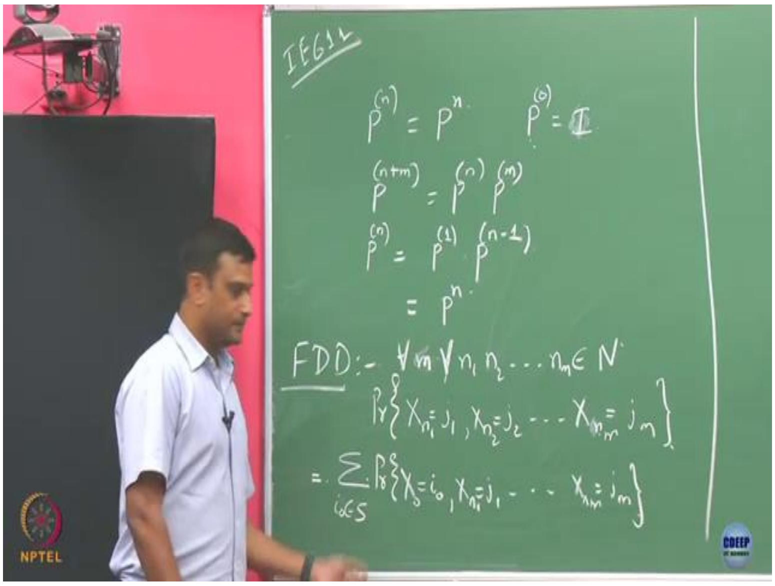
Introduction to Stochastic Processes
Professor Manjesh Hanawal
Industrial Engineering & Operations Research
Indian Institute of Technology, Bombay
Lecture 31 – Finite Dimensional Distribution of a Markov Chain

Now let us start. So, we were started discussing about Markov processes in the last class, so Markov process, we also called as Markov chains. And we are mostly interested in discrete time Markov chains. So, in the last class we defined what we mean by this Markov property and then discussed what we mean by time homogeneous Markov property and then try to connect Transition Probability Matrix with my n-step transition probability matrix.

(Refer Slide Time: 0:55)



So, in the last class we said that, he p of n equals to p to the power n. What was p superscript n means? This is the n step transition probability matrix and what is p here? One step transition probability matrix or we simply call it as transition probability matrix. And we have also concluded that P of 0 is 1, sorry, I identity matrix.

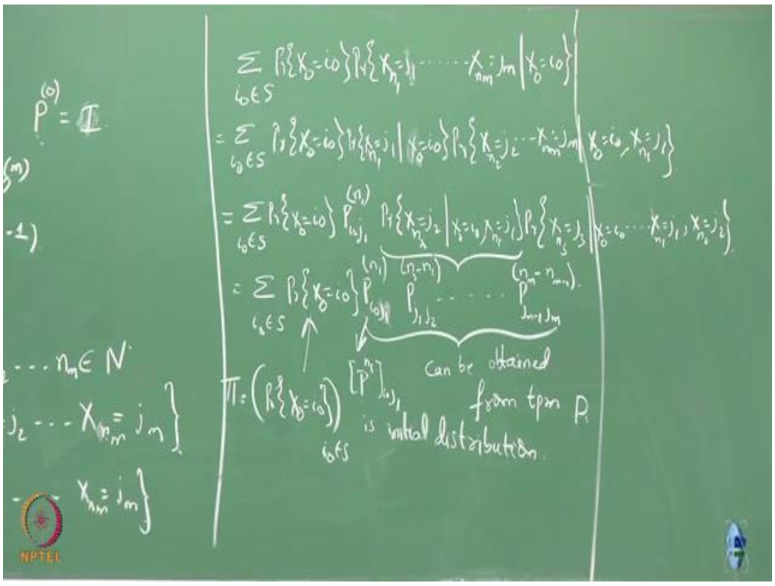
So, in the last class, we just shown that p n plus m can be written as P of n plus into this product. And then we said that we could write it as p of 1 and p of n minus 1 and then we said that so, and so p of n we are able to split it in this fashion and we could further split them into this fashion and we said that this can be finally result in p to the power n, fine. Now because of this so I think last time after completing this, I also wanted to introduce the notion of finite dimensional distributions, which I could not do last time. Let me finish this today before moving to another topic on this Marcov chains. We told that any process, a stochastic process to completely characterize this, I need to give his finite dimensional distributions, right?

