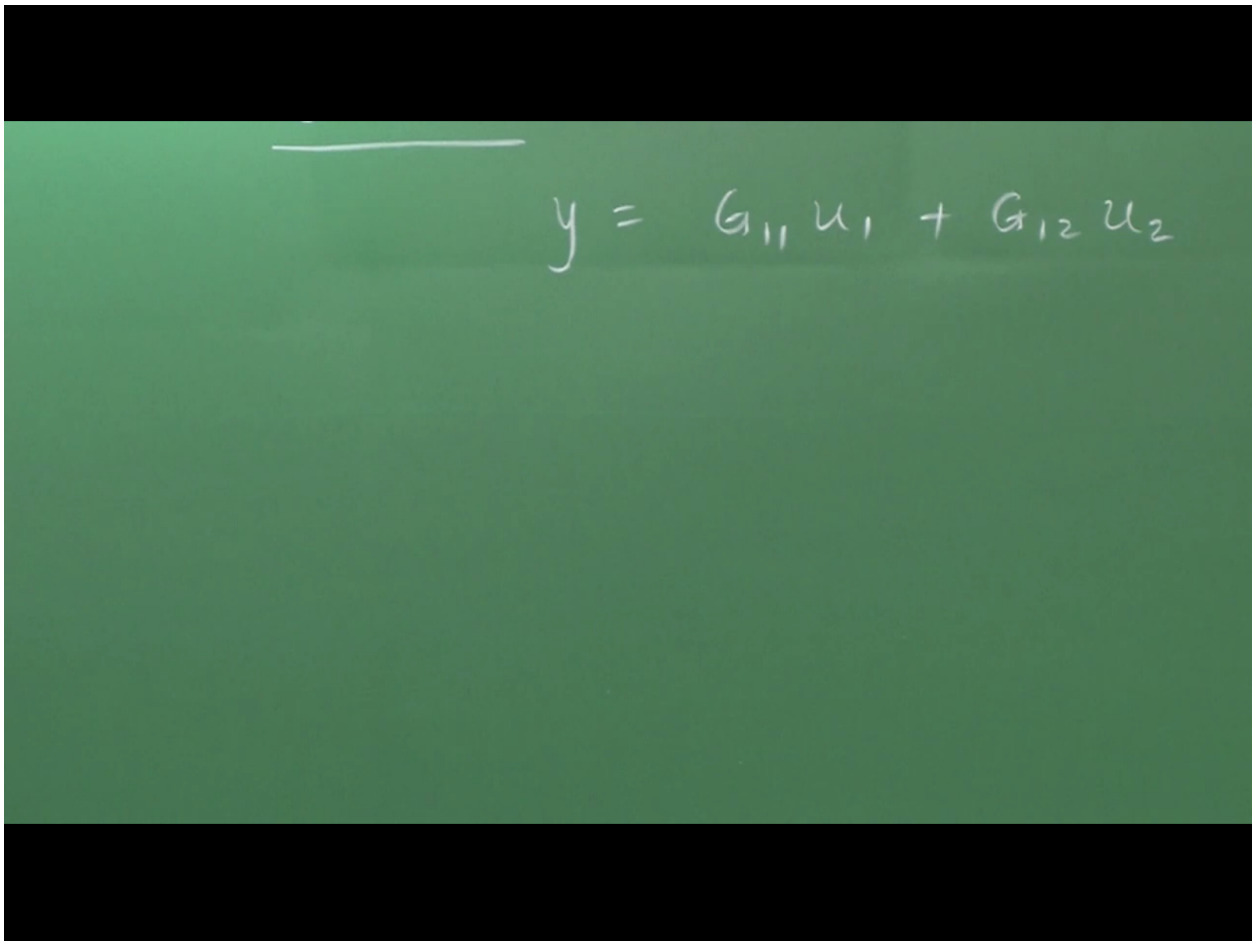
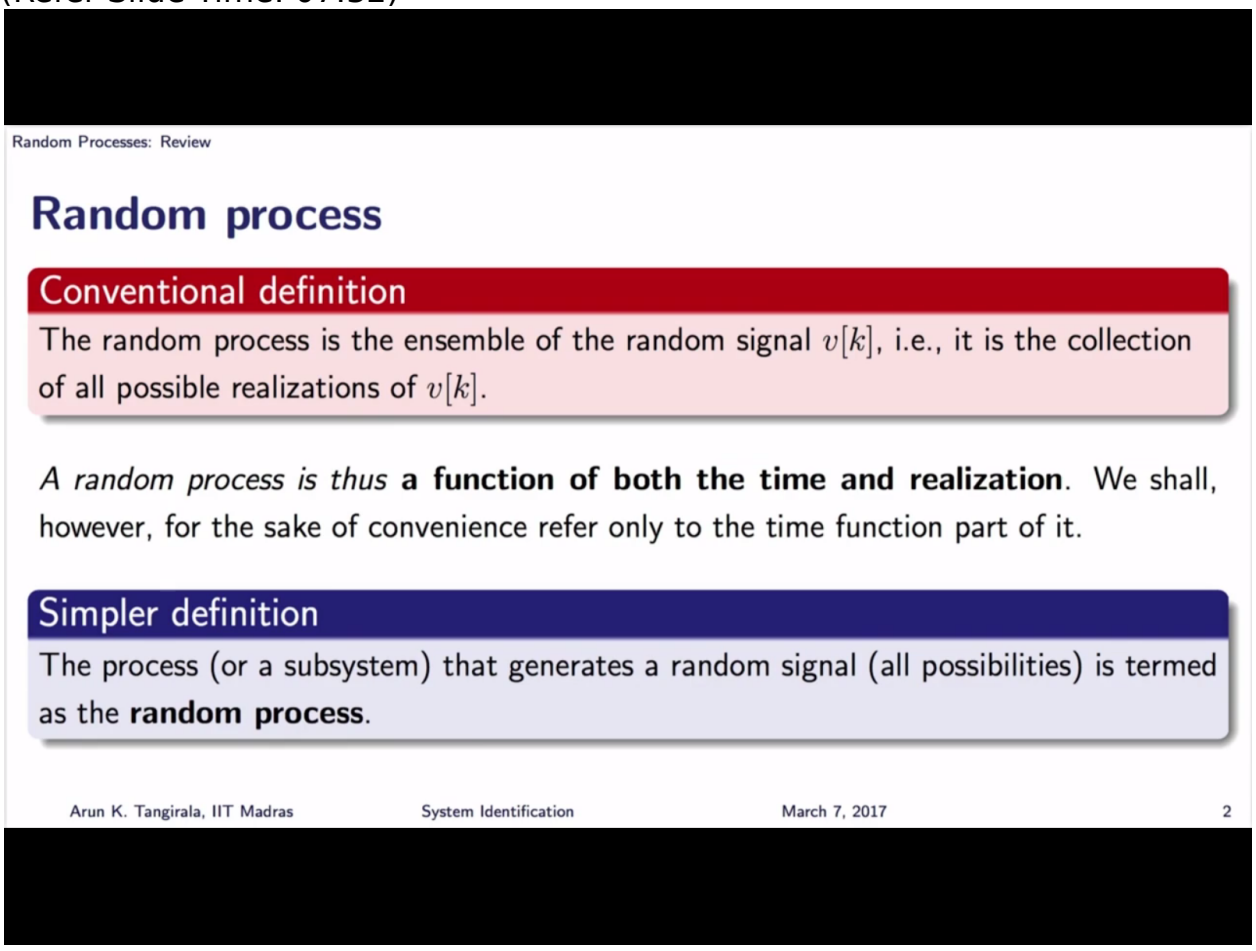
Very good evening. In the last class we concluded our review of random variables. Where towards the end I spoke about partial correlation. After having discussed correlation we realized that there is going to be confounding that will haunt us, and that's true of any data analysis. Whenever you are considering measurements of a set of variables, there could always be the... the scenario, where you have left out some other variables in your analysis, and that's not deliberately, it's simply because you... perhaps do not have measurements with you. And since typically that is the case, you will run into confounding invariably, unless you guarantee that the variables that you have included in analysis are indeed the set of variables that are of relevance to whatever, your process that you are looking at. But many a times we do have variables that we have measured, but have not taken into account, in let us say correlating to variables and that's when we can run into confounding partial covariance or partial correlation, essentially allows you to compute the correlation between two variables after discounting for either one or more confounding variables. The example that we went through was for the case of a single confounding variable, but you can extend that idea to multiple confounding variables as well. And in some sense you can think of correlation and partial correlation as analogies of total derivative and partial derivative. When you are evaluating partial derivative, you hold all the others constant, right all the other variables constant, and then you basically get to know how one variable changes with respect to another variable, that variable alone and that's obviously of interest because in a multi input, let us say you have multi input and single output process, right. Why is this correlation or partial correlation relevant, how is it relevant system identification, when I have a multiple input, single output system and I am trying to look at the correla... I am computing the correlation between the output and let's say one of the inputs. What is the danger or what is the risk that I run into, the correlation between Y and U1 let us say I am looking at the first input, can be confounded by other inputs if the other inputs are similar, they have some similarity with U1, right. So numerically therefore the correlation between Y and U1 is not truly the correlation... is not truly representative of the channel between U and Y1. It can contain the effects of the other input as well. So you can just see this, you have Y=G11 U1+G12 U2, and you correlate Y with U1, there will be a second term, which will contain correlation between U1 and U2, unless you have guaranteed in your experiment that there is nothing numerically similar or statistically similar between U1 and U2, you cannot assume that the correlation between Y and U1 is only representative of G11.
(Refer Slide Time: 04:00)



