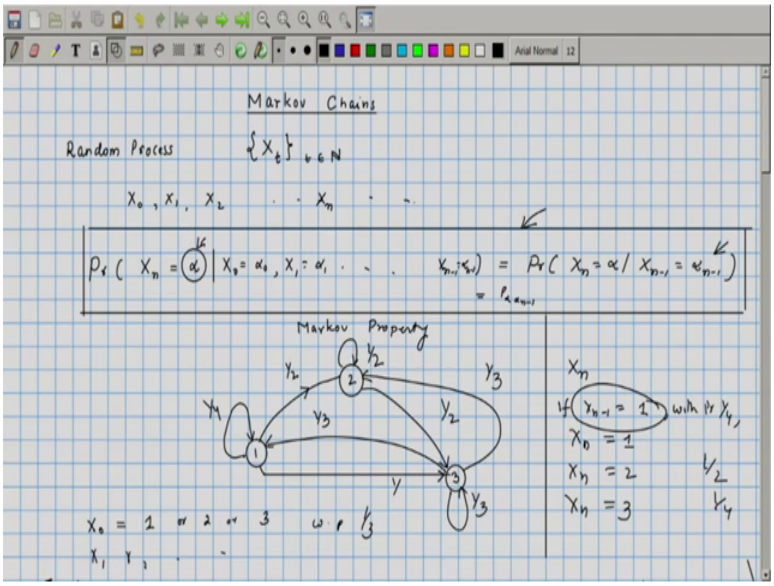
Randomized Algorithms

Prof. Benny George Kenkireth
Department of Computer Science & Engineering
Indian Institute of Technology, Guwahati

Lecture – 16
Introduction to Markov Chains

In today's lecture, we will learn about Markov Chains. So, what are Markov Chains?

(Refer Slide Time: 00:36)



So, it is a special kind of random process. By a random process what we mean, as a sequence of random variables indexed by some parameter ok. So, here for Markov Chains, we will assume that these index belongs to the set of natural numbers ok. So, we have random variables of this kind X 0 X 1 X 2 and so on, defined on some appropriate sample space. Now, these sequence of random variables, we will say that it forms a Markov Chain if certain conditions are satisfied.

So, let us say we are looking at this random variable X n ok. There can be very complex dependencies between these random variables. We want the dependency between these random variables to be of a certain kind, if that kind of dependency is there, then we will call this kind of a random process as a Markov Chain. So, let us describe that property. So, we will assume that these random variables takes discrete values or it takes values from a discrete sample space. So, we can conveniently assume it to be again from the set of natural numbers.

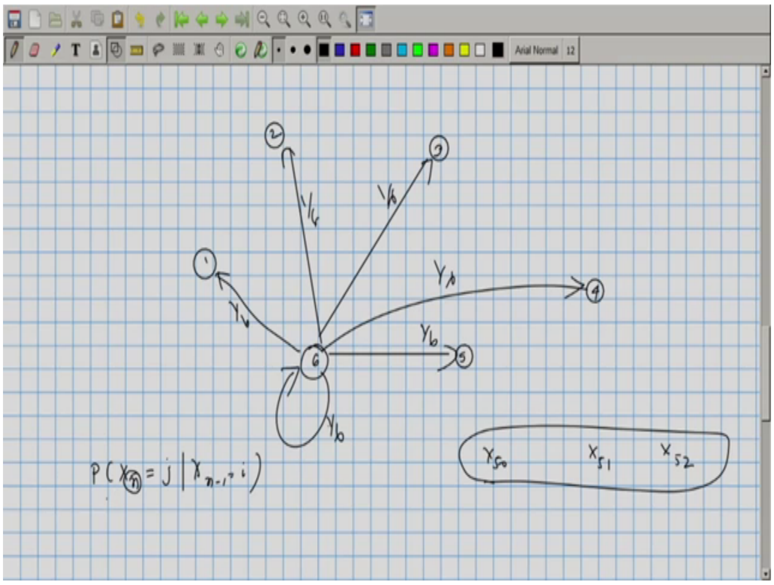
So, let us look at this probability that, this conditional probability that X n is equal to, let us say alpha given. So, when you are looking at the value that the nth random variable takes conditioned on the values taken by X 0 to X n minus 1, we are interested in the probability. So, probability that X n equals alpha given X 0 is equal to let us say alpha 0 X 1 is equal to alpha 1 and X n minus 1 is equal to alpha n minus 1. This conditional probability must be equal to the following conditional probability that is X n is equal to alpha given X n minus 1 is equal to alpha n minus 1. So whenever, so this is called as the Markov property. So, this is the Markov property.