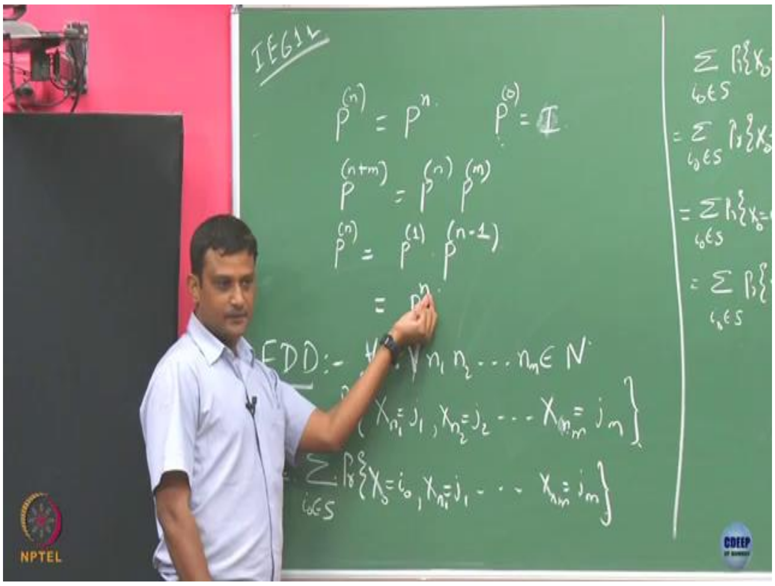
That is, I need to for any n if it is a discrete time process, I need to characterize joint distributions or joint probabilities of this format, assuming that I am looking into random variables that take that are discrete. That is, let us say let me write it as m. So, this is a joint distributions involving n random variables, sorry involving n random variables here and I need to give such joint distribution for all possible collections of my random variable to completely characterize my random process.

So, here I have taken some particular m random variables at time index n1, n2, and n m. You can choose any time index you want, okay? And this could be again for all time indices, you take any number m, and take any time index as m and let us say I have such a distribution, how we need to characterize this, I should be able to do that, then that will give me complete characterization of my random process. And this j 1, j 2 are the possible realization this random variable can give, so I need to also define this for all possible values my random variables can take.

Now let us see if this what all the things that govern this finite dimensional distribution in the case of a Markov chain, okay? I have this, what I could do is I am going to just introduce this initial state and then keep everything. So if I do this, nothing changes. These two quantities are the same like I have just introduced this x not random variable and just summed it over all possible values it can take. We have said that all possible values are coming from this state space S.

(Refer Slide Time: 6:12)





Now you can make this probability, what I call as j 1 given x not equals to i not and further we can keep on doing this, here it is x n 1. So I just first draw this out and then conditioned apply on that, then brought x n 1 out. And then in the second time, I just conditioned this random variable on that.

Now, you can keep on doing this. Now let us try to understand this part. So what is this? Now what I am asking, you were in the beginning in the 0th round. I am in state i not and then I am going to jump to state j 1 in n 1 rounds, this is you are basically jumping in n states from state i 0 to j 1, what is this probability in our notation? p i not, j 1, what is the superscript? So it is going to be basically n 1 minus 0 like the number of steps is n, I will just write it.

Now if you look into this, does this probability here depends on x not equals to i 0, no because if once I condition it on something beyond i 0 that is x n 1, this guy does not depend on this. So I could just write it as, I can do on, I can further split it like this.