On the other hand if you have done an experiment, where you have designed your inputs in such a way that U1 and U2 are uncorrelated. It's possible. We know that correlation and un-correlatedness in statistics would mean our orthogonality in functional space. So you just have to ensure that U1 and U2 are some… in some sense perpendicular or there is no correlation. For example you can take a sin and a cosine, something like that, right. If you design your experiment carefully, then it becomes easy. Then you can straightaway say that the correlation between Y and U1 is indeed reflective of G11, there is no effect of G12 at all, otherwise, you will have confounding. Partial correlation between Y and U1, conditioned on U2, because you have the knowledge of U2, will make sure that the partial correlation, which is between Y and U1 is only due to G11 and quite a few methods in the literature do exploit this property that the partial correlation between Y and U1 conditioned on U2. Whenever you talk of partial correlation, you have to mention conditioning variables also. You have to tell the reader, what are the confounding variables that you have taken into account. So here if you were to compute the partial covariance or correlation between Y and U1 conditioned on U2, then you would have numerically eliminated the effect of G12 on Y. It is in some sense you can say effectively, what you are doing is, you are breaking this meso system into multiple seesaw systems. You are

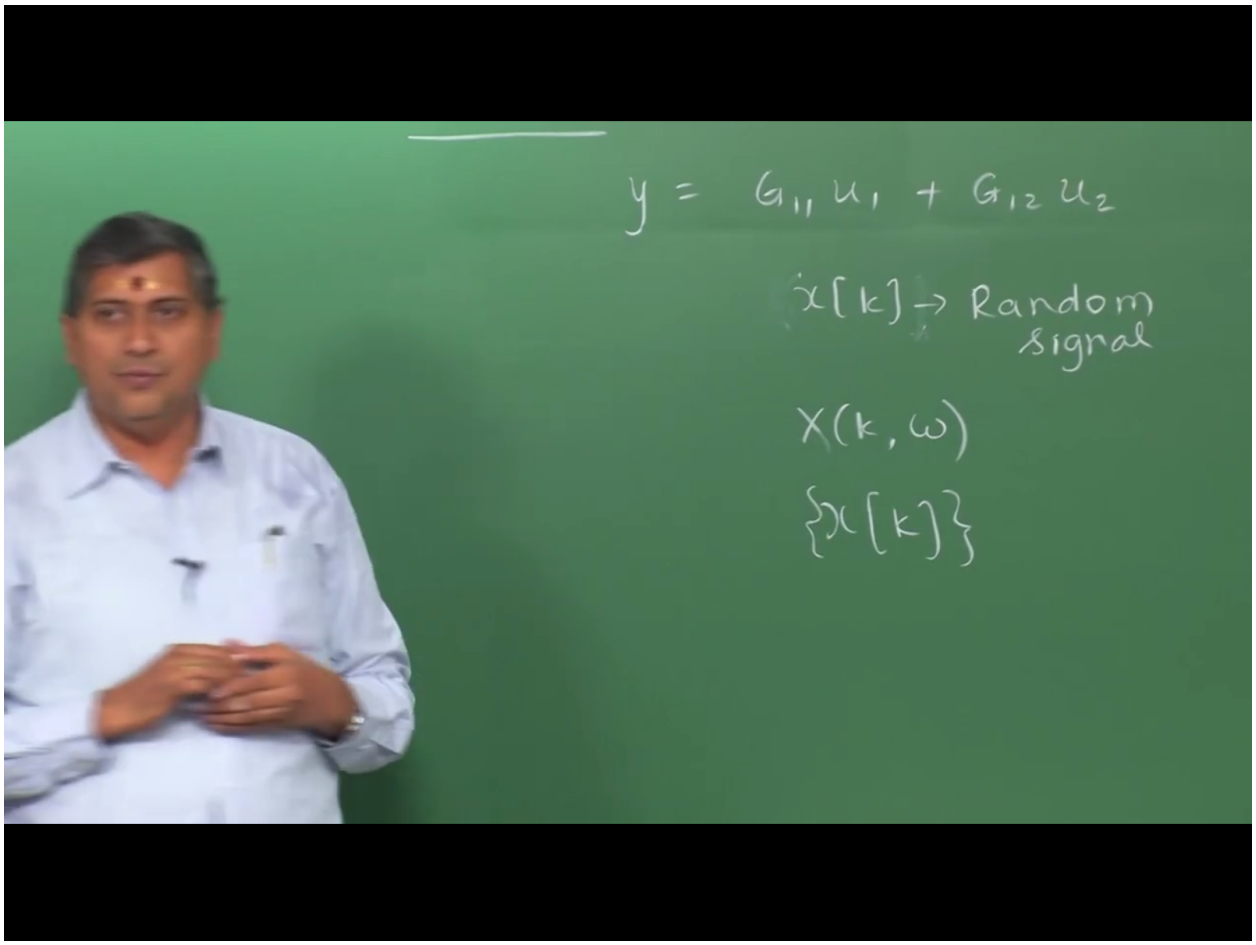
not breaking it physically, you are doing it numerically. This is what we call as decoupling numerically, okay. So we will... we will see the application of partial correlation today when we review now random processes, where we will come across a tool called partial auto correlation function. First we will get intro... introduced to auto correlation functions, where we will just quickly review the concept and it's properties and then also talk about partial autocorrelation functions, okay. So it's time to now move on to random processes. So we have to graduate now from random variables to random processes. And you must recall each time, why we went to the theory of random variables, because at each instant the random signal is a random variable. Now we know already that the random signal is nothing but an ordered collection of random variables, of course that exists infinitely in time. Associated with this random signal is a process called a random process. There are two ways of looking at a random process. A conventional definition says that the random process is the ensemble of all possible random signals that you can think of. See at each instant you have a random variable and we know that for a random variable there are many possible outcomes. Now when you are tying them together, you get myriad of possibilities of random signals. The collection of all such random signals is called a random process. So it's like an ensemble.

