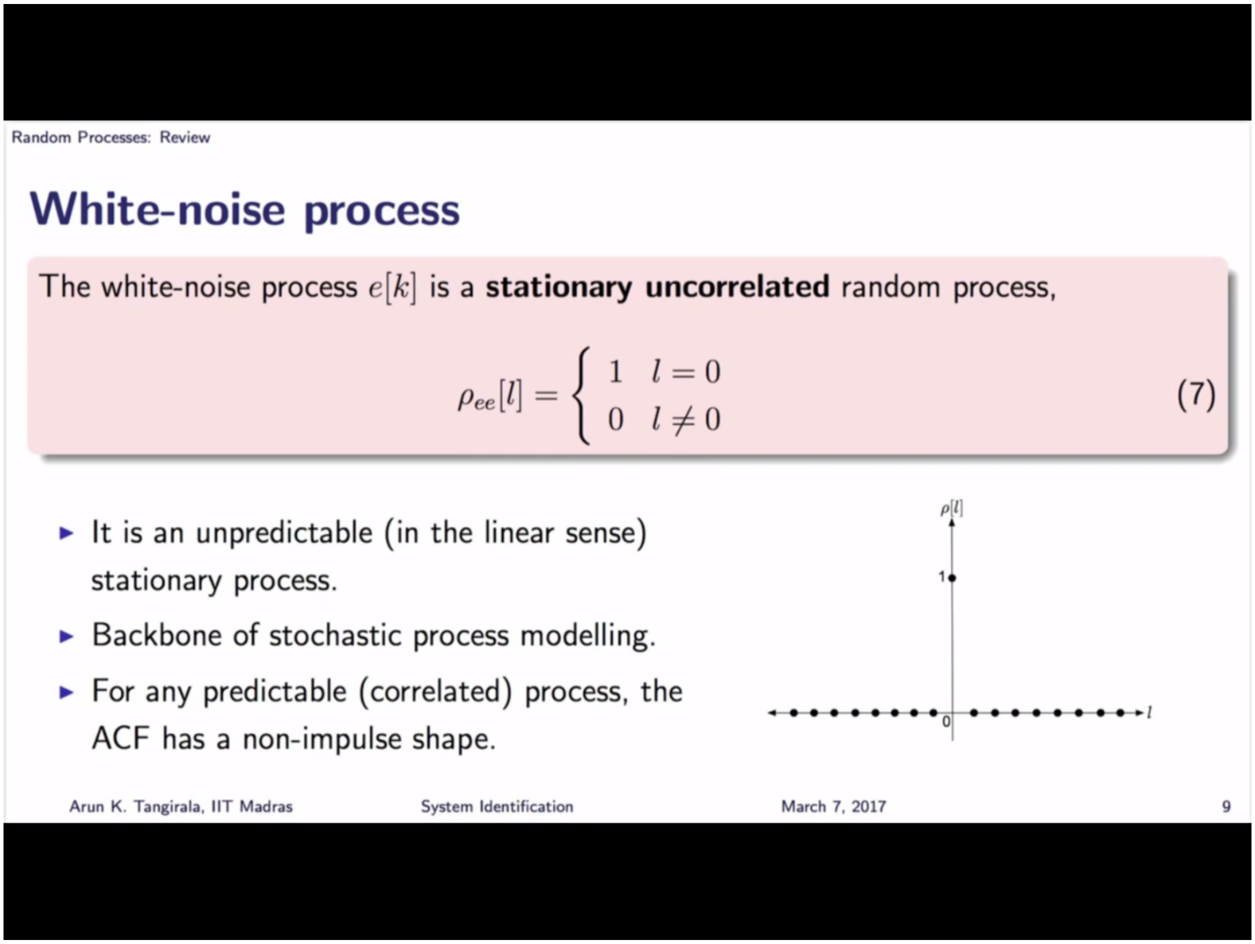
For us what this means is that not any symmetric sequence qualifies to be called as the auto-covariance function. There are many symmetric sequences, but not all of them are necessarily non-negative definite. Only small... or a subset of symmetric sequences are non-negative definite and therefore I need a special condition and only those qualify to be called as the ACVF of some process. So the question that you should ask yourself is why is this condition so important, why are... why do people keep highlighting the non-negative definiteness property of an ACVF? The reason is when I estimate ACVF, this is all theoretical discussion, when I estimate, what am I going to do, I am going to use a formula, I am going to be given N observations and I am going to use some formula to estimate ACVF. Here I have used a formula to estimate mean, this we call as a estimator. Likewise, I will use some other formula to estimate ACVF, then that estimator should guarantee, that all the estimates that I obtained of ACVF at different lags workout to be a non-negative definite sequence, if they don't, then unfortunately the model that I will build from the auto-covariances cannot be guarantee to the representative of the stationary process, you understand. It is important that the estimates also satisfy certain important properties of the theoretical quantities that I am estimating, then only the models that I build will make sense. So one of the important requirements for any formula that you come-up with to estimate, any method that you come-up with for estimating ACVF, you have to guarantee that it results in non-negative definite sequence, then only you can say, you can proceed further and ask, how good is estimator and so on, when it comes to ACVF. Likewise when we talk of spectral densities, by definition spectral densities are non… so look at this, it says that first this spectral density should exist, sorry, I didn't complete the statement here, the spectral density should exist and it should be non-negative valued at every omega, then only the sequence is said to be non-negative definite.