Now, what is this quantity here? This quantity I know that it does not depends on x 0 equals to i not. So in that case, what is this probability then? So, is it okay? Is it correct in our notation if I write it as, so this is j 1 here not i 1, j 1 and this is going from j 1 to j 2 in how many steps? N 2 minus n 1. And I can keep on repeating the steps. And finally, what last step I am going to get is j m minus 1 times j m and what is this guy going to be?

Now let us focus on the last term. So what are this joint distribution of m random variables, I am able to now express in terms of this. Now, each of the terms here P i 0, i 1 to superscript n 1, this is a n 1 step transition probability matrix. And similarly, this is n minus 2 n 1 length step transition probability matrix. So all of this quantity here, they depend on transition probability matrix of different different steps.

Okay, but what we have demonstrated earlier is whatever is step we are looking at, all that can be obtained by one single transition probability matrix by appropriately exponenting it. So, all these quantities, a claim is they can be computed from my transition probability matrix P, is that correct?

Student: It will become p less 12.

Professor: Right. But for, the so this quantity how can I write? I can write it as p raised to the power n 1 but which term? This is a matrix, p raised to the power n m if p is the transition. In that which term I will be interested in? i 0 j 1.

This the matrix I have to tell both row and column indexes. So, all these things can be obtained from transition probability matrix P. So fine, all these things could be obtained from transition probability matrix P, but still to complete this finite dimensional distribution, I need one more term which is what is the probability that x not equals to i not and this is initially called the initial distribution. So, x not that is your initial step and you want to know what is this probability that my initial step takes state i not and you want to know it for all possible states.

So, this is called so i not where your i not belongs to S is your initial distribution, or initial probability distribution. So, and often this is denoted as pi. And so, if you know this initial transition probability matrix and your transition probability matrix P, do you think you will be able to characterize all finite dimensional distribution of a Marcov chain?

So, now this, contrast this with IID process, then how IID process, to completely define my IID process, what I needed to tell you? Just one distribution, right? That one distribution that is common across all the random variables, but now here to completely describe a Marcov chain, I need to tell you two things, to completely describe a time homogeneous Markov chain, I need to tell you initial distribution and its transition probability matrix.

So, as you see like the initial distribution is important, because where the Marcov chain goes up to some round depends on way at which point it started. If it starts at different point, maybe the future I am going to see may be different. So that is why to completely distribute describe a Marco chain, I tell you the initial distribution as well as my transition probability matrix. So, initial distribution tell you, okay, where you are starting, and transition probability matrix will describe how you are going to evolve from that point.

So Markov chains, the way we have describes it, they capture actually a lot of things that happen, that we face in reality and they give an often nice characterization of those phenomena.

For example, let us say if you want to a model some let us say aircraft, so we want to model or you want to analyse where it is going to go and how it is going to go, suppose, let us say its trajectory is mostly defined by what? Its velocity, its acceleration, let us say. And maybe it also depends on how much fuel it is carrying.

So once you give it some initial trajectory and you want to start looking at this position at different different times, let us say you are going to look at this position at every second or let us say every minute. Now, whatever it reach let us say at certain point of time, let us say at every time I am going to measure velocity, its acceleration that time and also the fuel, amount of fuel that is carrying that time.

If you know that, based on that itself, you could further analyse further how far it is going to go. Right? I do not need to know what all the values of the acceleration and the velocity or the fuel it carried in the previous time instances, if let us say if I measured it at let us say t equals to 100 unit, then that is enough like all the things that has happened before it is kind of captured there. Now fuel has been consumed, consumed till this point, but what remains is of importance to me to know the future.

So, in the sense, the current time is kind of already capturing the summary of what has happened in the past. So if you tell me past entire thing, I can tell you what is the current thing. But if I tell you what is the current status, yes, it is a kind of summary of what has happened. It may not tell you what all happened, how the things happened in the past, but it contains the summary of the things that has happened so far that is enough for you to discuss the future. For example, if you just know at time t what is the fuel remaining and its current acceleration and all, that is enough for you to see how my aircraft is, how far it is going it in the future.

