Applied Time-Series Analysis

Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

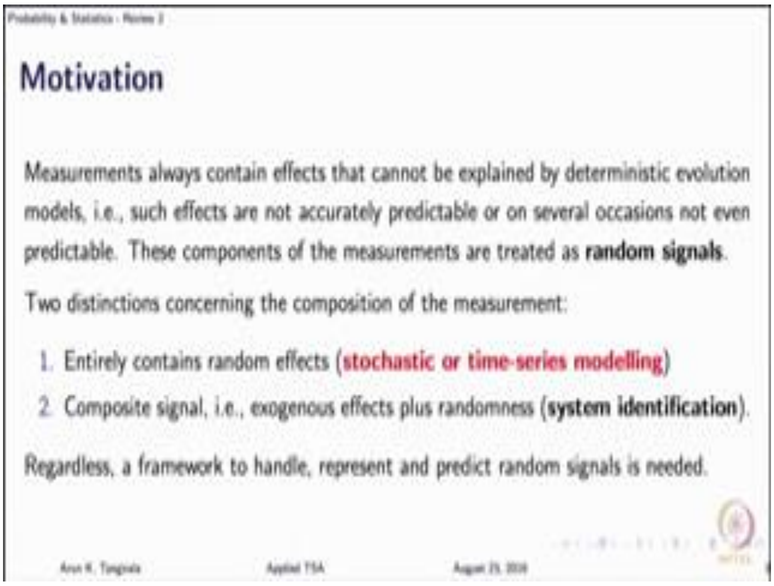
Lecture - 21
Lecture 09C - Introduction to Random Processes-1

Now, it is time to turn our attention to random processes, remember the reason for going through this review is we treat the random signal at every instant as a random variable. So, you have learnt how to analyze the signal at a particular instant in time. Now we have to learn to put together this entire chain of random variables and see how we analyze formally so that means, more complexities are awaiting, but all of that is founded on the concepts that we have discussed until now, therefore, you should be quite comfortable and thorough with this concepts. Any questions, before we move on to random signals any questions from that room in 2 4 1, any questions?

With this we will begin our journey into the world of random signals and random processes; they cannot be a better time than in the evening of a Tuesday, of course, will continue our discussion even tomorrow, fine. So, what will try to understand before we jump into these standard definitions? Especially the auto covariance functions and so on. Let us now understand what are random signal is formally what are random process is and then the notion of stationarity, strict stationarity, weak stationarity and also ergodicity then we will turn to the ubiquitous measure of auto covariance and auto correlation they gradually will build our theory on that.

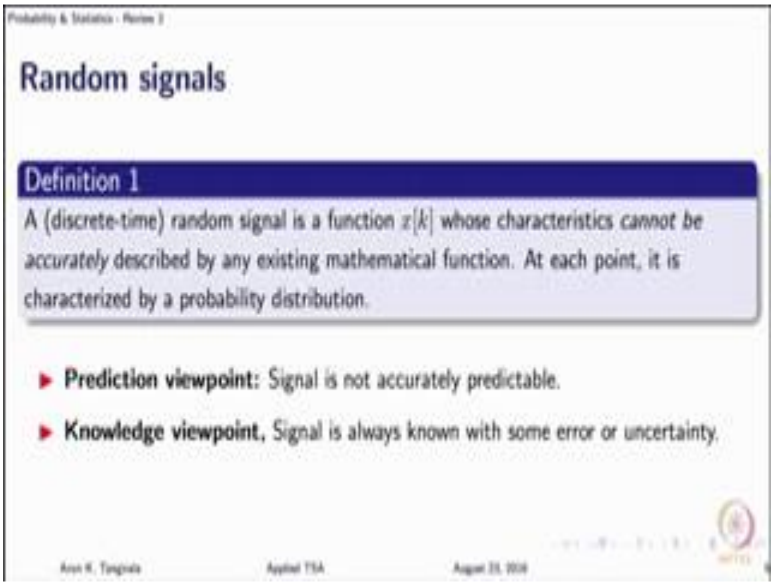
To recap again why we are looking at random signals because whatever I observe will always contain components that are that cannot be explain by known causes sometimes a causes are also unknown.

(Refer Slide Time: 02:23)



There are 2 distinctions as I have pointed out at the beginning of this course that when you look at a measurement, you can think of it to be made up of a deterministic or effects of some exogenous inputs plus a component that cannot be explained by any factors that we know or that there is nothing in that signal that I can explain using known effects I have to only actually relay on the history which is what is the focus of this course and that is essentially your category number 1 where we build time series models, in system identification we take up the second case that is where you have composite signals. So obviously, now whether we are looking at time series modeling or system identification, we need a frame work to first of all formally represent random signal, learn how to analyze it and then how to predict such signals.

