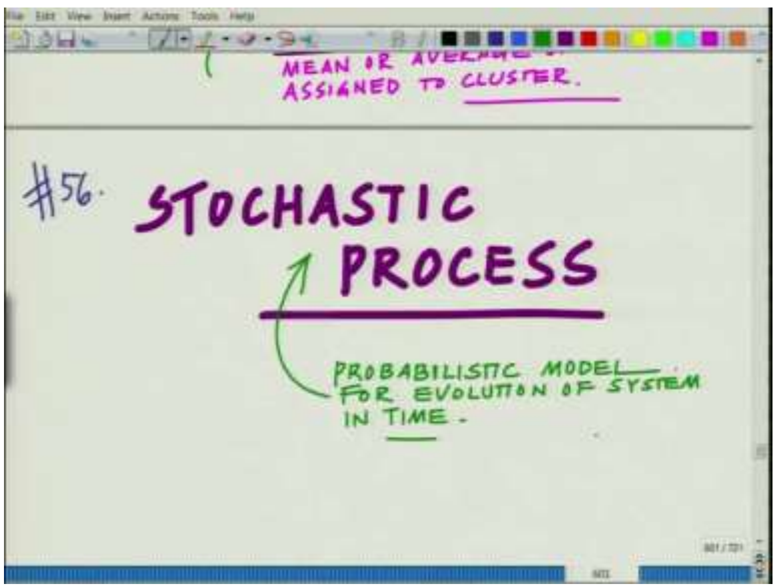
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor. Aaditya K Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur
Lecture No. 56
Introduction to Stochastic Processes and Markov Chains

Hello, welcome to another module in this massive open online course. So, in this module let us start looking into another very important aspect where one can employ the principles of linear algebra that is vectors and matrices and so on, including probability and the notion of probabilistic evolution, and that is in the very important area of stochastic processors. How does linear algebra help us better understand the evolution of something that is random with respect to, random in nature with respect to time.

(Refer Slide Time: 00:47)



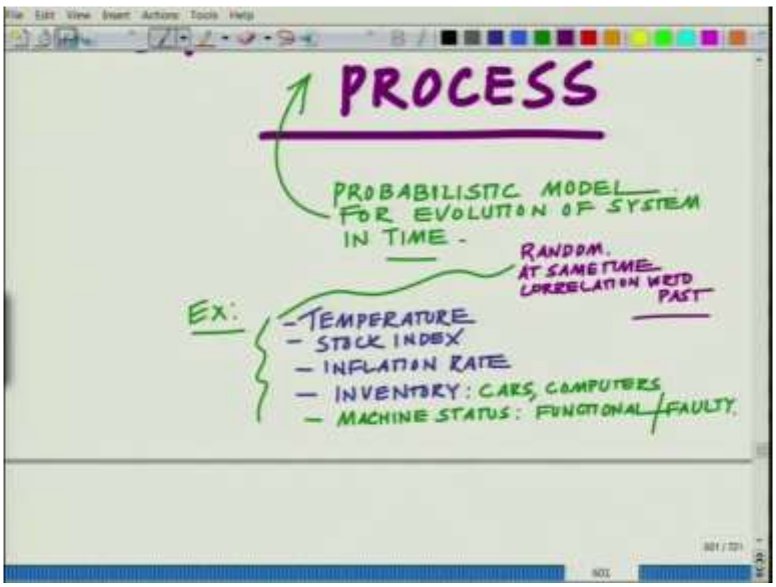
So, we would now like to understand the evolution of stochastic process and what is the meaning of that. This is the probabilistic model to describe the evolution of a system in time. So, this is essentially a probabilistic model for evolution of the system in time.

And for instance example one can think of several of these things that are evolving in time, for instance one can think of the temperature of a place, one can think of things such as the stock price, how is the stock price changing with respect to time, for instance the stock index, how is it evolving with respect to time, definitely there is nothing random about it, and the inventory for instance, the inventory of certain objects or the inventory of cars in a warehouse, the inventory of any number of things, the computers or phones at a shop and so on and so forth.

One can also think about the status of a machine, for instance you have an industry with a large number of operating machines, what is the status of this machine, how is this evolving, is it operating or is it broken, in the sense that does it need to be, is it faulty does it need to be repaired and so on, so how are these?

So, if there are so many things around us that are, who's status or who's state as we might say is changing randomly with respect to, and how do we characterize the evolution so that in the end of it we would like to make better predictions about how these are going to behave in the future and these have definitely a lot of use.

(Refer Slide Time: 03:11)



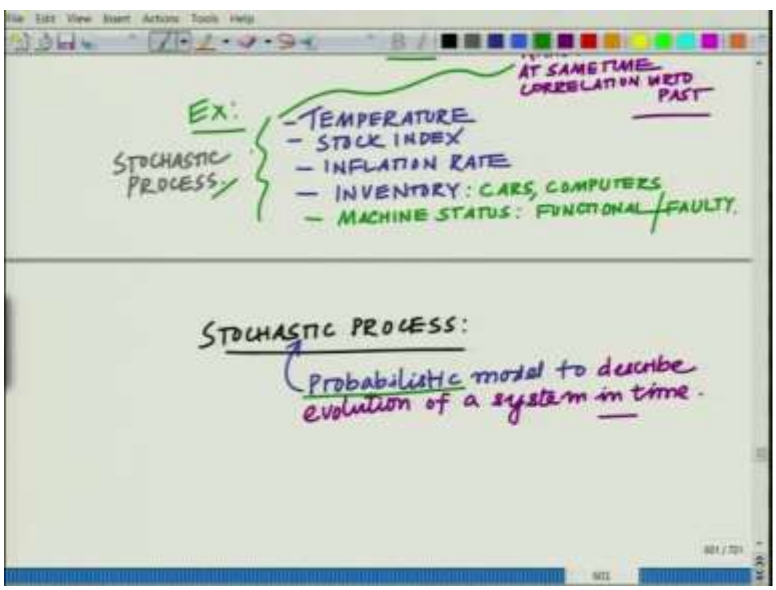
So, essentially the way to look at it is you have a large number of random things such as for instance you have the temperature, you have the stock price or you can say the stock index, other financial metrics such as the inflation rate how is this evolving? Then you have the inventory for instance, you can have the inventory of objects such as for instance cars in a warehouse, computers, in fact any number of items that are essentially being sold.