So, Markov chains are exactly doing this. They are just saying that, if in your scenario is such that the current state is capturing enough information, a kind of summarizing what has happened in the past, then you can ignore the past. And your future will be just you can analyse your future based on this current information. Initial x not….

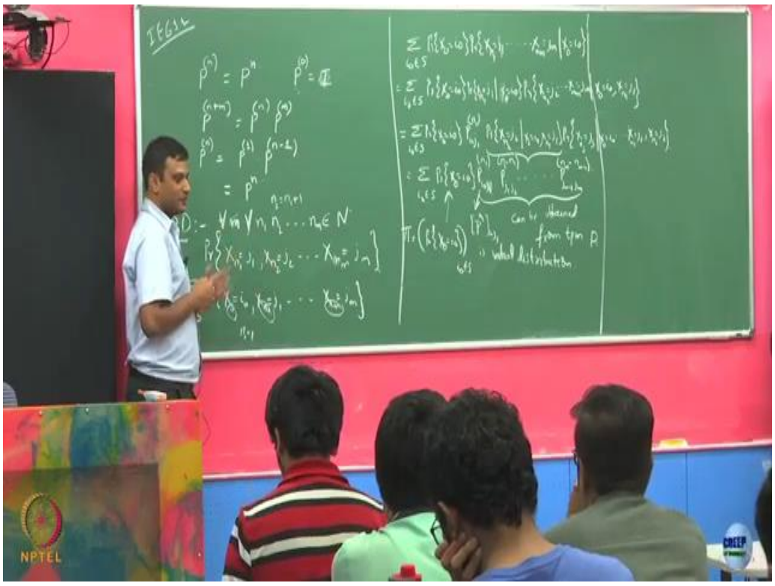
Student: Why it is depending on the x not? It is so because from the latest information available that is x n 1.

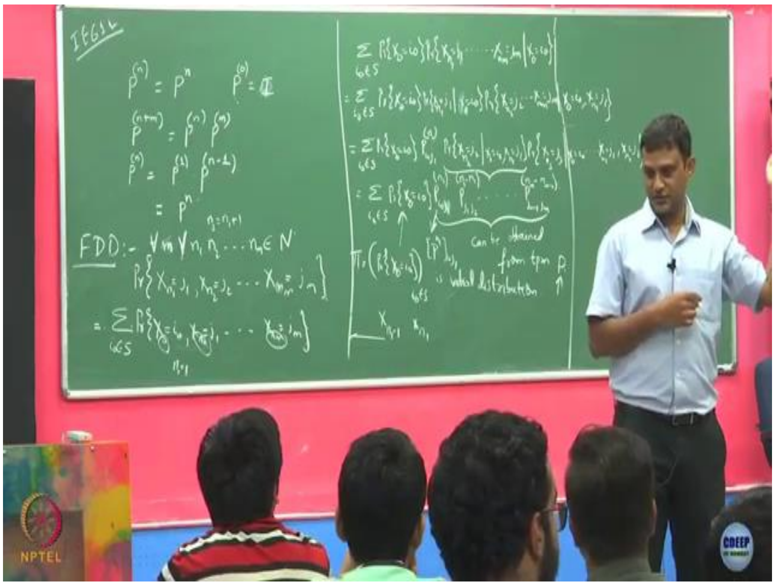
Professor: But I am, of course, it depends x n 1 depends on what happened in the x n n 1 minus 1 the previous plot. But here I am just trying to characterize the distribution jointly at different different time instance.

Student: So it should depend upon the, just depend upon the Cartesian, why the origin?

Professor: But the origin I am not interested, I am just trying interested in calculating the joint behaviour at this point. Indeed, what is happening going to happen here, we do depend on what is going, from where you are going to start. So that is why this behaviour is going to affect what is we are going to see at n 1. We are just trying to capture that summary not like what has happening in-between. Even when you are going to look at finite dimensional distributions, even you need to characterize that aspect also.