(Refer Slide Time: 03:45)

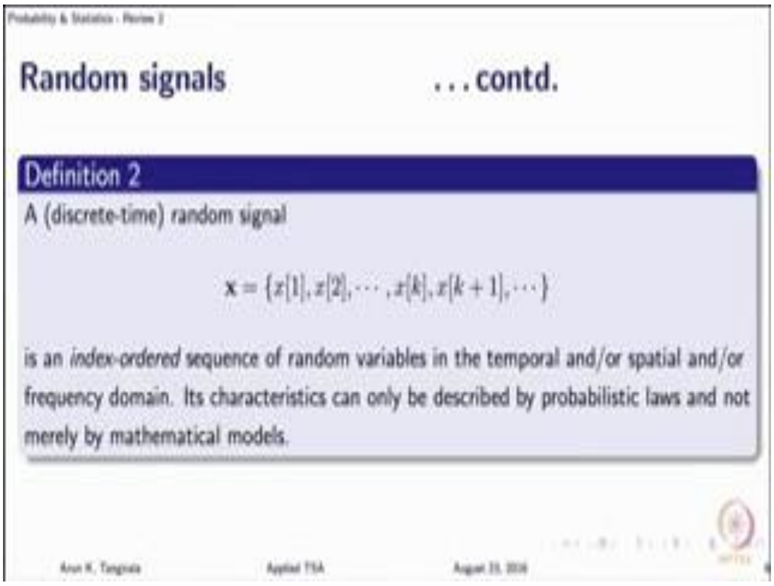


One of the standard definitions that you come across for a random signal is that it is nothing but a function whose characteristics of course, I am referring to a discrete time random signal here that nothing prevents us from looking at continuous standard random signals, but not of interest to us, we in practice we have sample data. So, I use x k here, but gradually will shift to v of k, I will probably change the notation in tomorrow's lecture itself. So, it is a function whose characteristics cannot be predicted by any known mathematical function and at each point the observation or the value of this function comes falls out of a probability distribution which means at each instant in time the random signal is a random variable.

Now, it is easy to remember this definition from a prediction view point for example, we can say that essentially the signal is not accurately predictable, there is no known mathematical function which means that I cannot predict accurately, it does not mean that I cannot predict at all, we have talked about that and from a knowledge view point that is when I am talking of a signal and I am talking of the observations in hand, what I mean by a random this observations being treated as a random signal is that every observation that I have is known with some uncertainty, there were you can say it is known in error or you can say that there is some uncertainty shrouding that observation that we have I am not so sure, if this is what exactly I should have obtained, I could have obtain something else that is what it means. So, whichever view point you take, essentially comes down to the same thing at each instant the value of the signal is a random variable.

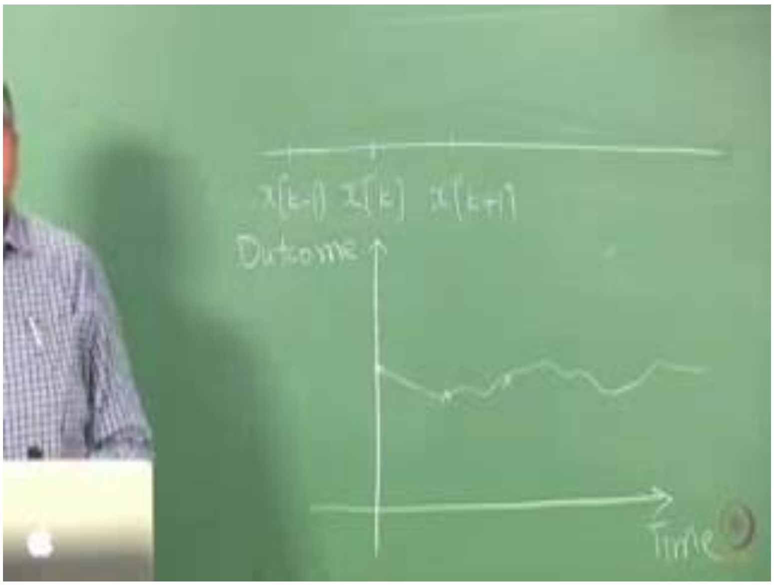
But the most important thing is that it is an ordered collection which is clearly reflected in the second definition.

(Refer Slide Time: 05:46)



Now you can say that a random signal is an index ordered sequence and that is the very important part of it.