Then you have the machine status, functional or is it faulty. If it is faulty the it needs to be repair with what, and what do you think how is it going to be evolving, how is the status of this machine going to evolve and so on and so forth. So, all of these quantities if you can see, now the characteristic of all of these quantities is that these are random in nature but at the same time there is a certain sense of correlation with respect to the past. What we mean by that is? That is of course one cannot make it completely, one cannot completely determine with 100 percent accuracy what the temperature is going to be tomorrow.

But based on the past history, based on how the temperature has evolved in the past couple of days, one can make a reasonably accurate prediction regarding what, therefore regarding what the temperature is going to be in the near future for instance the temperature for the past couple of days is around 30 degrees, let us say 35 degrees 36 it is varying around 35.

It is highly unlike that the temperature tomorrow is suddenly going to jump something like 50 degrees or something of that sort. You expect it to be, it is random we cannot completely determine it but at the same time, one can get a reasonably good idea regarding its next, regarding the next step one might say or the next state in this evolution based on the history and that is the key over here.

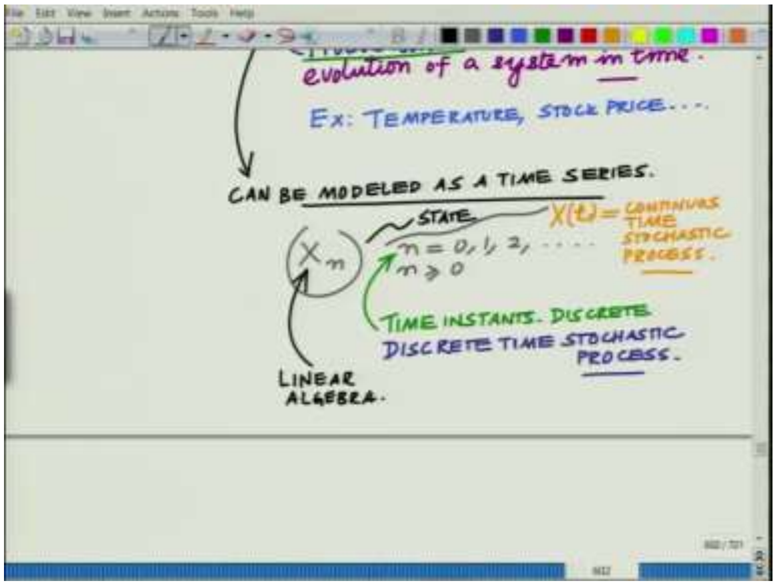
(Refer Slide Time: 06:12)



So, essentially the idea is, this is, all these are essentially what we call as stochastic processors. So, strictly speaking a stochastic process, what is stochastic process? So, a stochastic process is essentially, this is a probabilistic model, what is this? This is a probabilistic, to describe the evolution of a system in time.

This is a probabilistic model to describe; the key here is that it is a probabilistic model. Please understand this. It is a probabilistic model to describe the evolution of a system in time and all of these for instance. If you look at all these that is your temperature, stock price, inflation rate, all of these, these are stochastic processes. These are the examples of stochastic processes.

(Refer Slide Time: 07:47)



Now what we have just looked at for instance you have the temperature, the stock price, etc. Now the point here is, now this stochastic process can be modeled, this can be modeled as a, a stochastic process can be modeled as a time series that is we can denote it by for instance Xn that is this as for n equal to 0 1 2 so on that is n greater than equal to 0.

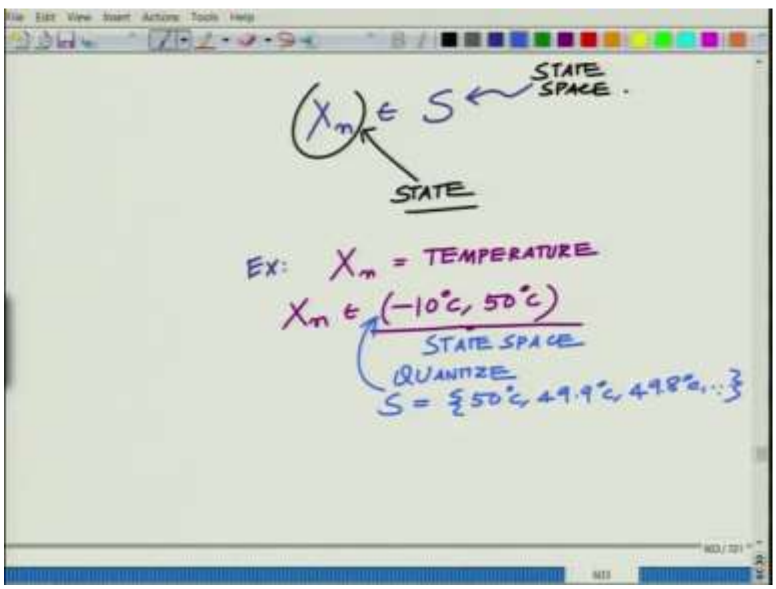
And more specifically this is essentially what we call as the state of the system or state of the process and these are nothing but your time instants, your Xn these are essentially the time instants which you can clearly see these are discrete in nature over here. So, this is essentially what is known as a discrete time stochastic process or you can call it is a discrete time chain or essentially a discrete, for instants one might look at what is the, n can be anything for instants unit of n, n can be either hours, minutes, seconds, days and so on.

For instance, if you are looking at the stock index, you would probably like to look at the opening stock index every day or you probably like to look at the closing stock index at every day that would be a discrete time. So here we have X0 on day 0, X1 on day 1, X2 on day 2. You would like to probably look at the inventory of cars on an hour by hour basis. So that would be X0 at the zeroth hour to begin with X1 after hour 1, X2 after hour 2 and so on.

So it depends, the time index, the unit of time can be anything, so this is basically happening from time, and its time to time and you can also have a continuous time stochastic process, but of course that is a different topic of discussion, that is evolving continuously in time that is X of t, that is instead of X of n, if you have X of t, then this becomes a continuous time stochastic process.

For us we will look at discrete time stochastic processes, because these can be very conveniently analyzed using liner algebra and matrices. Same can be done also for continues time, but it is a little bit more complicated, and we want to look at the applications of liner algebra in the analysis of these discrete time stochastic processes.