(Refer Slide Time: 07:32)



A more practical or a pragmatic perspective would be simply to say there is this random... there is this process that generates all possible realizations. We can say that one... one version of a random signal is a realization. So a process that generates all possible realizations is what we call as a random process, either way... either way is fine, but all you have to remember is that random process is essentially the one that is responsible for the generation of all possible random signals associated with that phenomenon. And the other thing that we should remember is that the random process is both the function of time and realization. That means it's a two dimensional thing unlike a deterministic process, which is only a function of time or space or some other independent domain. For a deterministic signal there is only one possibility and associated with that is a deterministic process, which is also only a function of time. Whereas with the random process, you have also the realization space. So many... in many texts you would see the random process is XK or VK, we will now move on to the notation VK, but suppose XK is the random signal that you are looking at, then the random process, this is your signal, the process is usually denoted as X of K times K, Omega, but this Omega should not be confused with frequency, unfortunately Omega is also a symbol that is used for realization. So strictly speaking a random process is a function of both time and realization, normally. We do not talk of the realization, we suppress that dependency, remembering in mind that a single random, course of random signal that we see is nothing but a single realization. If I were to observe, if I were to think of the random process, it will be collection of all such things. So the random process is actually a function of both, but normally we don't use this notation, we simply use this notation. To mean that there is a collection of random signals. Sometimes we may even drop the curly braces. You just have to understand that we are referring to the process, okay.

(Refer Slide Time: 09:46)



So one of the first things that we have to do, we have to think about in the analysis of random processes is this concept of stationarity. And why is this important. So what do we mean by analysis, primarily what we mean by analysis here is modeling, right. So when I build models for random processes, so I am given data, and I am going to build a model from this single realization and obviously this realization is going to be given to me over a finite time, I cannot have an infinitely long realization, some 100, 1000, 10000 whatever, it is, the number of observations is finite. The model that I build from this finite observations should hold good for the entire random process, that is requirement number one, right over that period of time and requirement number two, it should be valid at all times, otherwise, what is the point. As a simple example, suppose I give you record temperature, atmospheric temperature, in the morning, let's say between 6 and 8 o'clock, and I say okay build a model that will forecast the temperature at other times in the day, do you think this is a well post problem, or is there an issue with this statement, is there some issue, there is no issue, you can do it, what is the issue?

(inaudible)

Okay so therefore…

(inaudible)

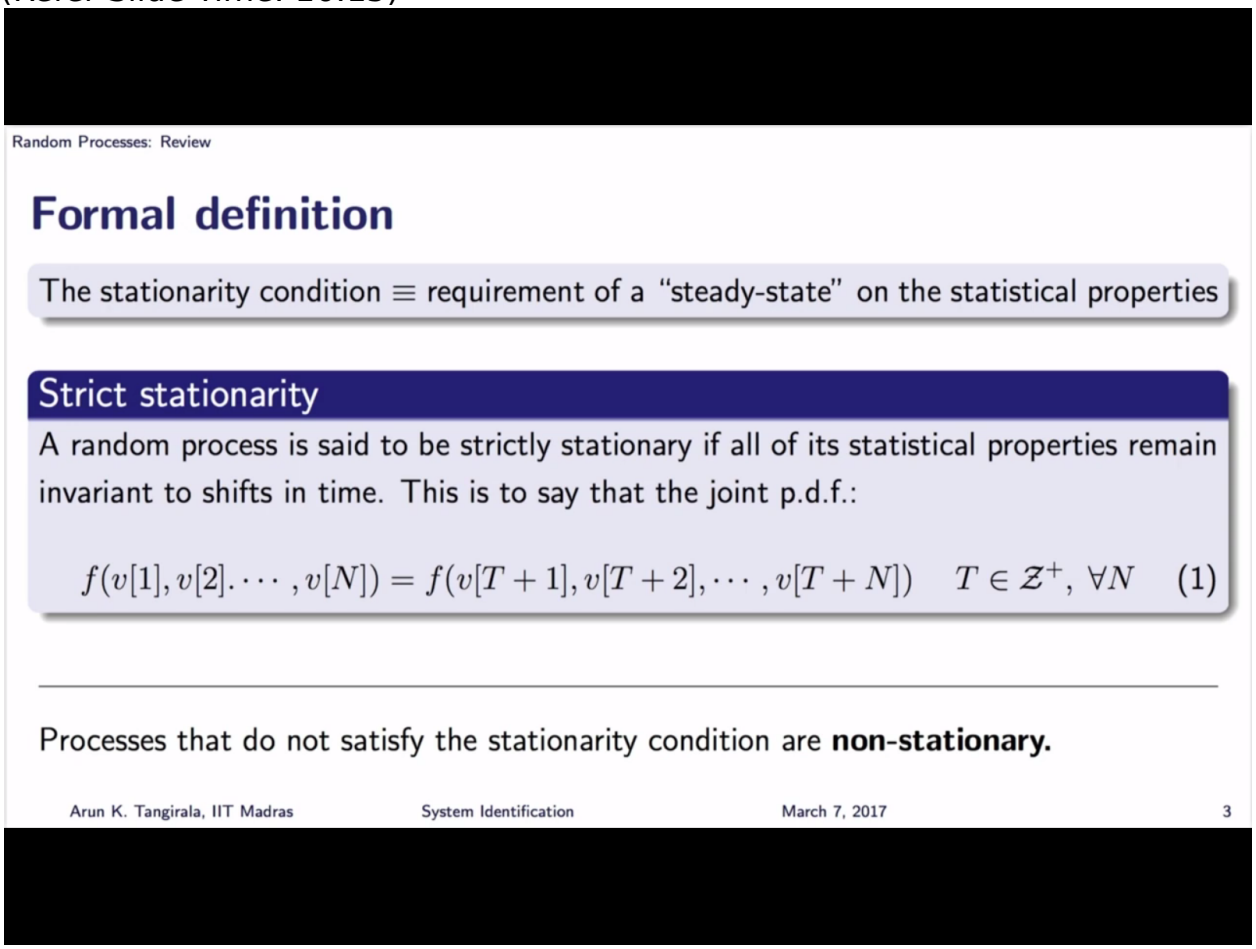
Different and therefore… it will always be different, you have to be more clear, now you have to be more technical in what… what you are… I understand what you are trying to say.