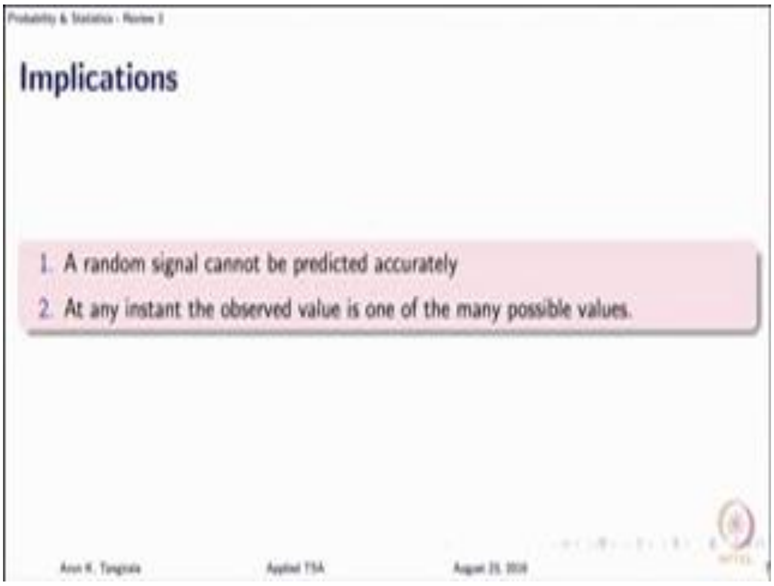
(Refer Slide Time: 05:56)



There is a thread that these observations are kind of hanging on to like the cloths are actually there is a thread in time and your variables for now I will use x, but very soon as I said will shift the notation. So, here you have x k, x k plus 1, x k minus 1 and so on. So, all these observations are hanging on to the same thread and what we intend to do in times series modeling is to see if there is indeed a thread that can I solve of them and then there is something that is not connecting which is like the unpredictable port of part of it, but the fact is that they are all connected by the thread of time, in time series modeling we explode the dependence is if they exist.

And once again this definition says that the characteristics of this random signal at any instant in time or even when you consider a collection of time instants, they can only be described by probability loss not there may be a mathematical function of course, but a mathematical function alone cannot describe the characteristics. So, what are the implications? It means that a random signal cannot be predicted accurately, we will remember.

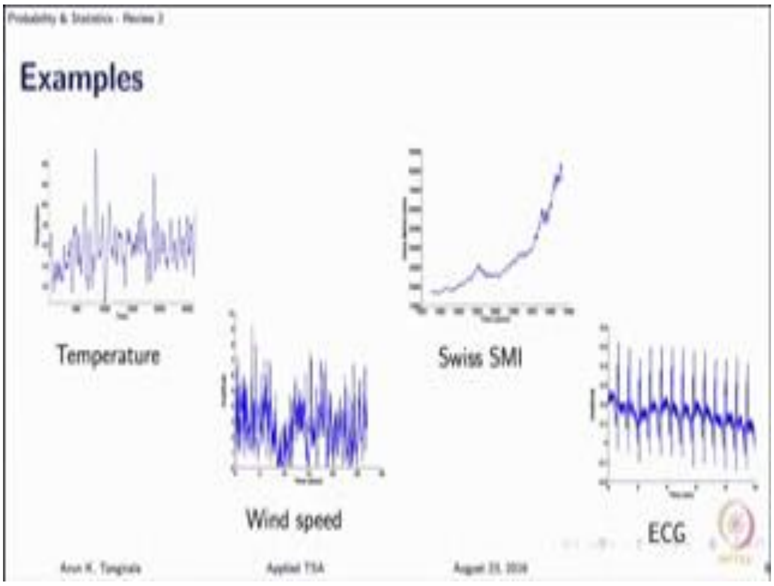
(Refer Slide Time: 07:11)



And at any point the observed value is one of the many possible values. Now what this means is if I have to stand at k here there are many possible values and this k here the value could have been here could have been this, could have been this and so on and then there is this dimension of time. So, here is your outcome or realization dimension whatever you want to name it, a random signal is now a in practice or sorry in theory it is a 2 dimensional signal, it is a function of the time or if it is not the time and space and so on and then there is this outcome dimension which is unique of a random signal then deterministic signal you do not see that.