(Refer Slide Time: 11:33)

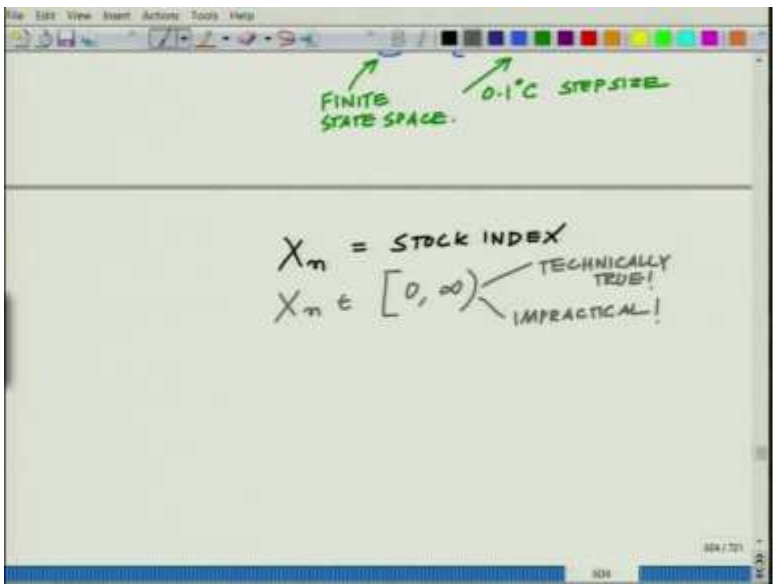


And now this Xn cannot take any arbitrary value but the key here is Xn can take values belonging to a set S or a space S and this is termed as the state space. So, X n can, is essentially the state out of the system. The notion of a state is very important over here, so you have the state which is given by Xn and this state can belong to a set S which is essentially the space of states or we will call it as the state space for instance once again.

Let us take X and let us take a simple, example Xn is the temperature like say Xn equals the temperature. Then 1 can say realistically Xn belongs to minus 10 degrees to 50 degree centigrade, although in principle it can be really low, go to the absolute 0, but that is, in practice typically that is not there for instance in a typical place you have temperature going from minus 10 degree centigrade to 50 and that is reasonably good enough.

And you can further discreticize it, for instant we might not be interested in measuring it to a 10 to the power of minus 6 or 10 to the power of minus 7 accuracy. We might be simply interested in measuring the temperature to point 1 degree centigrade accuracy, that might practically be sufficient. So, you can farther refine this, so let us say this is your state space. You can further quantize this for convenience and think of your state space as the set, for instance 50 degree centigrade, 49.9 degree centigrade, 49.8 degree centigrade and so on.

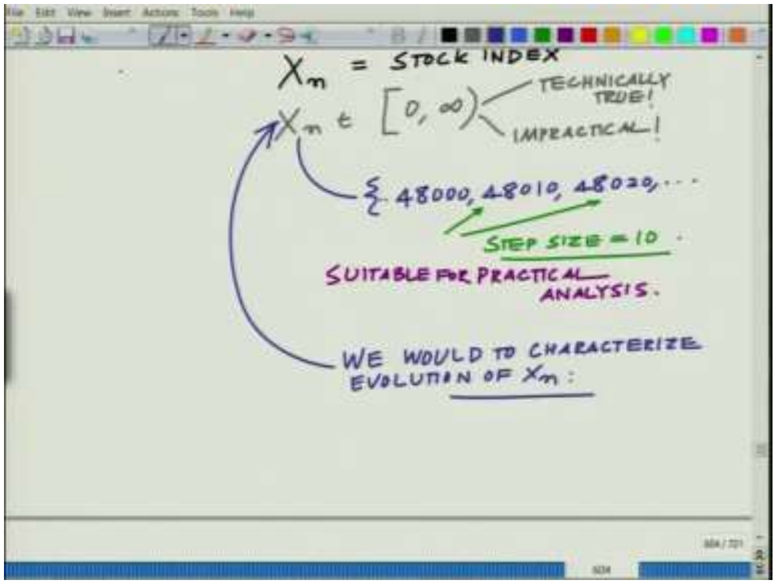
(Refer Slide Time: 13:49)



So, this is essentially a 0.1 degree accuracy, accurate to 0.1 degree centigrade that is what we call as the step size and so on and so forth. So, this, you essentially have a finite state space, that is the idea. So, of course this is, the original one is an infinite state space because it can lie anywhere between minus 10 degrees to 50 degree centigrade. From that by quantization you get a finite state space.

Now similarly, for the stock index, let us say another example, let us say Xn is the stock index then technically again Xn can lie anywhere between 0 to infinity including the value 0 this is, remember this is technically true but at the same time I would say this is impractical. Because we know the stock index typically takes values around a certain neighborhood where it is. So, let us say, and we can further quantize it.

(Refer Slide Time: 15:16)