(inaudible)

Sorry, correct, so there you will see the local levels average as being shifting, that itself is enough to put us off. Say whatever model I build between 6 to 8, may be valid for the next day perhaps, but not for the same day, right. So these… processes like these are said to be non-stationary. Their statistical properties are not invariant with time, this is one thing that slowly you should get used to, in deterministic world we spoke of liner time invariant processes. How did we define time invariance in the deterministic world, we said it doesn't matter when I look at the process, whatever, input I gave yesterday, if I give the same input today, or at any other time, it should evoke the same response. So we are looking at the response. Why are we paying attention to the response because for a deterministic signal, the value of the response has everything in it, it complete, yeah it's… I can place complete faith in it. Whereas for a random process, the value is not the one, even if the process is so called stationary, the value will be different, because after all you will get another realization. So what is it that I expect of invariance when it comes to stochastic processes. Basically I expect that the statistical properties, which are deterministic around, which are determined, the center of outcomes, the spread of outcomes or any other statistical property, like correlation and so on… they should not change with time. So there is something fixed about a random process that we expect. But we cannot expect the value of the observations to be the same, because we know even if I were to freeze time, if I observe with many sensors, I will get different values, therefore I cannot define or I cannot expect the process to remain same in the realization world, in the value, in the observation world, but I should expect it to be invariant, I can expect it to be invariant in some other domain or in… in some other aspects and those aspects are the statistical properties. Now if you stretch this imagination, we know mean is the first moment, now remember by the way, when you are dealing with a random signal, we are dealing with collection of random variables, and whenever we deal with a collection of random variables, we are looking at joint p.d.f., correct. So when we talk of variants or when we… when we talk of this collection of random variables, we are talking of joint… moments of this joint p.d.f. Strictly speaking we want all the moments of this joint p.d.f. to remain invariant to time, that is the ideal expectation, like an ideal gas. We know it's going to be extremely difficult for any process to fulfill that requirement, but it's good to start with that idealization, so that we know how much the

deviations, we are willing to withstand from this idealization. So strictly stationary process is one whose joint p.d.f.... no directly we state this requirement in terms of the joint p.d.f., we say... notice that now we have changed notation from X to V, okay slowly we are getting into the society notation. So V1 to VN are the N observations that I have. What we require for strict stationarity is that the joint p.d.f. of this N observations should remain the same at over... at whatever, time I observe. So if I shift this entire observation by a certain time T or you know certain time instant... number of time instants T, the joint p.d.f. should remain invariant for all times and for all sample sizes. That means if I look at the joint p.d.f. of three observations and joint p.d.f. of three observations may be in the night, they should remain invariant, or tomorrow or any time.

(Refer Slide Time: 16:15)