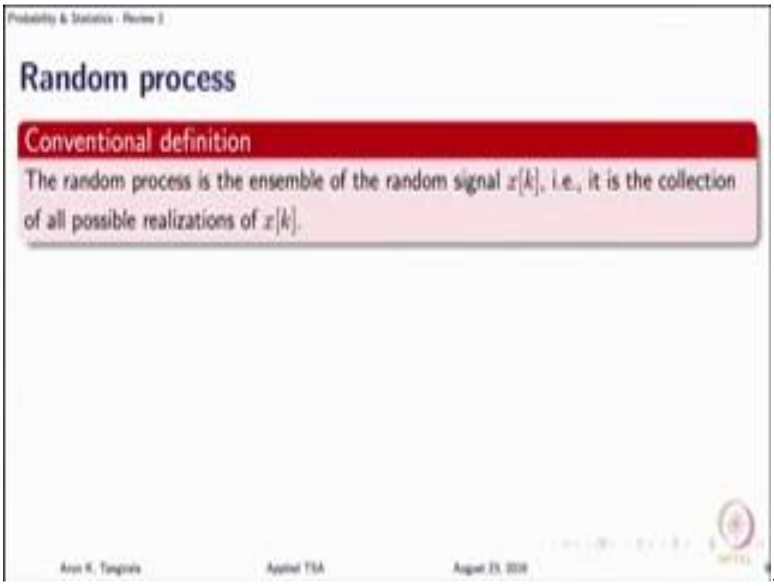
What we end up observing in general when I have a time series is some real (Refer Time: 08:09) at each case of here I have this realization than at x k plus 1 I have another realization and so on. So, I may end up actually seeing one realization of the many many possibilities I end up seeing only one and we have talked about this realization challenge, so some examples you have seen I can skip different time series. We have seen this in the introductory class.

(Refer Slide Time: 08:32)



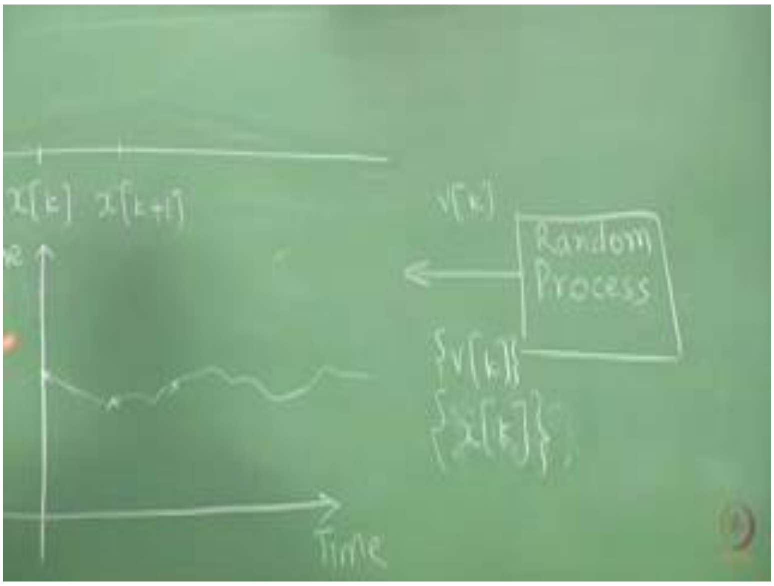
Now, we come to the notion of a random process, we have defined a signal, now we talk of a random process.

(Refer Slide Time: 08:34)



Something is generating the signal, there has to be a process that generates the signal a signal cannot exist by itself.

(Refer Slide Time: 08:51)



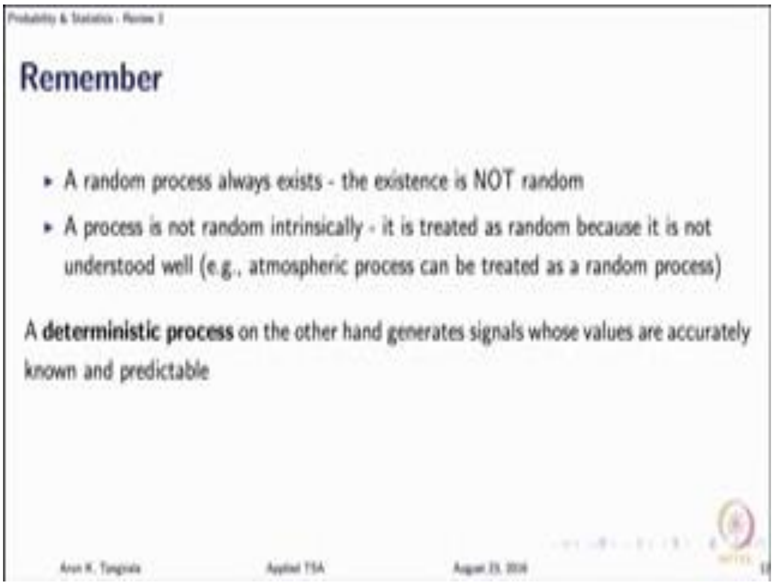
So, we think of this systems kind of prospective where we represent the process by block which is actually generating this bunch of random signals. So, one way of looking at a random process is it is that process which is generating all possibilities of course, of which I am only observing one, but this process is actually the ensemble of all possible collection this is the conventional definition that you will find in many standard and classical text books right, it see ensemble of the all possible realizations of the random signal.

Now, what this obviously means is that a random process or a signal is a 2 dimensional function typically you will find being this being written as x n sum omega well omega is a symbol will use of frequency later on, but very often in many text books you will see omega being used for representing the outcome which outcome that you are looking at or the realization that you are looking at sorry, k.