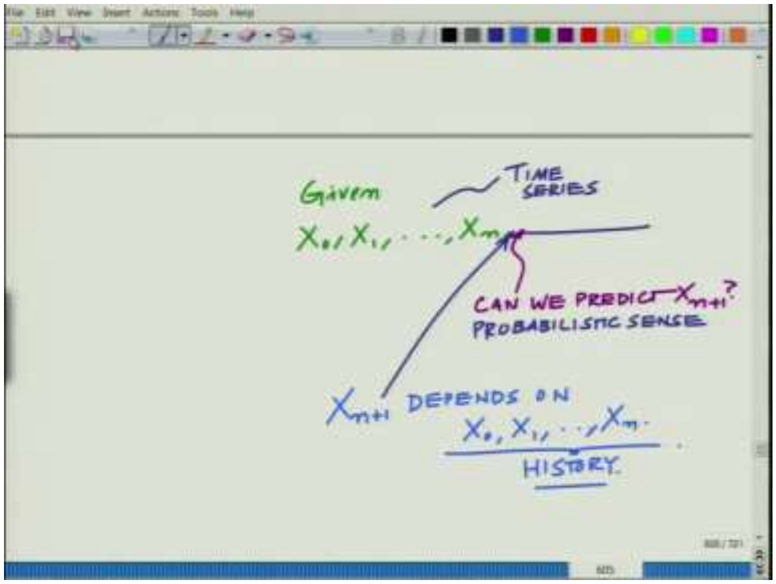


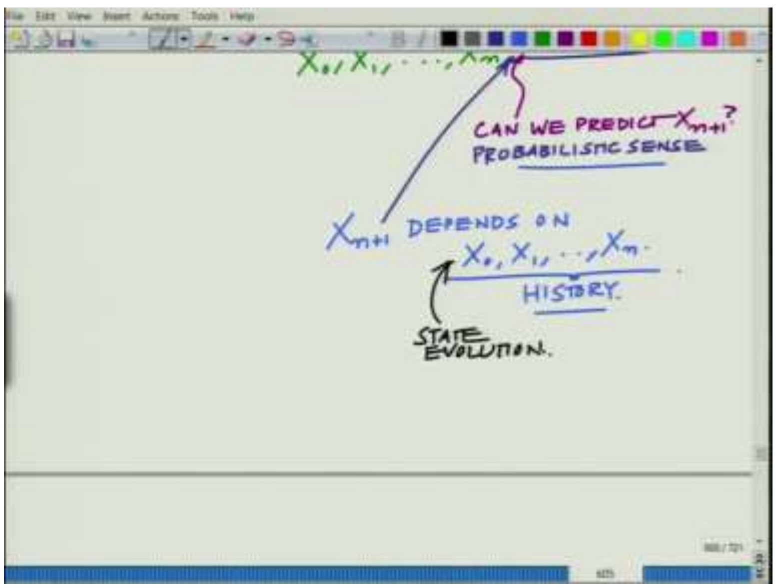
For instance let us say, we can say the stock index belongs to the set 48,000, 48010 steps of 10, 48020 and similarly you can take lower values also. So, you can take lower values also and this might be reasonably good enough. So, step size, and this might be very suitable for a practical analysis or practical case study because if you start with, so this is suitable for practical analysis. So, you might start with infinite state space but one can reduce it to a reasonable practical finite state space that makes it easier to study and make deductions in practice.

And the steps size chosen can be reasonable, so that your deductions might not be, it can be fairly accurate to the one that you would obtain using your infinite state space, as the step size goes smaller and smaller naturally you will have a large number of states. The granularity becomes finer and naturally this model becomes closer and closer to the actual one with an infinite state.

So, that is the idea, and now the point is we would like to characterize the evolution of this Xn. So, how to characterize? Now we would like to, the idea of this stochastic process is we would like to make comments or essentially we would like to make intelligent comments, we would like to characterize the evolution of Xn.

(Refer Slide Time: 17:43)





So, given the time series, more specifically given X0, X1 up to Xn let us say. The question we want to ask is, now this is a time series we all know, what is a time series, basically series of, observations or series of signal samples in time. Now given this time series now can we predict for instance and can, of course it is impossible to say Xn plus 1 is exactly going to be this.

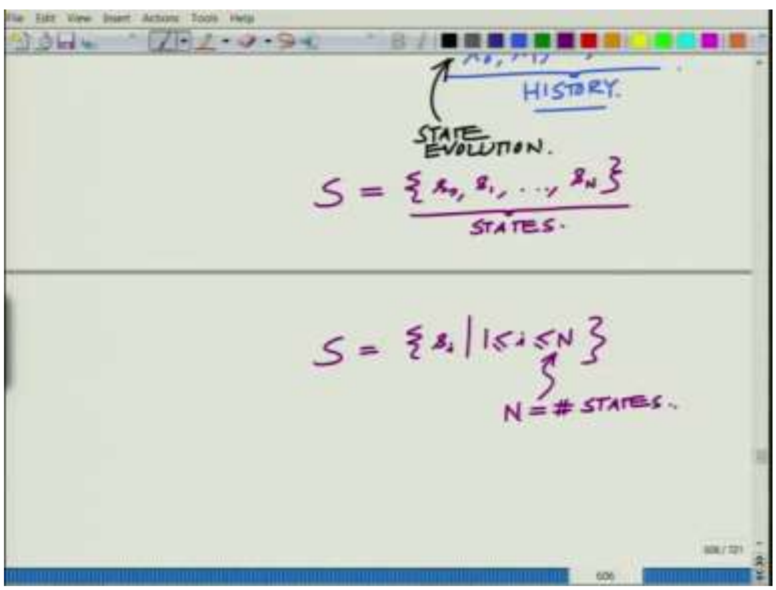
But can we predict with reasonable accuracy in a probabilistic sense that is the idea. Can we predict this in a probabilistic sense. I will just, I will simply, I will characterize it going forward what we mean by this. Now of course, the key idea here is that, it depends on Xn plus 1 depends on X1, X0, X1, up to Xn. This is essentially what we term as the history of the evolution, the past.

So, although Xn plus 1 is random, strictly speaking but the past evolution can tell us something about Xn plus 1. That is the wallet. So, how can you predict, the question, the more interesting and probably logically sound question or the question that can help us better characterize the Xn plus 1 is the following.

That is can you say something about Xn plus 1 having observed X0, X1, up to Xn, of course if you want to have to make an arbitrary prediction, then one can also make that but that is going to generally be way off but rather in many situations one is able to make better predictions because one, evolution and that is what we want to essentially capitalize on. So, what can you say about Xn plus 1 having observed X0, X1 up to Xn?

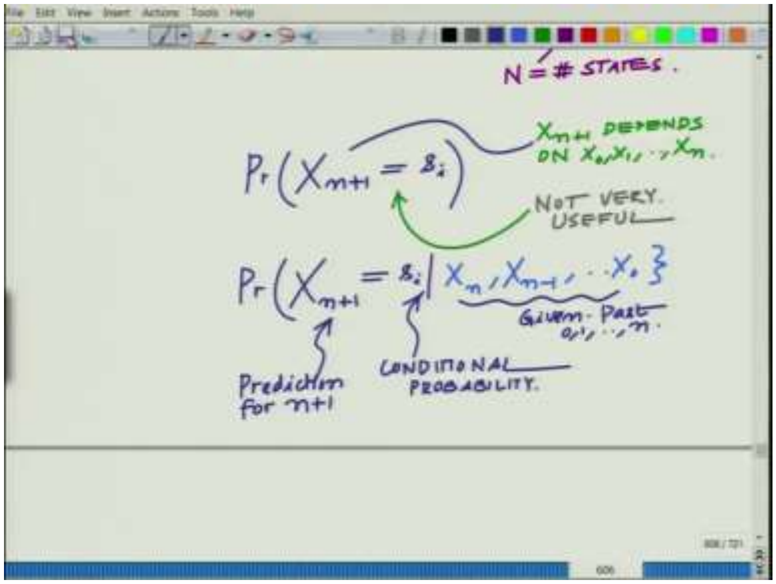
So, can we talk about this, can we make this prediction in a probabilistic sense. So, now consider the states. Now since we are talking about the state evolution remember this is nothing but a state evolution. And we have the time instance 1 up to n.