So, it states that the conditional probability that the nth random variable takes the value alpha, conditioned on the values taken by the previous random variables is exactly equal to the conditional probability, where the conditioning is with respect to the value taken by the previous random variable X n minus 1 ok. So, let us see some examples of Markov Chains. So, let us say that we have three a random variable, which takes three different values. So, it can take either value 1 2 or 3 and we are going to think of this as states. So, X 0 can take any of these three values with equal probability. So, let us say X 0 is equal to 1 or 2 or 3 with probability 1 by 3 and then the successive random variables X 1 X 2 etcetera are define in the following way. So, that is defined by means of this state diagram. So, let us define those random variables.

So, let us say if you look at X n, X ns value is dependent on the value taken by X n minus 1 in the following way. So, if X n minus 1 is 1 then with one fourth probability so, if X n minus 1 is 1 then with one fourth probability X n is equal to 1 with one half probability X n is equal to 2 and with one fourth X n is equal to 3. So, this is conditioned on X n equals 1 sorry X n minus 1 equals 1, if X n minus 1 was 2 then with probability half X n is going to be equal to half with probability half it is going to be equal to 3, if X n minus 1 was 3, then with one third probability X n can take the values 1 2 and 3 ok, with one third for each of these.

So, this sequence of random variables are well defined and their probability is that it they take particular values are also well defined, you can verify that this indeed satisfies this particular property, it satisfies the Markov property. So, in general, when we look at Markov say Markov processes or Markov Chains the index here, we assumed it to be natural numbers in general. It can be any set a countable set or an uncountable set and this values that the random variable takes can also be from an uncountable set, but for us we will look at special kind of Markov Chains, where the index we will assume belongs to the set of natural numbers. So, this is a discrete-time of Markov Chain and we will assume that the values, that the random variables can take comes from a finite set and that is finite set, we will assume to be a subset of natural numbers.

(Refer Slide Time: 07:27)

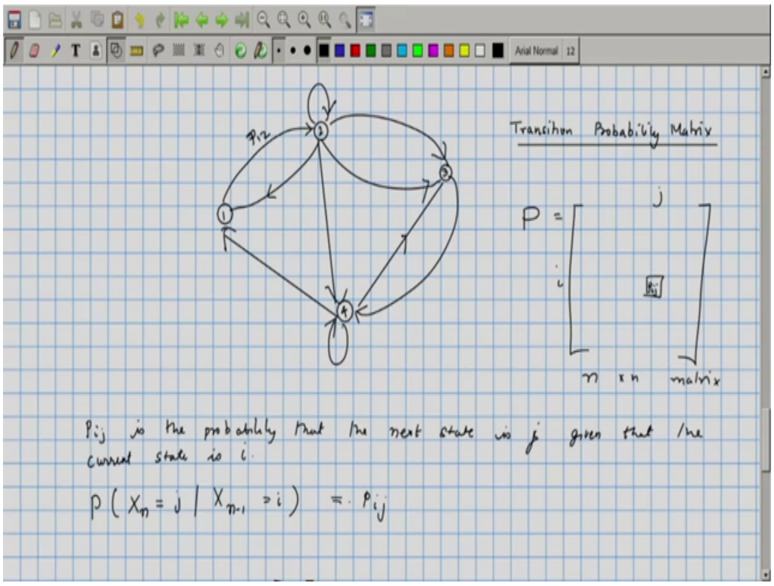


So, these kind of Markov processes, we can essentially depict them by a state diagram, where the values of the state indicates the values taken by these, the values that these random variables can take and then for each value; so, we can write these this probabilities X n is equal to j given X n minus 1 is equal to i ok.

So, at the n state, if you are in state i then you can draw these diagrams with edges to each of these other word, states with the weights on the edges indicating the probability with which those values are taken. For example, if X n minus 1 most 6, if I put values 1 by 6 on all of them, it means X n is going to be 1 of these values with probability 1 by 6. Now, you have to describe, this for every node in this diagram. Now, here we may have to do this for each and every value of n, if X 50 was something then for X, X 50 that will be one such diagram X 51 will have another diagram, X 52 could have yet another diagram and so on, but we will again restrict our attention, where all these diagrams are going to be exactly same. These are what are called as time homogeneous Markov Chains.

So, in addition to the Markov property, we will insist that this probability, this conditional probability depends only on alpha and alpha n minus 1. So, it can be written as P alpha n minus 1, whatever way the values alpha and alpha n minus 1 that alone decides this probability. Those kind of Markov Chains will be called as time homogeneous Markov Chains. So, when we are looking at time homogeneous Markov Chains, where the index is the set of natural numbers and the values that the random variable can take is a finite set, they can be conveniently represented by these things known as state diagrams.