Normally we suppress the second dimension that is we do not talk about the outcome anyway we have only one realization. So, we suppress this and we generally use x k and you will see this different kinds of notations x k or tomorrow we may use v k and so on and occasionally you may also see simply v k 1 has to understand whether we are referring to a random signal or a random process, but you will see this slightly different notations typically for processes you would enclose in the curly braces indicating that you are looking at all possible realizations. So, all of this is formalization of concepts we you know keep this on the back of our mind.

Now, a simply definition is what I have drawn already, it is that process that generates all possible realization and what we are doing in time series modeling or time series analysis is trying to understand this process depending on what I want to do if I want to actually predict or forecast then I try to arrive at the mathematical equation that will allow me to build forecast or if I am actually doing pattern recognition or some kind of a feature extraction and classification then I am again looking out for some features all based on the single realization typically and that is the biggest challenge right. The reality is that I mean reality; we do not know whether it is actually random or not, but within the frame at that we are looking at this random process generates many possible realizations and I am suppose to find out what that processes from a single realization.

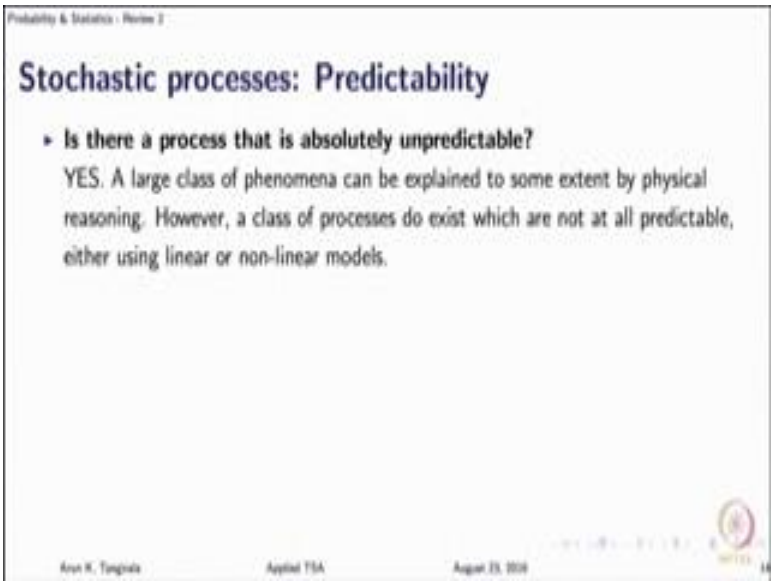
(Refer Slide Time: 12:06)



Now, some of the things that we want to remember is that a random process always exist, there is nothing randomness about its existence it always exist by definition a random process existed from times memorial and will exist forever because if it seizes to exist at some point in time then it is a contradiction. So, we will assume that it always exist engineering prospective is it exists for so long that 100 generations before me and 100 generation after me will never see would not have seen its beginning or would not be seeing its cession and again the process may not be random intrinsically it is only our ignorance that is making at looking forcing us to treat it as random that is forcing as to think that there are many possibilities out of which I have obtained one now a deterministic process in context is the one process that for which you can fit a mathematical function that exit and you know it and there is no ambiguity or uncertainty surrounding its state at any point in time.

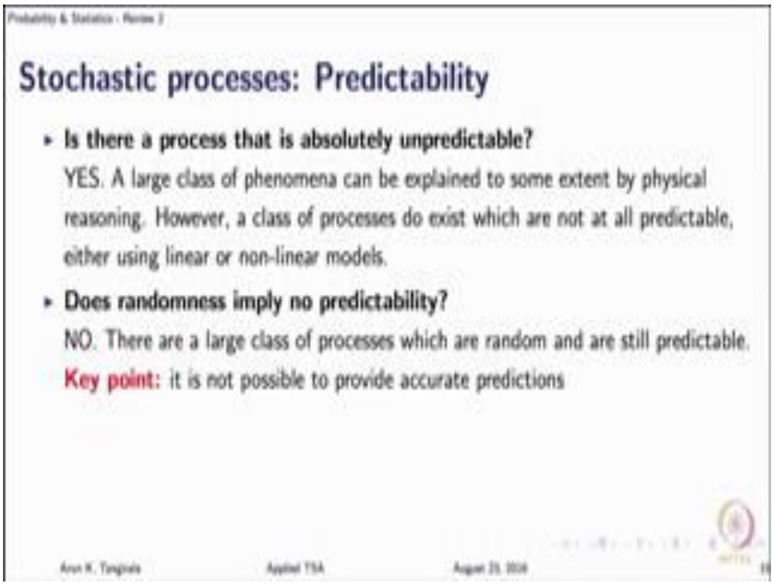
Now, of course, we will use the words random signal and process inter changeably there are 2 questions that we quickly want to visit before we move on and these are the 2 common question that students ask is there a process that is absolutely unpredictable? Yes.