(Refer Slide Time: 20:50)



Now let us say S has the states. It is finite state space. These are the states that is you can say S is equal to si, such that 1 less than equal to i less than equal to N where N is the, capital N is the number of states. This capital N is the number of.

(Refer Slide Time: 21:33)



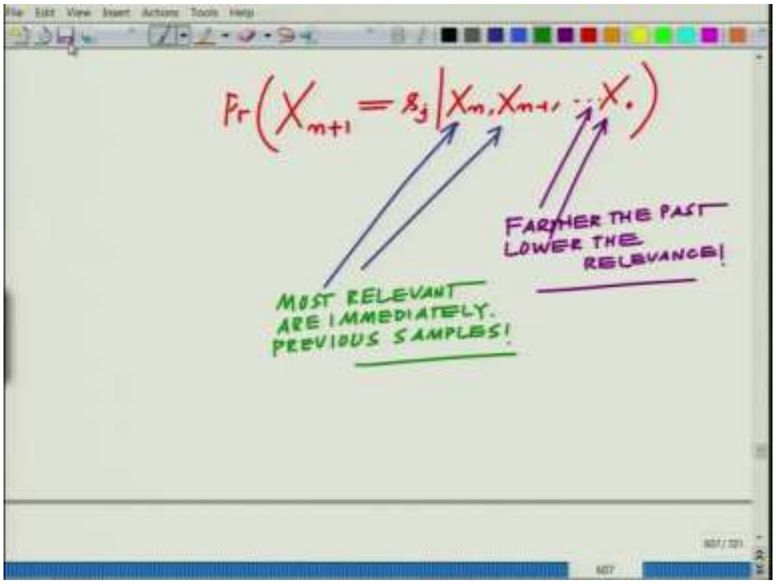
And now one can ask the question. What is the probability Xn plus 1 equal to si, this is the question that one can ask but this is not very meaningful. Because Xn plus 1 depends on X0, X1, up to Xn, so this question itself not very useful it is not, it is meaningful. I would not say it is not meaningful, it is not very useful to begin with alright.

One can also ask this question no doubt and one can also answer this question as we are going to start with but it is not very useful because it is, one can look at it as the following because you have a past history and you are not capitalizing on that past history, you are not using it to your advantage.

So, the more useful and interesting question that can allow you to better answer this question is the following, which is what is the probability Xn plus 1 equal to si given, remember there is a conditional probability, given Xn and remember I am writing this in a particular order, I am writing it in inverse order of time given? Given.

So, remember there is a conditional probability. You might have seen this before that is given these past. What is prediction for time instant n plus 1? That is essentially what we would like to, given the past evolution of this what is your prediction regarding Xn plus 1 and of course here you can see for any such time process.

(Refer Slide Time: 24:24)



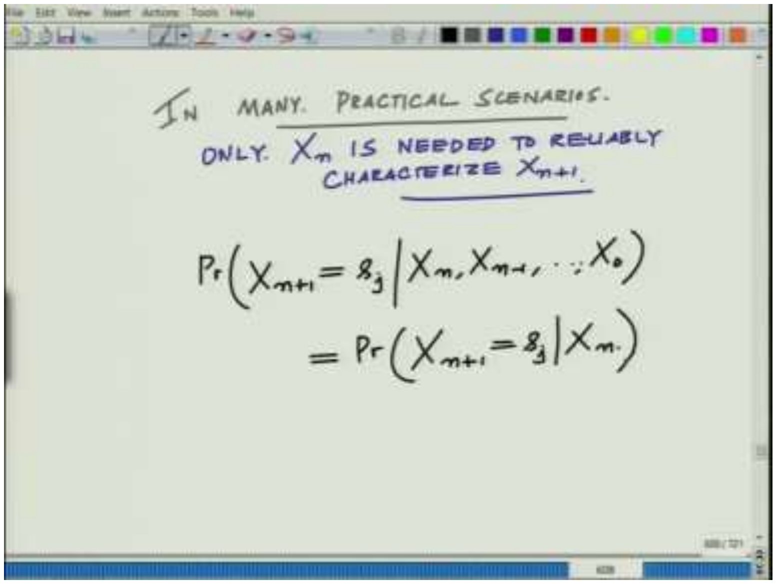
Now the one thing that you have to observe is interestingly, the farther the past, the less relevance it has to Xn plus 1 although the past is relevant, probability Xn plus 1 equal to sj. Now the point you have to ask is farther the past. So, for instance these samples are farther in the past. Farther the past, lower the relevance and this is the interesting aspect. Because of course the entire past is relevant but speaking about the degree of relevance for instance if you are looking at the stock price evolution, you can consider the entire past of the stock index.

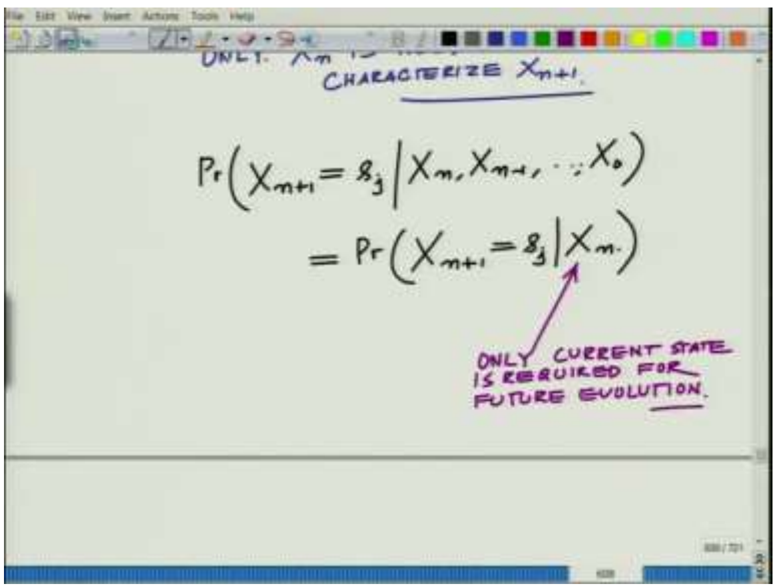
You can consider it yesterday, day before yesterday, 1 month ago, 1 year ago, 5 years ago, 10 years ago and so on. But realistically speaking how much influence is the stock price that is let us say 10 years ago going to have on the stock price tomorrow. So, you can see that there is, although there is infinite memory to so as, in this system so as to speak because it depends on the past but the farther you go into the past, the past relevance is lower and lower.