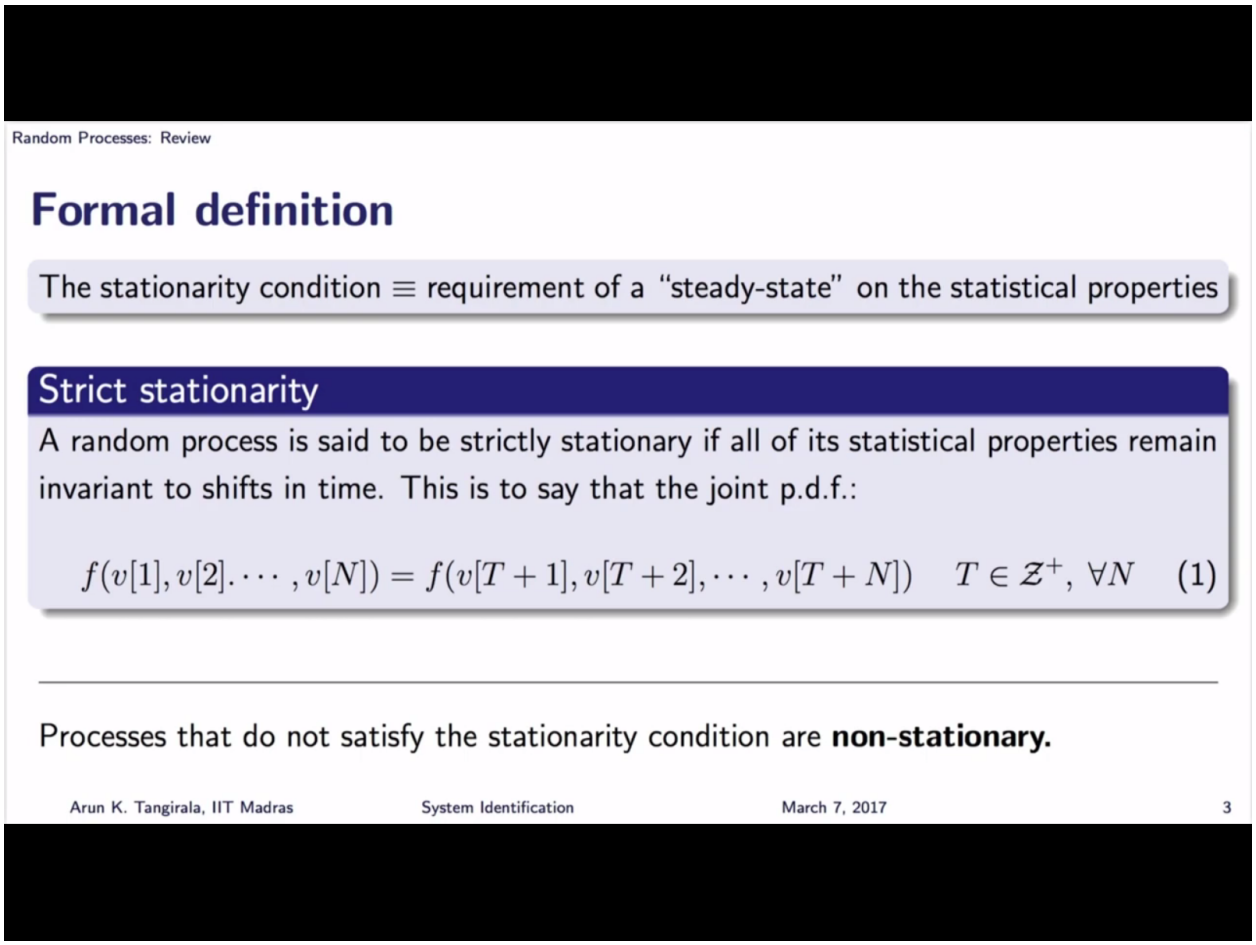


Likewise I look at joint p.d.f. of 100 observations, now, later or any other time, they should remain invariant, which boils down to saying, if I shrink the number of observations to one, at each instant the p.d.f. should remain the same, because it should be... it should be satisfying for all N and for all T. Every observation should fall out of the same p.d.f. It's like, suppose I tike a dice that I am rolling, that we use for ludo and other games, what happens, I am looking for some face value 1 2 3 4 5 6, there is a certain probability

associated with the roll of a dice and the surface on which I am rolling, let's say the surface is very good, very smooth, no issues, so we assume uniform probability, that is all values are equally likely 1/6, but is that going to be the same. If I am looking at a random process, where I am only observing, at each instant what is the face value, now I am going to construct a random signal out of it and I am looking at associated random process, is it going to be strictly stationary? Yes or no, if yes why and when?

(inaudible)

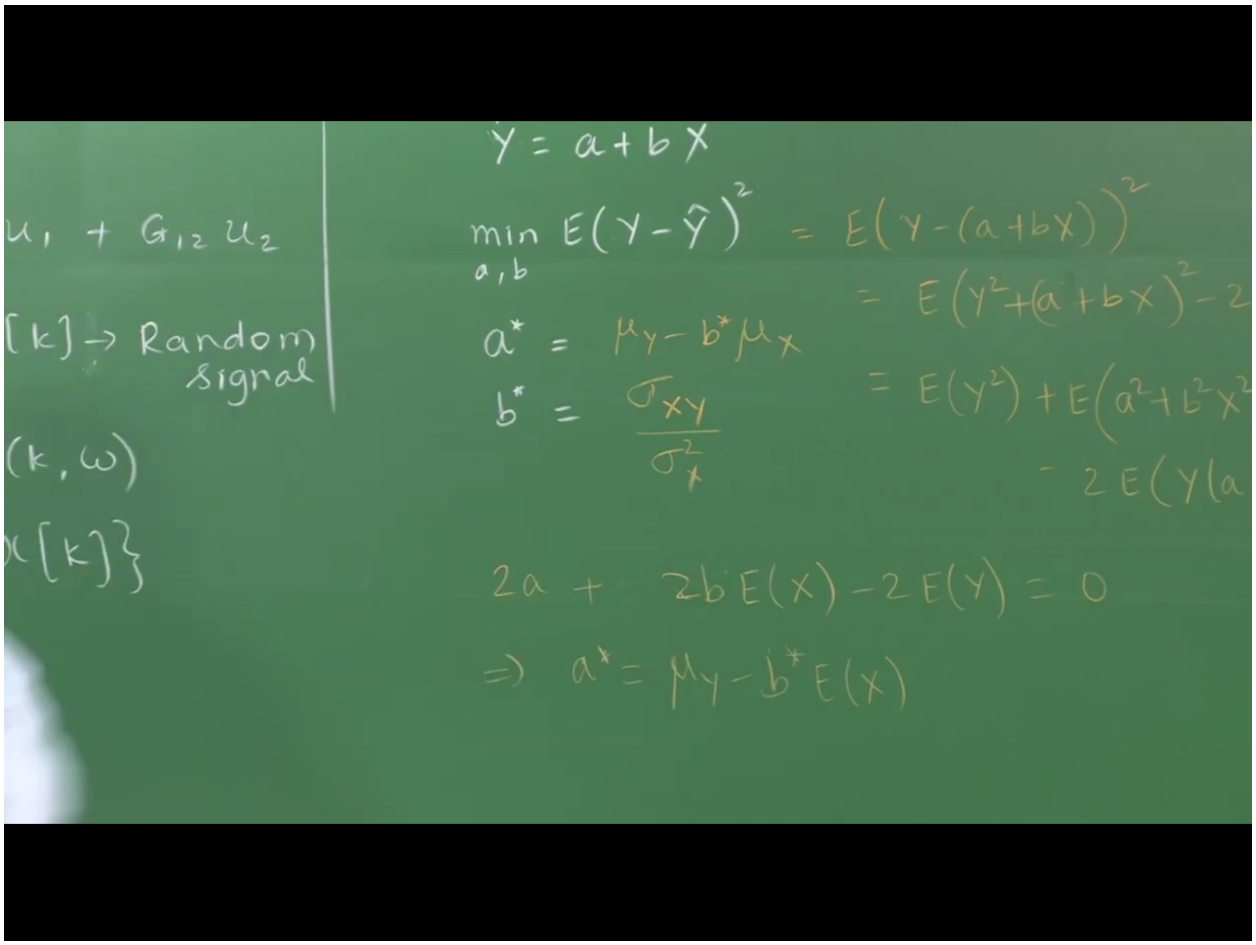
Correct, but after considerable time, right, it depends on the dice. If it is… if it is very well manufactured, then I can assume it to be strictly stationary for a very long time. But if it is a fake one, right, then probably I cannot assume it to be strictly stationary. Now this is an idealization. When we say that at all times and at all observations it is impossible for any process to full fill this, but over considerably large time scales, compared to the observation scale that I have. So if I am performing the experiment, let us say over a few hours and the dice take years to deform, then for all practical purposes I can assume it to be strictly stationary. But if I buy it today and in a week it deforms, then I cannot assume the resulting random process to be strictly stationary. In general it is hard to meet this requirement, but we will start with this requirement. And by default any process that does not satisfy this condition is said to be non-stationary. So as you can see the strict stationarity condition is on the joint p.d.f.
(Refer Slide Time: 18:55)