(Refer Slide Time: 13:32)



I have being saying that random process does not mean you cannot predict, but they do exist random processes that are absolutely unpredictable, there are this real life processes that are in some sense, unpredictable, it could be in linear sense or in a non-linear sense. So, given whatever amount of data you will not be able to predict. So, you do not want to run into such a problem when you are working in a company, but when you are dealing with an assignment yes you want because you do not need to build a model.

(Refer Slide Time: 14:08)

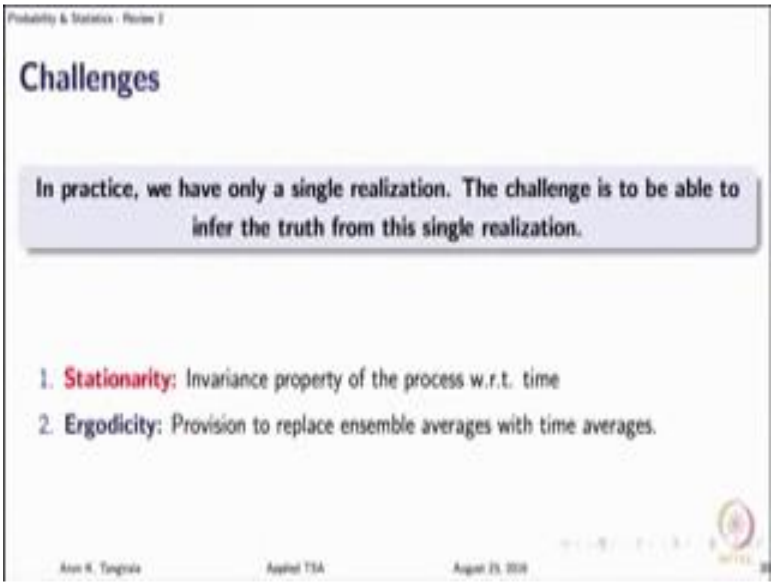


So, the other question that we want to ask is thus randomness imply no predictability, no we have already said that it is possible to predict, but we cannot predict accurately and our goal is to extract the juice right in sugar cane all of it is not juice of course, we do eat, but then still throw away what we cannot digest. So, likewise here we want to see if there is juice in the data to extract if there is in then will throw away the data I mean will keep aside the data and say this is nothing it is too difficult. So, the key point is it is not possible to provide accurate predictions.

Now, you will see on familiar websites like cora or some other website and so on whether there is a difference between random process and stochastic process this is like the confusion in nomenclatures? Well there are some historical things will not going to it, but in some parts of the literature you will find this answer that typically random process gives the kind of implication that it is not predictable at all and whenever there is an element of predictability we would try to label that as a stochastic process, but of course, you know all there is some historical account as I said depending on which country you are looking at, if you when you look at the history of this theory of random processes you will invariably run into Russian literature thanks to (Refer Time: 15:35) and then also some literature from you know (Refer Time: 15:40) and so on who was not in Russia.

You will see that there were also political reasons and then you know psychological things and human opinion as to why random why this word stochastic came up and so on will not get into all of that, but you can read that when for those of you who has said wrong registration probably go and have a time to read that. So, will skip this slide on deterministic process we have talked about it. Now in general any process that we go out see out there is not a deterministic process fortunately otherwise we do not have any job. So, the challenge in times series analysis we have a single realization.

(Refer Slide Time: 16:21)



And we want to infer the entire process that that is a process that generates all possible realization from the single one.

It should be kind of clear based on the foundations that we have given and also based on the description that when we set out to model this random process what we are actually seeking is to not describe or not to be able to predict that particular realization because that is an make sense by chance you have that realization if you had some other person observing or some other sensor observing you would have a different realization. So, the focus in random process modeling or stochastic process modeling is not on a particular realization, but on so called the average properties of this random process.

At any instant I am not so much worried about the value per say that I am forecasting, but I am more worried about the averages more worried about the variability. So, the statistical properties are of interest to me more than the values per say of course, in forecasting values are important, but when we are modeling we are more worried about the averages.

Now, as we had discussed in overcoming this challenge, we will be forced to assume certain properties of this random process and these are the 2 key properties that we will encounter which will stay in the background and keep coming up whenever things fail the first property is that of stationarity, we talked about it, it is got to do with the invariance of the probability distribution and the other property is ergodicity which has got to do with how we perform our measurements how we collect our measurements, how we perform the experiment and so on and generally we do not talk of ergodicity if the process is non stationary. So, what will do is in the next 2 3 minutes and then will wind up talk about stationarity briefly and then continue our discussion tomorrow on stationarity, non stationarity and ergodicity then will also introduce the auto covariance function.