The immediate impact, the greatest relevance, the greatest impact is basically of the immediate previous samples. So, the greatest relevance here you have to understand, most relevant are the immediately previous samples, we would like to say Xn, Xn minus 1, may be a couple of time delays. Of course if you go to Xn minus 30, Xn minus 40 they might also have an impact but the degree of impact, the level of impact on Xn might be significantly lower.

In fact in an extreme situation on the most useful scenarios and one of the most tractable cases is where Xn depends only, significantly only on Xn minus 1, so let us characterize that aspect.

(Refer Slide Time: 27:22)

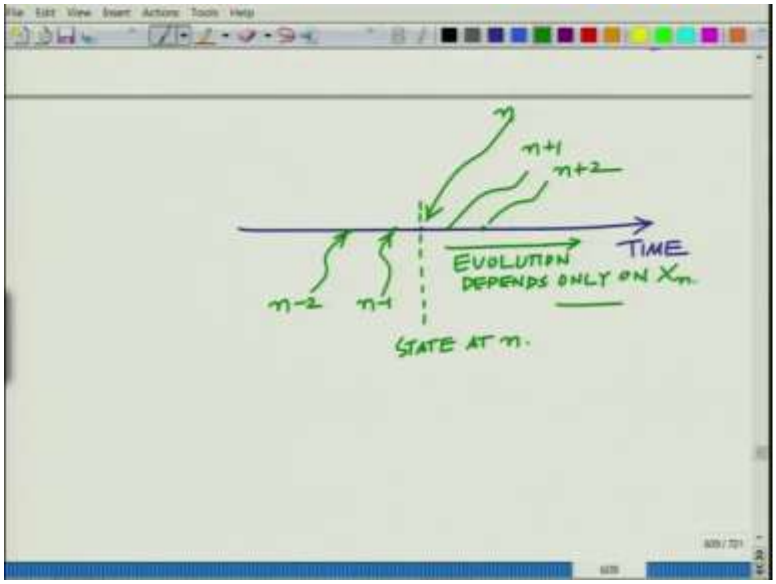




So, let us say in many situations we have or we are reasonably happy looking at the immediate, dependence on the immediate past in many I would say practical scenarios. The probability, what happens only Xn is needed to reliably characterize Xn plus 1 that is we would like to say probability Xn plus 1 equal to sj given Xn, Xn minus 1, X0 this is equal to probability Xn plus 1 equal to sj, given Xn.

That is only the past, only immediate past or only current state we would like to say. Only current state is required for future or required to characterize. Only current state is required to characterize the future evolution that is, we require only knowledge of X and Xn rather than Xn minus 1, Xn minus 2, so on to characterize how it is going to evolve from this point onwards. That is the future, that is the current state serves as a partition or you can say current, stands as a decoupler.

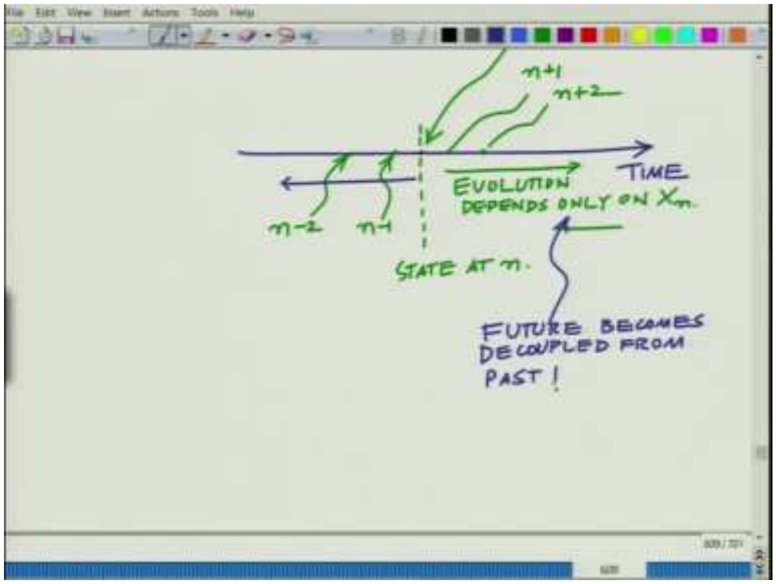
(Refer Slide Time: 30:09)

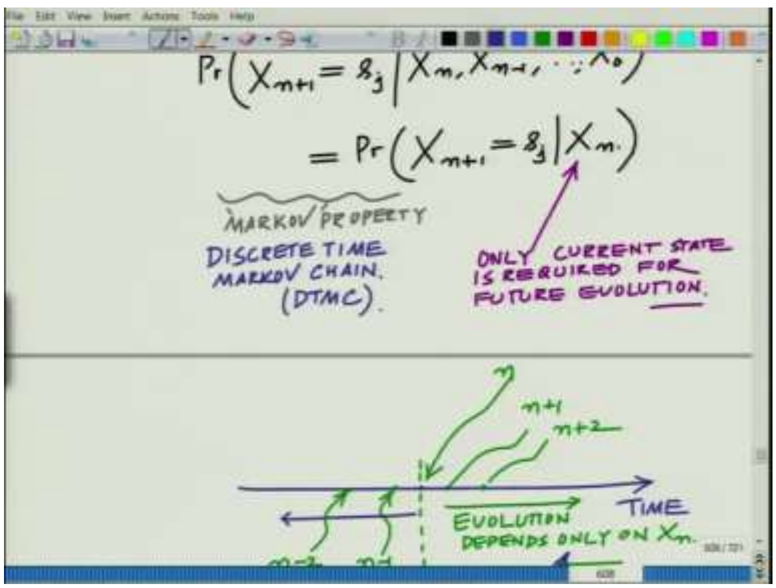


So, essentially if you look at as the time evolution of this system, so you have a system that is evolving with respect to time and here you have time instant, your time instant n and then you have the n plus 1, then you have the n plus 2 and here you have the n minus 1 and you have the n minus 2 and this property what it says is that if you know the state at time instant n.