(Refer Slide Time: 20:28)





So, for example, here n 1 could be just 1. I am considering all possibilities of the indexes. I am not ignoring that fact. So for example here, n 2 could be just n 1 plus 1, not only one step, how it is going to affect all and I am also going to look at jointly all the distributions. What you are asking in this?

You are asking to find all these joint distributions, right? You are not asking it just like how it depends on the previous one. What we are trying to describe here is when we have a Markov chain property, when you have Markov property, all these joint distributions are going to be just captured by my transition probability matrix and my initial distribution, this transition probability matrix is already trying to capture all you are asking it for from this state, what is the probability that I go to next state?

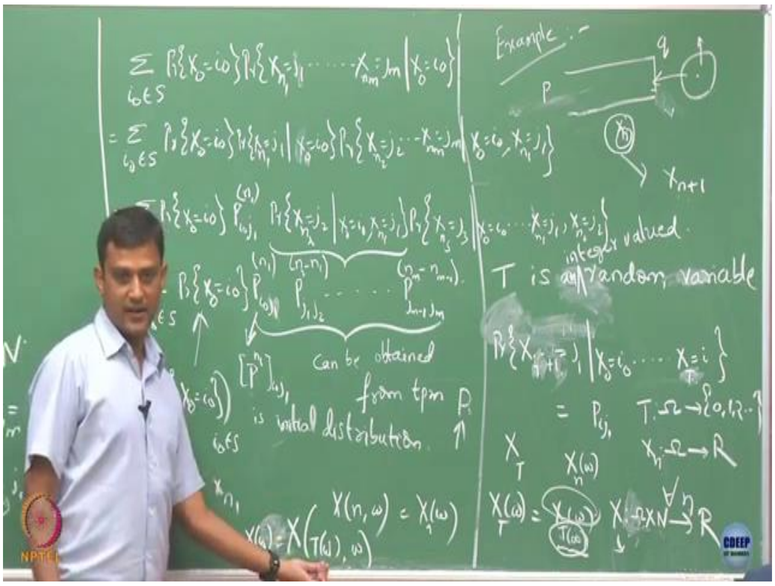
But that is not the only thing I will be interested in, maybe like to completely understand the process I need to know all this joint process, joint distributions. For this we are saying, if you want to just what you are asking is okay, n 1 will only depend on what has happened in the previous step. That is fine. Because I am saying that it do depend on where you are going to start from. If you are going to just okay, you start, you only going to give me from n 1 minus 1 and you are asking me, okay, what is going to happen? But that is not going to completely describe my process, I may be interested even to know what was before that.

That is what the meaning of FDD is, all possible description I need to give about the process. We are just trying to model this, okay, somebody has given you and then I want to analyse the system, how to obtain this transition probability matrix is a different question. For that, one has to I mean really do big big experimental setup, do real measurements. So, in this course, like right through we are putting all these things under carpet.

Okay, how we are going to compute this initial probabilities, how we are going to obtain this transition probability matrix. This is all there is a separate thing how we are going to find this like you have to do simulation and then try to estimate this parameter. So these are all parameters for us. P is a transition probability means it is a parameter, it is a matrix with so many parameters, I do not know those parameter, what I am just saying that so if you pass on these parameters to me, I am going to analyse and tell this is how your system is going to behave.

Okay and so for example, that is how like if you want to evaluate 2-3 aircrafts, each aircraft will come its own parameters, then as a system analyst, you will say okay, take all this inputs and then you try to analyse the system and who is going to give the parameter for this system. So like the vendor will give you the system, he would have said okay, these are all my system. So, it is fuel efficiency is these, it can sustain this much of turbulence and it can go at this speed and this it can take this much acceleration at all. Using those things you will come up with this set of parameters and then use it to understand your future.