Let us briefly talk about stationarity to the next few minutes, why are we looking at this notion of stationarity? We have already said that I am going to work with a single realization which is a collection of observations and I am going to build a model as an example I am going to collect temperature data in the morning let say on the campus between 6 and 7 am right or you can do this generally students are awake during night times. So, you can do this even in during night times any hour of the day favorite hour of the day you can pick and then let us say you sampled and you build a model, it is obvious that the model built let us say over 6 to 7 in the morning cannot be used, can it be used to forecast the temperature in the noon? Let us say peak time, why? Something is different; do you think we can use that model? Why can you spell out why we cannot use.

Student: The conditions that conditions that give raise to the (Refer Time: 19:59) change (Refer Time: 20:02)

What condition?

Student: Like temperature will be different (Refer Time: 20:04) there is sunlight.

Physically itself, the position of the sun is different that gives rise to probably varying you know heat transfer to earth correct. So, physically itself you see, but statistically correct your right physically itself the suns location is different. So, the kind of heat that earth receives is different and so on and therefore, you expect the conditions to be quite different, but statistically would you be able to describe what would be different between 6 to 7 in the morning and let us say 12 to 1 in the afternoon, anyone from that room?

Student: Scale of the variable and (Refer Time: 20:50).

Sorry.

Student: Scale of the variables is (Refer Time: 20:53).

What do you mean by scale, the range?

Student: (Refer Time: 20:56).

The range you mean.

Student: (Refer Time: 20:57) shifted (Refer Time: 20:59).

Anybody from that room?

Student: Sir the probability distribution function had changed.

And why would that change?

Student: Because of the physical causes like maybe at a given time like it may appear assume Gaussian at sometime, like there are mean and variance would have been change for some other time like in the morning. So, like you cannot say (Refer Time: 21:33).

Not bad, well all of you are trying to tell me the same thing, thank you. So, all of you are trying to tell me the same thing, but let us put it in a simple way that the set of possible values itself are changing like you said the range is different you cannot and let us hope that is not the case that between 6 to 7 in the morning, you get to you expect to see temperatures in the range 35 to 40 that would be brutal, already we are facing scotching temperatures. So, we do not expect to see that kind of temperature range at all when we talk of range what is a statistical measure that we are referring to.

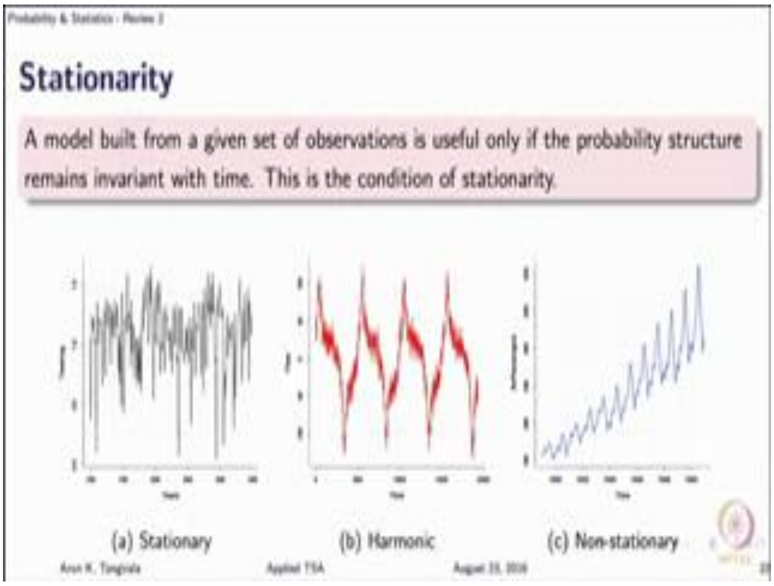
Student: Variance.

Variance, we are looking a spread of possible values and then the average also; obviously, you know when we are talking of range itself being shifted there is a likelihood not always necessary, but there is a possibility that the average is also different. So, which means that the model that I have built over one duration of my experiment is not going to be useful and that may be the case for many random processes. So, we want to know for sure that the model that I built from one set of observations is indeed going to be useful we will first restrict ourselves to those class of processes at least, we will learn how to model in such situations where the process does not pose such kind of challenges and that is the motivation for defining stationarity, very often stationarity is confused for steady state of the values, realizations are not going to reach a steady state, what is going to reach a steady state?

Student: The Statistical Property.

The statistical property, so, you see that is what exactly I have spoke about earlier when we speak of random processes, we are not talking about realizations per say, we are always we have frequently referring to the statistical properties. Here as came the answer from the other row we will fix initially, will come up with this idealism that the probability distribution should not change with time.