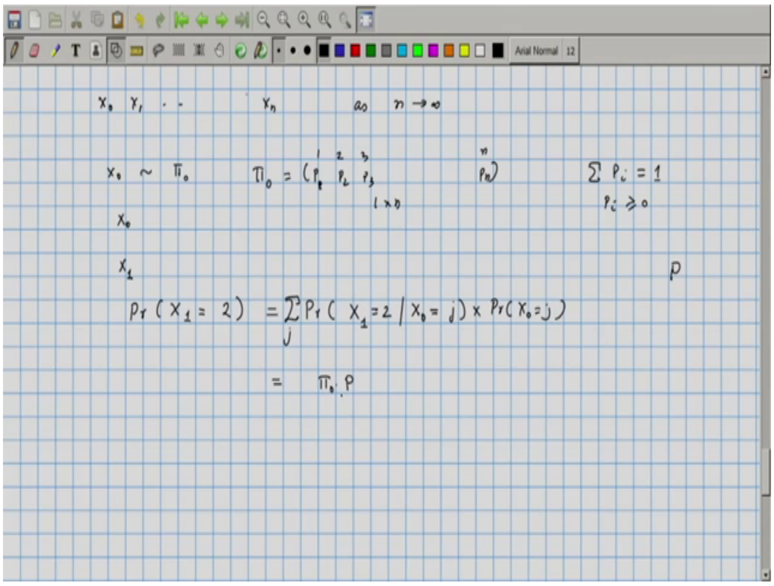
(Refer Slide Time: 09:54)



So, essentially it will be something like this, let us take an example. So, this is a Markov Chain with four possible states, which we will number as 1 2 3 and 4; that means, the random variables can take values 1 2 3 and 4 and let us say ok; so, we could have a state diagram like this where the probabilities are given. So, this is P 1 2 and so on for every particular edge, we can represent this conveniently by a matrix. So, we can think of the transition probability matrix ok. So, the transition probability matrix P consists of, this is an n cross n matrix, where n is the number of states or the number of values that the random variables X 1 X 2 X n etcetera can or X k can take and the i jth entry denoted by P i j. So, P i j is the probability that the next state is i given sorry. The next state is j given that the current state is i. So, it is a probability of transitioning from state i to state j ok. We can write it as probability that X n is equal to j given X n minus 1 is equal to i.

So, this is what we denote by P i j and a matrix with all these P i j for the different values of i and j is called as the transition probability matrix.

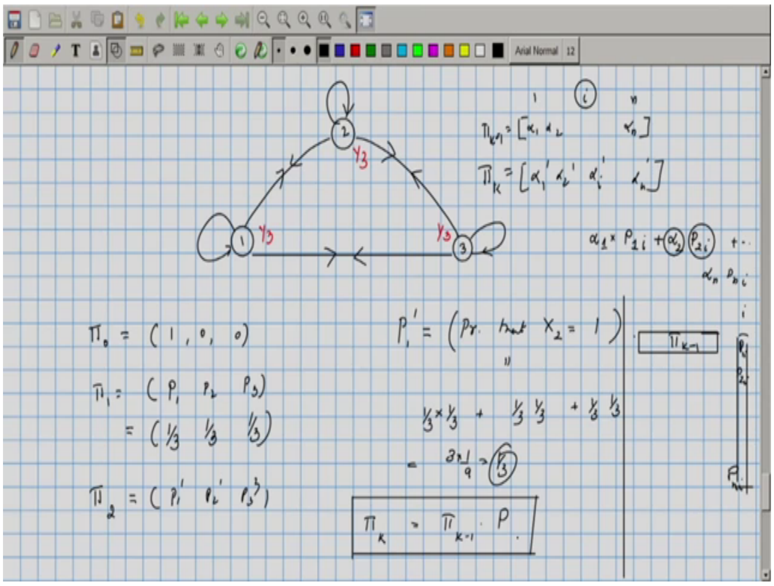
(Refer Slide Time: 12:37)



So, now let us look how, how these random variables behave? If you look at these random variables X 0 X 1 so on X n and as n tends to infinity. How does these random, what is the distribution of these random variables? So, let us assume that X 0 that is the first random variable whatever is the value it takes, its distribution is given ok. So, let us say X 0 is distributed as Pi 0; that means, X 0 takes the value. So, if you think of this as a vector of size is a 1 cross and vector say let us say our states were 1 2 3 up to n, so Pi 0. We can think of it as P 0 sorry P 1 P 2 P 3 up to P n. These are numbers says that summation P i equals 1 and P i is, are all greater than equal to 0 ok. So X, so Pi 0 is the distribution of X 0; that means, the probability that the random variable X 0 takes the value 1 as P 1, the probability that the random variable X 0 takes the value i is P i and so on.