(Refer Slide Time: 24:11)



Okay, let us understand another example. All of us know queues like we wait lot of time in the queues. So let us say we have a queue and there is a server. Let us take a simple idealized case where time is slotted like at every time a guy may come and join the queue. So, in every slot, there is with some probability a new guy will come and join the queue. It may happen that in that particular slot, nobody will join, but with some positive probability he may also join.

Now and also meanwhile, there is a server here which keeps serving you, let us say you are just in a, going to watch a movie and you are in the line to get the ticket, so somebody is preparing a ticket for you and as you get the ticket, you will move out and lines may either grow longer or grow shorter depending on people are joining or people are exiting the system.

Now, let us say at some point, let us say at time x n, so n is going to tell you how many people are there in the queue at time slot n, so how we are going to define time slot is up to you. You may want to count the number of people in the system at let us say every second or maybe at every minute. It depends on the say, I mean it depends on the scenario like if people are coming very very frequently, you may want to take up smaller time scale. If your people are coming slowly, you may want to take a larger time scale or time being let us say this is just n equals to second, and second 1, second 2 like that. Every second a guy will join the system at some probability and you want to see how many people are there in the system.

Now so time being assumed that you know this service rate the way this guy is preparing the tickets, then by that you can kind of understand at what rate people are leaving the system, getting out of the system. So, the guy who get the ticket like he will just rush inside the theatre and it may go down. But that it could be that, that process the rate at or the rate at which this guy serves the people, it could itself be something stochastic.

For example, some guys will be very annoying, may ask the receptionist the ticket guy a lot of questions and keep him engaged. So that guy will take more time to get served or some guys just want to grab the ticket and go, so that guy may take less time. So there itself some stochasticity can happen and also we are saying that then the people are joining, there itself is a stochasticity every time a guy is going to. Now if you know all, suppose if you know the description of the stochasticity, at any time if you know how many people are there in the system, can you at least give probabilistically how many, what will be the next state of the queue here? Time being let us say, people are entering with probability q in every slot and every slot a guy gets out and leaves the system with probability q.

If your system currently at n, there are n guys waiting in the queue in slot n, can you say how many people, what are the possible values for x n plus 1? So, what are the possibilities? The next so, the new value of this number of people waiting in the queue will increase by 1 if a new guy joins….

Student: And nobody leaves.

Professor: Nobody leaves, it is going to decrease by 1 if nobody joins but one guy leaves, it is going to remain same if either nobody joins or nobody leaves or one guy joins and one guy leaves. So to describe this new number of people that are waiting in the queue, does it matter like starting from the beginning, how many people are there till around n or only the previous one? Previous slot is enough there like if I know what is my status currently, I can describe what is going to happen in the future.

In a way the number of people waiting in the queue is already kind of summary of the system, what has happened so far, it is capturing the summary of the system and that summary is itself enough for me to define the future. I do not really need to know what has, how happened till x 1, x 2 all the way up to x n minus 1, all that is enough is x n.

So, you see that Markov chain is exactly capturing such kind of system, where my current is already capturing or kind of summarizing all my past that is enough to describe my future. And in many cases, this will arise. For example, this q model and the aircraft example, I gave like just that is a toy example, but many realistic examples can be thought of that follow such kind of scenario.

Student: How many people are in the queue with Markov Chain?

Professor: Yeah, in that case, if I go to think of this x n as a process, x n process, then this is going to follow.

Student: And what will be the service time?

Professor: So right now rather than, rather than saying whether it to satisfy the Markov property, what I want you to understand here is the motivation. In this example, is it enough for me to know the current state of the system to know the future? Or is it like I need to know the entire thing that has happened so far to know the future? That is enough. In a sense the current state is kind of capturing the summary that is enough to describe my future. So that is where we want to start using these Markov chains.

Student: How can go about it without knowing ‘p’ and ‘q’? They can be variables, they might change.

Professor: Yeah. It can change but see, this is what I like. I am not saying you without knowing p and q, you give me p and q and then I am going to as an analyst, I am not so, there are parameters. So, there are two things. There are parameters, given these parameters as an analyst you are sitting and analysing and telling what happens, who is going to give these parameters is a different question.