So, we know state at n, then the prediction of evolution so far, the prediction that is the evolution, henceforth the evolution depends only on Xn that is, the future is decoupled from the past that is, if I know what is, why state at time instant time and I do not need to know about the past, I do not need to know what is the past history, I do not need to know how the system arrived at the state Xn because having now known Xn the future depends only on Xn irrespective of the past, irrespective how it arrived at the state Xn.

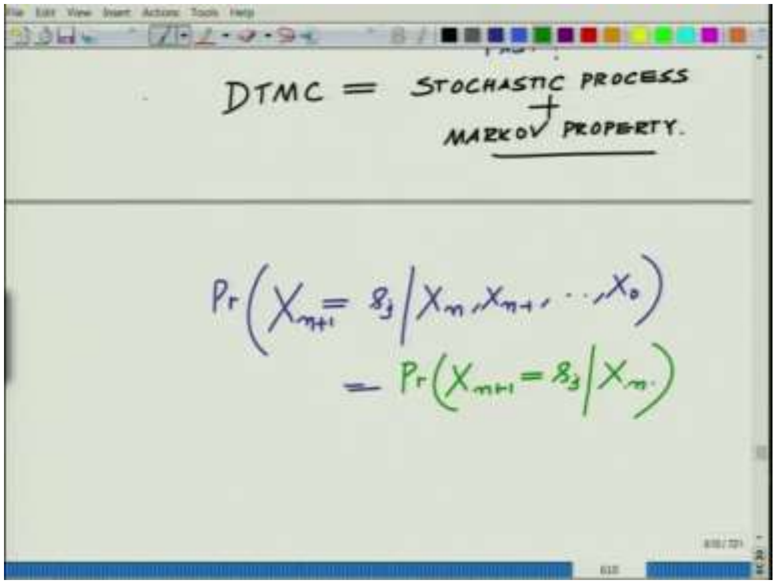
(Refer Slide Time: 32:09)

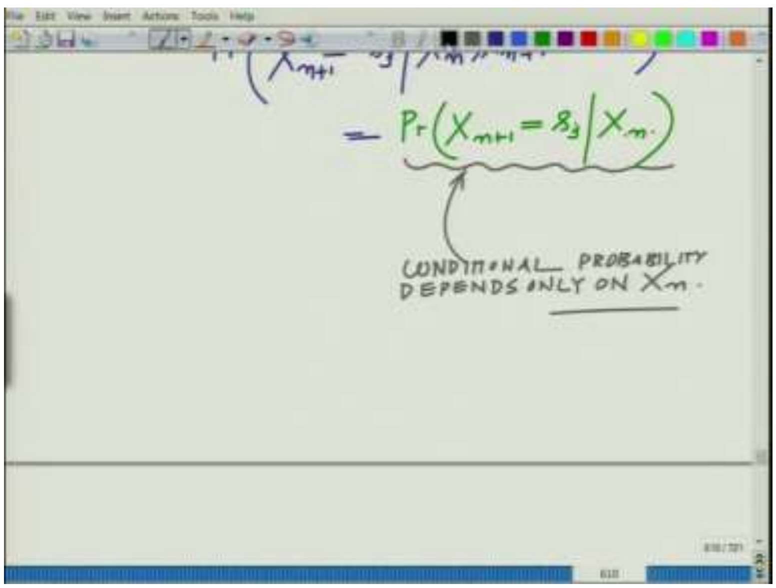




So, the future, so one way to say this is the future becomes decoupled from the past. So, the future becomes decoupled from the past, and this property is termed as the Markov property. This property is termed as, it is a very important property; this is termed as the Markov property and this is the name of a person, name of a scientist and such a discretion, such a stochastic process is termed, because we are looking at discrete time stochastic process, this becomes a discrete time Markov chain or what we also call as a DTMC. So, this essentially becomes a discrete time Markov chain. So, what is your DTMC or Discrete Time Markov chain?

(Refer Slide Time: 33:32)





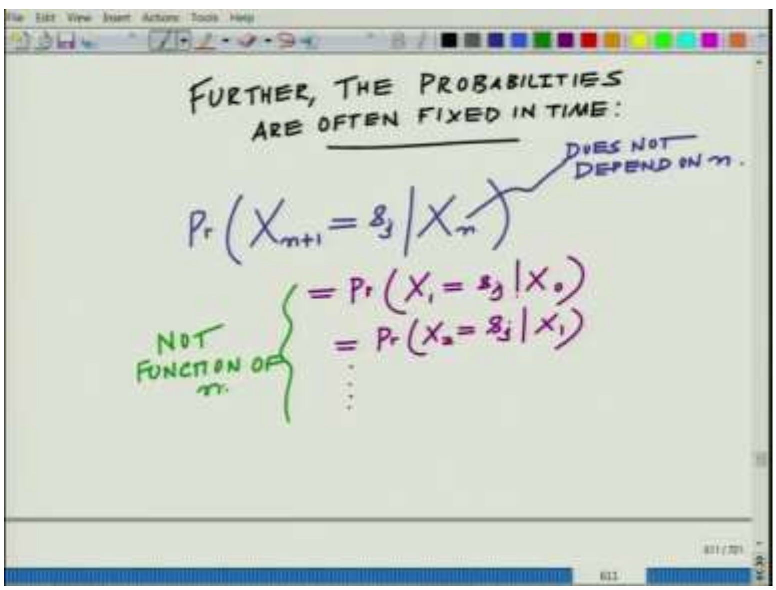
A DTMC is nothing but two things it is essentially a system evolving in time, that is your stochastic process plus the all important Markov property. So, a stochastic process with the Markov property and what is the Markov property? That is, remember if you want to, it is a very important and very easy property to also understand intuitively that is probability that X plus 1 equals to set sj given Xn, Xn minus 1, X0 this depends only on Xn that is equal to probability of Xn plus 1 equal to sj given Xn.