But we very well know even from the theory of random variables, I don't have to worry about the p.d.f. all the time, I can work with the moments, right. We said that, in linear world… in the linear world I don't have to worry about the p.d.f. all the time, it suffices to worry about the mean and variants and covariance, when I am looking at two variables, it's covariant. How do we prove that, suppose I have a variable Y, that I am predicting, random variable Y. Let's say I am predicting it using some random variable X. What is the statement that we made, when we were reviewing random variables, we said for linear… in the linear world, I only have to worry about the first and second order moments, right. I don't have to worry about p.d.f.s. So let us say, now this the model that I want, this is a prediction or predictor that I want to build for two random variables X and Y and I want to predict Y using X and let us say I am going to minimize, I am going to find Y… sorry A and B in such a way that this is going to be minimized. Here we don't have any observations, I am just saying there are two random variables, I am going to predict one using the other. Obviously I know there will be some error, because it's a random variable. Now I want to fit a linear model, although I have written a linear predictor, for now assume that's a linear model that I am fitting, I want to find A and B in such a week that I get this minimum mean square error, it is a standard criteria. What are the optimal values of A

and B? Can you give me the answer? In terms of the statistical properties of Y X. You should get used to this slowly now, you already have questions similar to this. You should be extremely comfortable with computing the so called theoretical or optimal estimates. So just proceed stepwise, you should get the answer. So what I want is A star and B star as optimal estimates of… If for simplicity assume, X is 0 mean, so… it's up to you, doesn't matter. Anyone with the answer. Please make sure that you know how to solve this problem. This is the most basic problem that is solved in system identification. Where you have only two random variables, we are not even looking at signals and we are fitting a linear model, very simple, that's what you meant. Okay so your issue is, you don't know to expand this subjective function. What is there in expanding, it's a quadratic. So you have expanded, that's not a issue. Expectation of Y-A+BX to the whole A+BX there is whole square here, so how many terms do we have? And if you were to look at it, you would have three squares and then three cross terms, correct. So let's write that, I mean, if I don't know how to do things, let become a baby and do the… take the baby steps. So I have here expectation of Y square + A square or you can keep it this way, you can say A+BX to the whole square – 2Y times A+BX, fine, that's it and therefore I have here expectation of Y square + expectation of A+BX to the whole square, which is A square + B square X square + 2AB X – 2 times expectation of Y time A+BX, fine, is it clear? Now I just have to differentiate this with respect to… partially with respect to A and B, that's also now when you differentiate with respect to A, this term doesn't prevail. This term vanishes, then this term also vanishes and therefore if you look at the equation that you get for solving A star, you get 2A, because expectation of A square is A square plus what do you get, 2A times B times expectation of… and then what else minus what do you get 2 times expectation of Y and this we set to 0, am I right? Is that correct? What happened, that's fine. Therefore, is there a problem with this, no there should, no A here, I am sorry, yeah it's 2B, correct. Therefore A star, if… simply if you take it to the right you get Mu Y minus well this B is the one that's at optimality, so we say it's… this also has to satisfy, both have to be satisfied, so this B is the solution to the other equation and that's what you have here, this Mu X, and likewise you setup for B star.

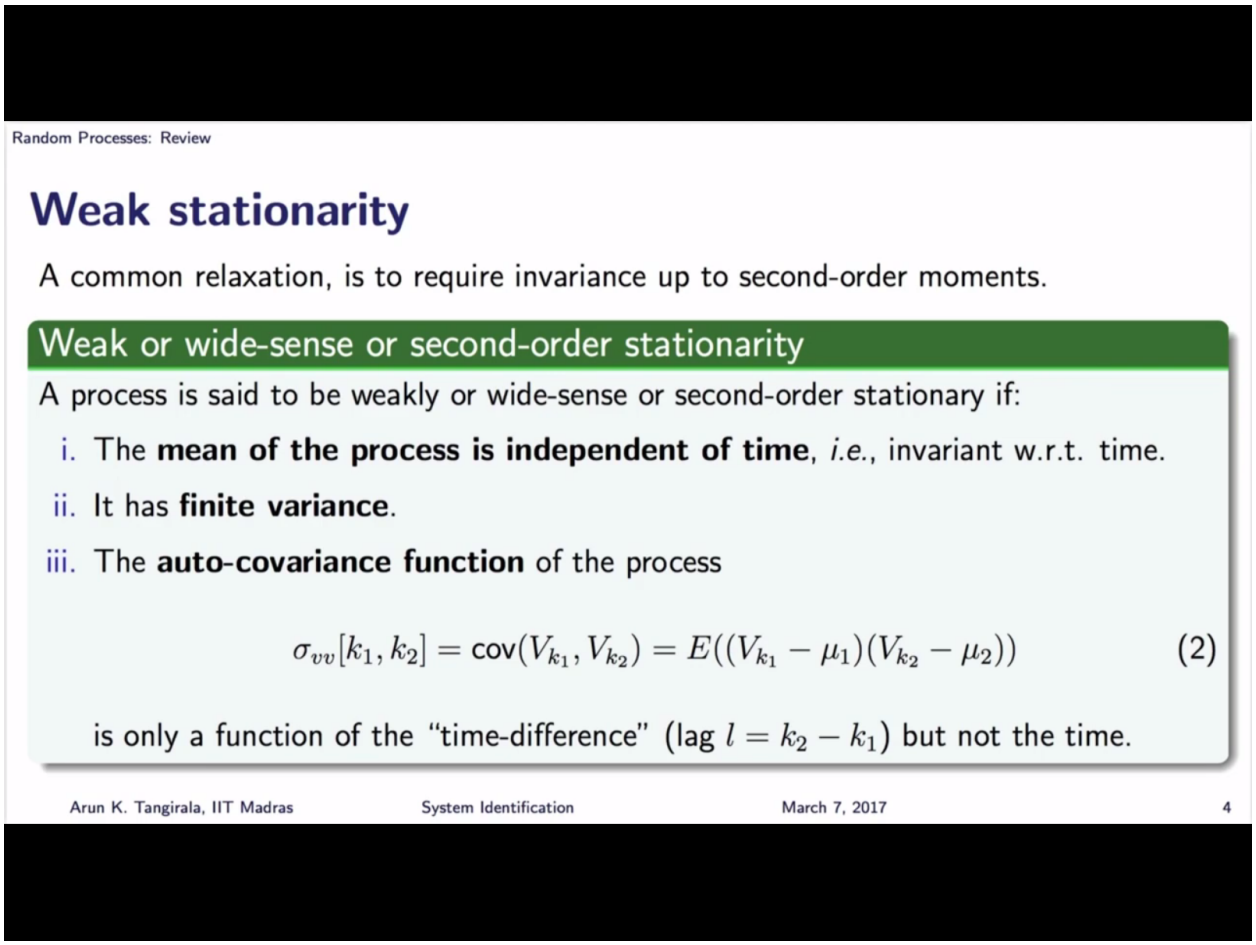
(Refer Slide Time: 25:42)