(Refer Slide Time: 26:22)



Mere existence alone is not… that any way is guaranteed by absolute summability, but it should be non-negative value. Here we are talking of values, when it comes to ACVF, we are not talking of non-negativity of values, we are just talking of non-negativity of the sequence. But when we are looking at spectral density, it's a density function, this gamma of omega is a density function, density function have to be non-negative value, and that is what is the requirement. This has… this particular result that you see is at the heart, you can say it's fundamental to the entire world theory of random processes. You cannot even imagine, you know, in what ways it applies. So anyway, I will just conclude the class with white-noise, and we

will continue the discussion. It’s a very simple one, we have already spoken about white-noise processes and you know white-noise is… this notion of white-noise is extremely important. Don’t worry about non-negative definiteness so much, you just have to remember that any ACVF is… sequence is a non-negative definite when it comes to random… stationary process. Okay so the white-noise process is this idealization that serves in many different capacities in time series modeling. It is this ideal unpredictable process, why… why do I need to define this, because when I am going to build a model, let's say pure time series model, the first thing that I am going to check is, if the given series is white, if it is, then if it is an assignment problem, I am happy, I don’t have to build a model, but if I have been given a job, and I am being paid for it, then I better you know, impress my boss with a model, then you are disappointed. So that is just one use, remember when we build a model, we have said, if you recall in the liquid level system, when we talked of goodness of a model, what did we say, as to when the model is good? when the residuals, that whatever, the model could not predict. One of the conditions is that the residuals should be white, other condition is, it should not have any input effects, but the other… so there are two conditions, one is that the residual should not have any effects of input left, that is for goodness for G hat and the goodness of H hat is measured by looking at the whiteness of the residuals. So again there the notion of whiteness comes in. So for many reasons the white noise process conceptualization is extremely important and it is defined as a stationary un-correlated random process. It is only specified in terms of the ACF, not even the variants. It says the ACF should look like an… is like an impulse, that’s all. It doesn’t say, what should be the variance and so on. It doesn’t say, what should be the distribution, that is the p.d.f. is also not specified, so I can have a Gaussian white-noise process, I can have a uniform white-noise process and so on, all that it says is, it should be stationary, second order stationary and un-correlated. These are the only two requirements. And any process that satisfies these two is said to be a white noise process.
(Refer Slide Time: 29:51)



So tomorrow we will complete our discussion, at least of the time domain characteristics, where we will look at partial auto correlation function and cross correlation functions. We will of course begin the class tomorrow with some more discussion on the white-noise process, in what way white-noise process is going to be serving as an extremely important tool in system identification or in modeling, okay. Thank you.