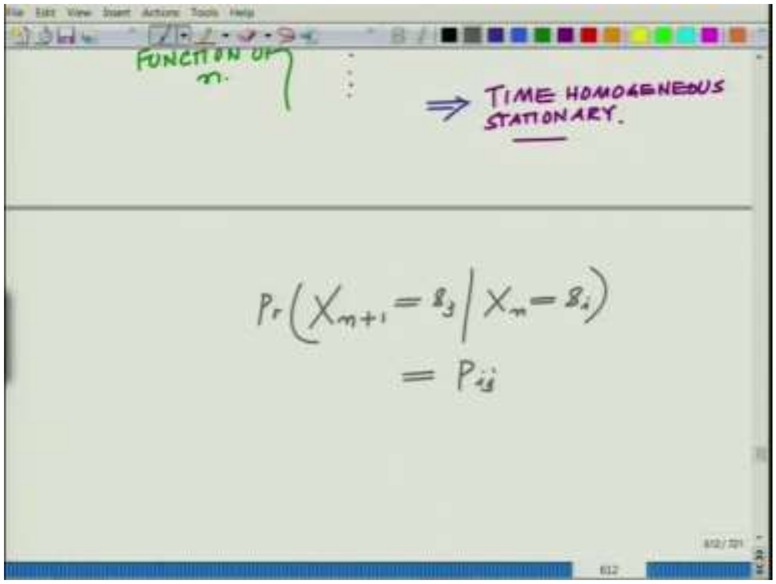
That is the temperature tomorrow, that is if you know, what it means is, for instance if you know temperature today that is sufficient to make a prediction about the temperature tomorrow. You do not need about the predict temperature yesterday, the day before yesterday and so on. Well, although it might seem practically, how do you put it, although it might not hold complete. Sometimes it might hold true, sometimes it might be an approximation that is made practical.

What it says is this, if this property, good assumption, because having known the temperature today you can make a reasonably good prediction or reasonably good inference regarding the temperature tomorrow. Although, the point here is sometimes it might be accurate and sometimes it might simply be an approximation for convenience but essentially this property, if it holds this is what is termed as a Markov property. And it is very useful approximation in many practical situations and also holds in many systems.

So, essentially the conditional probability, what it says is this conditional probability depends only on the, the conditional probability depends on how you look at it. Depending on if you look at Xn as a current state, conditional probability depends only on Xn, the conditional probability depends only on Xn.

(Refer Slide Time: 36:25)



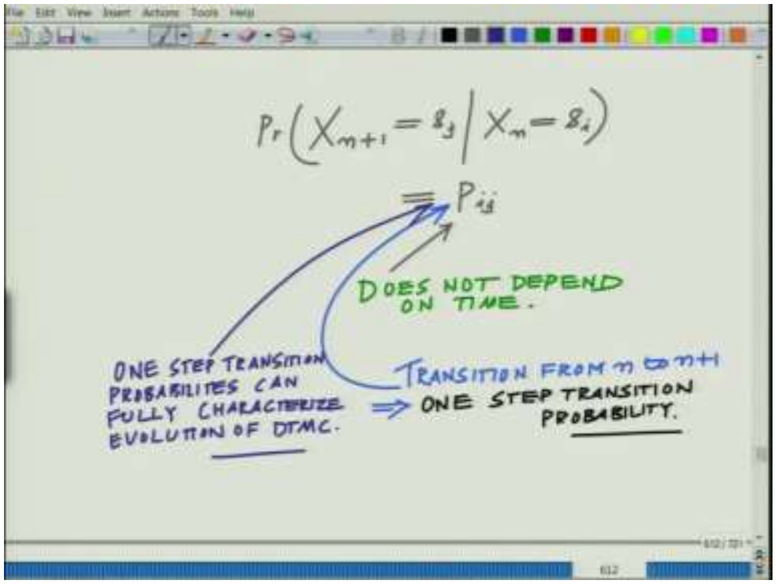


And further, these probabilities are often fixed in time, which again is a reasonable assumption to make, that is what we are saying is the probability Xn plus 1 equals sj given Xn does not depend on n. That is we will have, this is equal to probability if you set n equal to 1 or n equal to 0 this will be probability X1 equal to sj given X0 which is probability X2 equal to sj given X1 and so on and so forth. That is, these are not function of n that is what we are trying to say this is not a function of n. So, not a function of n.

So, this is known as the Times Homogeneous. So, this is known as essentially such as discrete time Markov chain, this is known as essentially being time homogeneous, homogeneous with respect to time and you can also say that these are stationary. These probabilities are stationary with respect to time and this is what we can denote as, we will come to that later, for instance if we have the probabilities I will come to that later.

In fact if you have these probabilities, probability Xn plus 1 equal to sj given Xn equal to s i for a time homogeneous system. These can be derived by the quantity p ij which you can see does not depend on time.

(Refer Slide Time: 39:16)



You can see it does not depend on time. These are also known as the One Step Transition Probabilities. So, these are known as, if you are looking at, that is looking at Xn plus the prediction. That is looking at the probability for Xn plus 1 equals lies Xn plus 1 is basically the state. The state at time n plus 1 is sj given the state at time n is si that is denoted by the quantity p ij. It does not depend on the specific value of specific time instant n. This is known as the one step transition probability because you are looking at the transition from n to n plus 1.

So, it does not depend on time and transition from this looks at the transition from n plus 1. So, these are known as basically, these are what are know as the one step transition probabilities. This is known as the one step transition probability. So, this Markov chain, the discrete time as we are going to see in the subsequent module. This DTMC can be effectively characterized by these one step transition probabilities because we said in a Markov chain only the current state is enough to make a prediction or enough to characterized the future revolution of the system.

And therefore, one can ask the question given the current state, what, given the current state is si what is the probability that the next state is sj that is characterized by this one step transition probability. So, these one step transition probabilities can characterize, fully characterize the evolution of a DTMC that is the essential. So, this one step transition probabilities can fully characterize the evolution of the discrete time Markov chain, probabilities transition probabilities can fully characterize the evolution of this discrete time Markov chain.

Now how can one do this? We will discuss that thoroughly in the subsequent modules that is starting with this one step transition probabilities how can you infer the various properties and how can you make various predictions and how can you characterize the various aspects of the evolution of this stochastic process which is also a Markov process so this is a discrete time Markov chain. So, we will stop here and continue with this discussion in the subsequent modules thank you very much.