Anybody has any difficulty? What was the difficulty in arriving at this? Typically I think there is some fear in the mind, that oh I will not be able to solve this, nothing like that. You just follow things procedurally, you will be able to crack it. When you are in doubt, follow the steps. If necessary take baby stems, doesn't matter. Is everyone, I am asking this repeated, is everyone comfortable at arriving at the answer... with arriving at the.... Because it's extremely important that you are comfortable with taking expectations, this is what you will be solving in your assignments and may be your quiz and part of your final exam and so on. You should be really comfortable with deriving this theoretically optimal models. Clear, can I go... go ahead further, or anybody has any question? Fine, so I assume that all of you are comfortable. Now what we observe is that this optimal estimates of A and B are only dependent on the first and second order moments. Does the p.d.f. come out, you can say yeah, this first order moments are dependent on the p.d.f., that's fine, but explicitly does the p.d.f. make its way through. I only need to know the first and second order moments to estimate the linear model. Now although we have done this for the bivariate case, that is for two random variables, in general even if we were to extend it to the multi variate case, suppose, suppose I had X1 X2, multiple random... multiple regressors, even then... then the conclusion would be the same, you

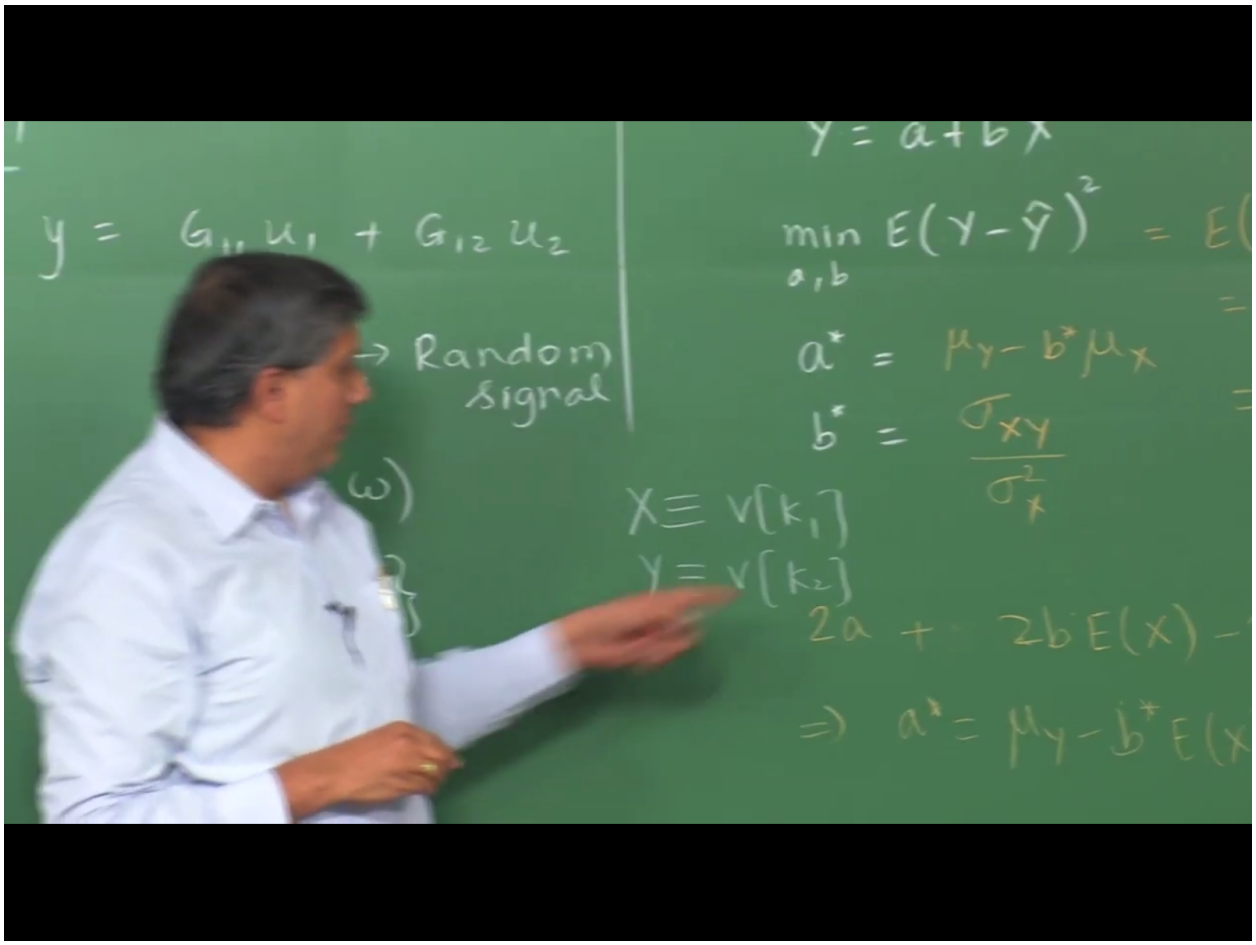
only need to know the covariance, it's up to the covariance, you need to know. Which straight away goes to show that as far as linear models are concerned, I do not need to have a knowledge of the p.d.f. I just need to estimate the covariances and the means, correct. Now the other thing that you should remember here is, this is a very generic problem. Y and X are some, two random variables, they could mean different things in different situation. So if I am looking at, let us say building a model for a random signal, single signal, Y could be an observation at the Kth instant and X could be an observation in the past, you understand? Because at each… each observation of a random signal is a random variable, so what Y and X did stand for varies with the application. Since we are interested in random signals, we can think of Y as some… of the random signal, the Kth instant or you can say K+1, doesn't matter. I am predicting using the past… predicting the Kth… at the Kth instant using may be K-1 K-2 and so on. In which case the sigma XY that you see in the numerator, which we call as a covariance, then becomes what is known as the auto-covariance, because you are now evaluating the covariance between… because we said X is some signal… the signal in the past, at some K-1 or K-2 whatever, instant and Y is the signal at the Kth instant. So you are looking at the covariance between the… between two observations of the same signal at different instants. So you are looking at what is known as internal correlation, which is also known as auto-correlation. There are many names to this, the more prominent name is the auto-covariance or associated with that is auto-correlation. Now we will get into that. So we will say here, that as far as linear models are concerned, it is… it suffices that the first and the second order moments remain invariant with that. And what we are essentially doing is relaxing the condition. We said oh p.d.f. doesn't have to be change… I mean it can change with time, so long as it changes in such a way that first and second order moments does not change with that. The other moments are most welcome to change, I am not worried about it, why, because I am going to build linear models, alright. So this weak stationarity says, the mean of the process remains invariant with time, the variant should be finite. Why should the variant be finite? Why we are requiring this? Look at the screen, look at the board, you will get the answer. If I treat X and Y as two observations at different instants of a random signal and I am fitting a model for the evaluation for predicting the signal, clearly it says that B star will depend on the variants of the signal, sigma square X would be the variants of the signal at some instant, that should better not be infinite, right. So it should have finite variants. And the numerator here.