So, for example, let us say if you are a planner, you want to plan something, somebody will say I have ten manpower to do this. Somebody says, I have this much budget to do this. And somebody will say I have this much time to do this. So, this has been given to you and now you will be coming up with a plan or analysis. Exactly like that. So here this description is given to you and now you are analysing, what is that? If somebody is going to say, okay, the time I have is less and somebody says money I have is less and somebody say I have less manpower, then you are going to reanalyse with this new parameter and say okay, this is in the performance.

Right now I am just saying like a big boss, so you tell me what is all happening, what all the things you have and then I am going to plan and or analyse what is going to happen. Okay. So fine, okay now, the way we have defined Markov property, it says that if I tell you at this index, I know the status of the system, then from that point onwards, I will, to explain my system I do not need to know anything about my past.

But most of the time it so happens that this time index that we are going to deal with may not be deterministic, itself could be a random thing. Okay for example, I want to see how much my aircraft goes further when its fuel tanks become half. So, now, the point where the fuel tanks become half itself could be a random thing because so, the fuel consumption that is going to happen it depends on so much of environment that you are going to face that you may not have control that itself is a stochastic thing.

So, your fuel tank may get half in just two hours or one hour, 50 minutes or maybe one hour 55 minutes, whatever. It could be a random time and from that point onwards you want to understand, you want to understand how the things behave in the future. So there this n here could itself be a random quantity. So for example, you want to understand so let us say you just randomly dropped into some cinema hall and at random you dropped in, it is not necessary that when the counter opened.

You may have, after that you may enter the queue after 10 minutes or one hour or whatever. So the time when you are joining the queue itself could be a random time. Now you want to see, okay, once I joined this, how the things and what are the things I am going to say like what is going to happen. For example in this case, let us say you may want to analyse if I joined the system at some point, how much time I need to get before I get my ticket.

So, to analyse how much time I need to get the ticket, you want to understand what if I, my entry time is not exactly deterministic, but that itself is a random. So, all of us like go and join queue at some in rush and at some point we join and from that point we want to analyse. Yeah?

Student: Is that something an arbitrary time?

Professor: This is an arbitrary time but this still a deterministic This is all index, this n I said, this is some n. I know I could make it random by saying that, okay, this is x t, t is a random variable. Now, t is going to decide which index I will be looking at.

Okay, so repeat, this example you can imagine. You are joining this queue and the amount of time that you need to get in this queue before you get a ticket, that is a random thing. Because it depends on how many people were before you and also depends on the people before you how fast they are getting served, how much time they are getting served. Now if you know that you have exactly entered at a particular time, you know this, but you may not know, you may be entering the system at random point.

From that random point now you want to analyse, you want to understand how much time I may need to get served. Okay, but so I am saying a priory if you know exactly at what time you have entered, it is fine, but if you do not know this time of joining that itself could be random, how that that is going to affect your future? So, now the question is the Markov property we have discussed so far, does it carry over to the case when I have such a random times?

So, I want to basically, we want to basically analyse. Suppose, let us say t is a random number, t is a random variable, which is integer valued. Now I want to know probabilistic one. So, now I want to know, so now see like I have now put this, I want to ask you if I tell you the observation of the states of the system till this random time t, now the next state it is going to take in the state j 1 it is going to take in the next round, can I write it as simply is one step, p i j 1?

So if this t is some fixed n, this is not our description, p i j, so this guy is independent of everything else except this x n equals to i n that is the definition of p i j 1. But now I want to ask the question, what happens if this time is not deterministic, is random? So notice that when I make it random, this t could be taking any possible values that this random variable can take. Okay, so now let us make this notation bit more clear. What I mean by x t?

How to interpret this term x t? Okay. So, I know that what is x n of omega, the value taken by my n x nth random variable at sample point omega. Now what is x of t means?

Student: The value taken by random variable at index t.