(Refer Slide Time: 23:38)



If you look at this examples here, 3 examples on the left side you have that is for the figure label stationary it is a tree ring data of the number of tree rings that you can actually get from trees, is it stationary, how do I say that when you should not believe me and just saying from visually there is nothing that is telling me that is non stationary, but that can be quite deceiving.

Let us think of it as an example of a stationary series, there are formal ways of determining whether some series is non stationary or not and so on and then in the centre you have a figure you have a time series which is so called harmonic, it is harmonic signal in the random world is nothing but a periodic signal, on the face of it thinking of a random signal as periodic seems to be contradictory, but do not try to apply the definition of periodicity in the deterministic world, straight away to the random signal world, remember we said for random processes, we do not really you know medal with the realizations in the sense we do not really work with the realizations we work with the statistical properties. So, now, you should be able to guess what periodicity means in the random signal world or the random process world the statistical properties have some kind of a periodicity not the realizations and we will talk about it later on.

And then on the extreme right, you have a time series which is the airline passenger data you will find this everywhere in many of the time series text, this series is an example of a non stationary signal you can see clearly what kind of non stationarity do you see? In other words which statistical property do you see changing with time?

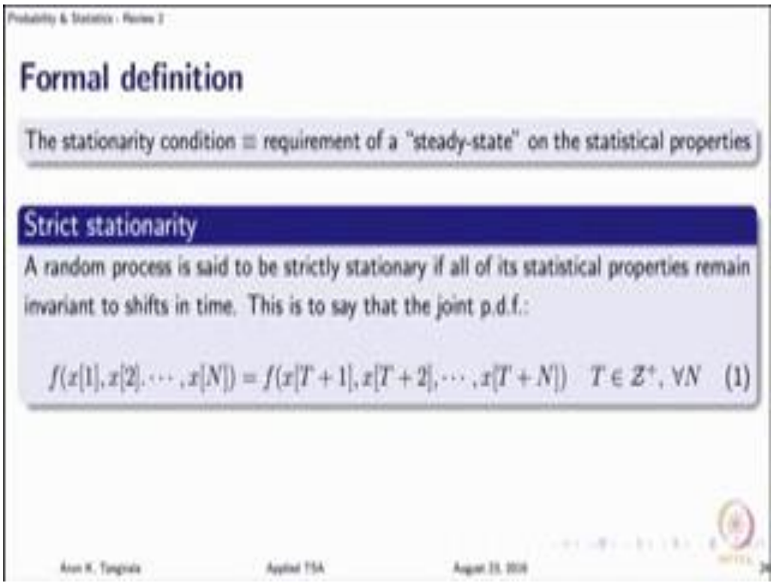
Student: (Refer Time: 25:38).

Mean.

Student: Variance.

Variance, both, the set of possible value seem to be changing, seems to be shift in time. Now having said this, this is not the straight away going to give us the definition of stationarity, we still have to be more formal, but we will first understand that stationarity means some kind of steady state on the statistical properties.

(Refer Slide Time: 26:04)



To begin with, we will say all statistical properties meaning movements should remain invariant with time and that can only happen if this joint p d f; this is the formal definition that you see and with this will end today. So, this gives raise to the notion of strict sationarity, it is a very ideal kind of definition, what is it saying? The joint p d f; why are we looking at joint p d f?

Student: (Refer Time: 26:33).

Because we are going to analyze the collection of observations, a set of observations so, we want to say that this joint p d f of the n observations that I have collected, does not matter I begin at t1, look at the statement it says there are 2 things, the joint p d f should remain invariant with 2 things, what are those 2 things? When I start collecting and how many I collect. For all t that is shifts in time and for all n although I say for all t and for all n its understood that it does not matter where I begin and it should not matter as to how many observations I am looking at jointly which means it should be even valid for a single observation, every observation should follow out of the same probability distribution imagine that at every instant there is this big bag of possible values all kinds of colored balls whatever you want to imagine, but there is this big bag of possible values and that the possibilities and their distribution should not change with time.

Do not you think it is too stringent requirement, it is I mean we do this kind of idealism we adopt this kind of idealism in every subject in thermo dynamics with begin with an ideal gas, but there is no ideal gas correct and in every situation you know when you get married you think what all idealities you soon realize that in nothing like that I do not know some of you already realized, but anyway. So, this kind of idealism is necessary to get into this world of random signal analysis and then slowly will relax this will relax this condition recognizing the this is 2 stringent, but it is good to start with this will say you know I wish the process was like this, but it is not.

Tomorrow will talk of more relaxed requirements of stationarity and then move on to ergodicity and auto covariance.

Thank you.