(Refer Slide Time: 31:15)

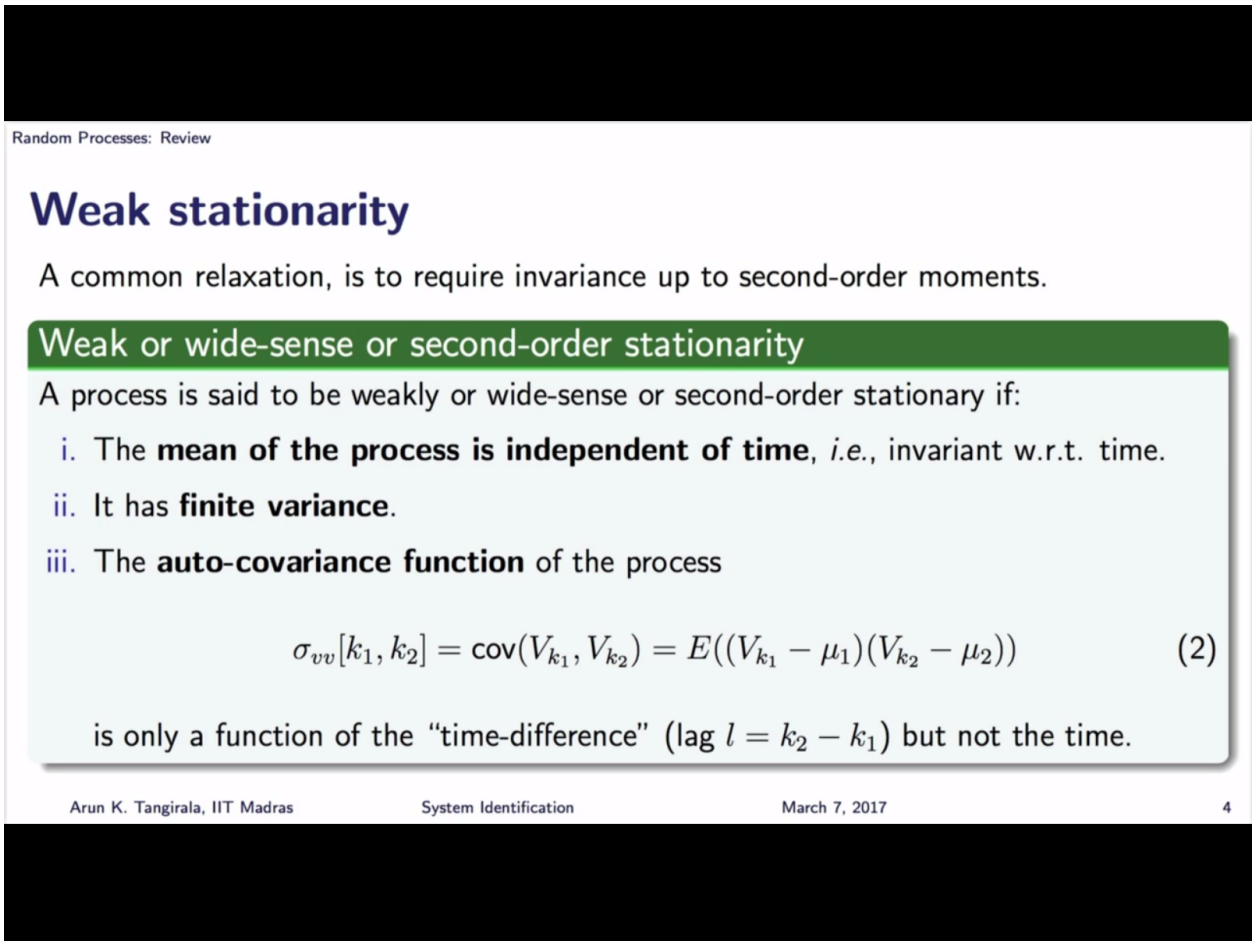


See what I am requiring is that this model that I built, based on hysterical data should hold good for future. What does it mean, A should not change with time, B should not change with time or A star and B star should remain invariant with time. When will A star and B star remain invariant with time? When the means are constants, when the variants is finite and when this numerator here is invariant with time. What is that numerator when I think of a random signal, it is the auto-covariance, it's a covariance between two observations, any observations at K1 and K2 instant, that should not depend on time. What should it depend on, it should only depend on the distance between the sampling instants. So if I think of X as V at K1 and Y as observation of the signal at K2, this sigma XY we call it, say auto-covariance as you see on the screen.

(Refer Slide Time: 32:18)



It should only depend on K2-K1, not on K1 and K2 separately. What does it mean?
(Refer Slide Time: 32:31)



It means that it should... the covariant should be such that, whether I look at the covariance between K1 and K2 or K1+T and K2+T, it should not be... it should remain the same. Otherwise, what does it mean, the internal structure is changing with time, all... that means the way for example V4 influences V6, if that is going to be different from the way V6 is going to influence V8, and there is something, then that means my B star is not constant. On the other hand the way V4 influences V6 can be different from the way V6 influences V7, but that's because you are looking at two successive observations, versus two observations, two instants apart. But what this third condition means is, doesn't matter where I am positioned in time, as long as I pick two observations that are separated in time by the same interval, their dependency should remain invariable, right. And that makes perfect sense, because you want... then only this model will remain invariant with time, you understand. Any questions? So that is the most important thing and we will of course dwell quite a bit on auto-covariance. This actually is opening doors to the auto covariance function. This we call as weak, stationarity, or wide-sense stationarity, or second order stationarity. There are so many different names associated with this. And we will assume hence forth that the process is second order stationary, when we say process is stationary, typically we will imply this kind of stationarity. There is another

form of stationarity as I keep saying called Quasi stationarity, which we will not worry about now, we will talk about it couple of lectures later.
(Refer Slide Time: 34:28)