Professor: So this t here, here and we are going to denote it as x of t of omega. So t itself is a random variable. So, on that sample point, just first tell me what is the index. And from that index, what is the value taken by this x t of omega on that sample omega? Yeah, they will be same. So, when I write x of t, the way you have to read it is, so this is a random variable x t. So this is again, so t is from omega to 0, 1, to all the way. And I know that x of n is also from omega to r, this is for all n, all my x n's are defined on the same sample space.

Now, when I say x of t, it is just the term not specifying you which index n, I am just specifying your random index. So, when I write, so this is basically x t of omega. When I write, this means, first look at the index at the sample point omega and look at its value at this omega.

Okay, another example. So just let us say if this fits well here. So let us say that I am interested in 100 stocks, okay? Or whatever number you take in the stock market. And what I am interested in is how the values of the stocks change on every day. Okay, so let us call the stocks 1 to, up to 100. I can think of this as my sample space, omega is this stocks.

Now on each day, I can focus on this stock and then see how its value is going to change. If you are going to fix a stock, let us say that corresponds to omega then you know the value it is going to take on different different days that is one sample path you have. Now, what I want is I would say that okay, the value of my particular stock when its value becomes less than certain number on the day, when its value goes below certain number, let us say. So, now what is t is going to be? t is going to describe you the day, how you have said that particular stock when its value become smaller than some number, so that its value may become smaller than some number in some day, you do not know that could be a random quantity.

You go to that day and then look at the value taken by that stock on that day t of omega, so, the value that stock has. So, in this case this t of omega basically gave me the day on which the value of that stock omega went below certain number and then this is going to tell exactly what is the value of that stock. So, you realize that why this random I may be interested in knowing my value process value at a random time because you may want to put some conditions like that. For example, you may want to, in the stock case you may say that okay, what is the value of my stock when it is exceeded 500 rupees?

Okay, so you will look for the time when it exceeded 500. What I am saying is it is exceeded 500 but its value itself need not be 500, it may be 600 but what I was interested in when it exceeded 500, it may have exceeded 500 rupees on the hundredth day, but when it exceeded what is the price? That will be given by this and when it exceeded 500, that day is given by this t of omega. So, and here omega corresponds to that particular stock that I would be interested in. x will go from…

Student: Will not that x go from either function from 0, 1 2, the said 0, 1, 2, to r?

Professor: No, x it is not like x exactly so, x is actually basically I can think of x as index by and with index and sample. So, this same thing we have written as x n omega. So, you understand the meaning of this random variables here. For every n I am defining x n taking value on omega and assigning values to omega on real line. So, instead of fixing n I can just make it a joint one. So, I can think it as so, I think….

Student: n belongs to capital N.

Professor: Yeah, n belongs to capital N. N is the set of natural numbers. So either so, you first so to define a random variable, the process what I asked, give me an index and then give me a sample. Then I will tell you what is the corresponding value. So here I am instead of separately defining for each index, then I can just say give me the sample point and give me the index, I will tell you what is the value that it takes.

So, in this case, this x is a process, we have discussed this when we were talking about a stochastic process, stochastic process can be thought of as a collection of indices, sorry collection of random variables, where every index corresponds to a random variable or alternatively we can also think it as random process, x as a map which gives value R for a pair, omega and n. So same thing in that case, we can take this as if this n, so here this can be thought of as, for so x of omega in this case can be thought of as x of t omega and omega.

So, sorry, yeah, x of t as this will go into tell you what is the index we will be looking at, and what is the…So, for example, when I say stock right, so your the t is defining so, on a particular stock omega, let us say you have defined when it exceeds value 500. The t of omega will tell you that time when it exceeds 500, that index.

Student: Sir, if there are two indices that again exceed 500?

Professor: So that is fine. I mean, that ambiguity is there. For time being for example, let us assume or maybe we can refine it as the first time when it takes value 500, we can do all such refinements. Then it is going to be unique in that case, and then we can define like this.