Now, once X 0 is given, we can think about what will be X 1, because X 1 is a random variable, which is dependent on X 0 in a Markov sense. So, what is the distribution for X 1? So, we need to compute probability that X 1 is equal to let us say 2 given, the value. So, we need to determine the distribution of X 2 X 1 equals 2 and so on. So, we have to determine the distribution of X 1, what? How do we do that? We know the distribution of X 0. So, probability that X 2 is equal to 2 is nothing, but probability that X 1 is equal to 2 given, X 0 is equal to let us say j into probability that X 0 equals j summed over all values over of j. So, if X 0 was j conditioned on that. What is the probability? That X 1 is equal to 2, this summed up over all values gives this probability ok. So, if we had the transition probability matrix P i j, P whose i jth entry is a probability of going from state i to state j in any single step, then this probability will be nothing, but Pi 0 into P why is this so; let us take a simple one, simple example

(Refer Slide Time: 15:38)



Let us think of a Markov Chain with 3 states and let us assume that all the edges are present and they are taken with probability 1 by 3 each ok. So, I am not drawing the reverse edges, just draw it on the same. So, all these transitions happens with 1 by 3 probability each and let us say the initial distribution Pi 0 is equal to 1 0 0; that means, I am starting at state 1. Now, Pi 1 will be, let me call this as P 1 P 2 and P 3. Since, I have started at state 1, there is only a one third probability that I will be in state 1 after the first transition. So, this is going to be 1 by 3 and there is a one third probability that I will be in state 2 and there is a one by third probability that I will be in state 3.

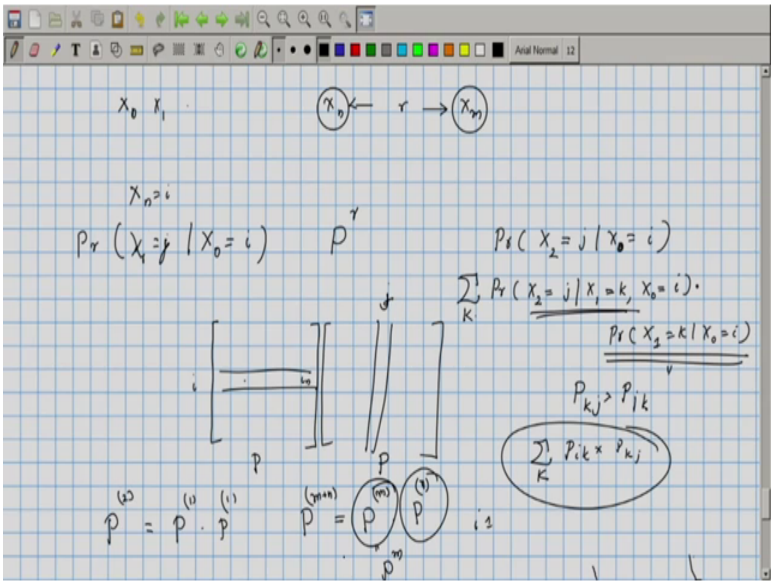
Now, what will Pi 2 be, like could be in this state with 1 by 3 probability, I could be here with one by third, 1 by 3 probability, I could be here with 1 by 3 probability given such a state if I make one more transition, one might take one more step, if I look at the random variable X 2 then what is the distribution of X 2. The probability that I will be in state 1 is equal to; so, let us call that as P 1 prime P 2 prime and P 3 prime. P 1 prime is equal to probability that X 2 is equal to 1, this is the probability that X 2 is equal to 1.

So, how is X 2 going to be equal to 1 well you could be at state 2, in the previous step that could happen with one third probability and then from there you could jump to state 1, that will again happen with one third probability or it could be in state 3 with one third probability and from there you could jump to state 1 with one third probability or you could have started in state 1 and remain there which again happens with one third probability. So, this is going to be equal to 3 into 1 by 9 that is 1 by 3 ok. So, this is going to be the same for every, I mean every transition, because this is a very symmetric case. So, Pi k in general is going to be equal to Pi k minus 1 into P. So, how do we see this?

Suppose, this was your distribution Pi k minus 1 at some stage ok. So, let us, let me call it as alpha 1 alpha 2 alpha n. The probability that I am in state, I at the next step. So, that will be given in the vector Pi k. Let me call this is alpha 1 prime alpha 2 prime alpha i prime alpha n prime ok. The probability that I will be in state i at the next instance is, probability that I am in state 1, which happens with alpha 1 into probability of going to state i from state sorry, going from state 1 to state i plus alpha 2 into P 2 i that is you are at state 2 and from state 2, you transition into state i and this summed over all values alpha n into P n i ok.

So, this is, so this expression is nothing, but take the vector corresponding to Pi k minus 1 multiply it with the column vector corresponding to so, the ith column P 1 i P 2 i and P n i ok. So, this is that product. So, that is true for any particular i. So, we can write this expression Pi k is equal to Pi k minus 1 times P. Now, we can so, we talked about going from 1 state to another in 1 step. We may also look at these random variables.

(Refer Slide Time: 20:21)

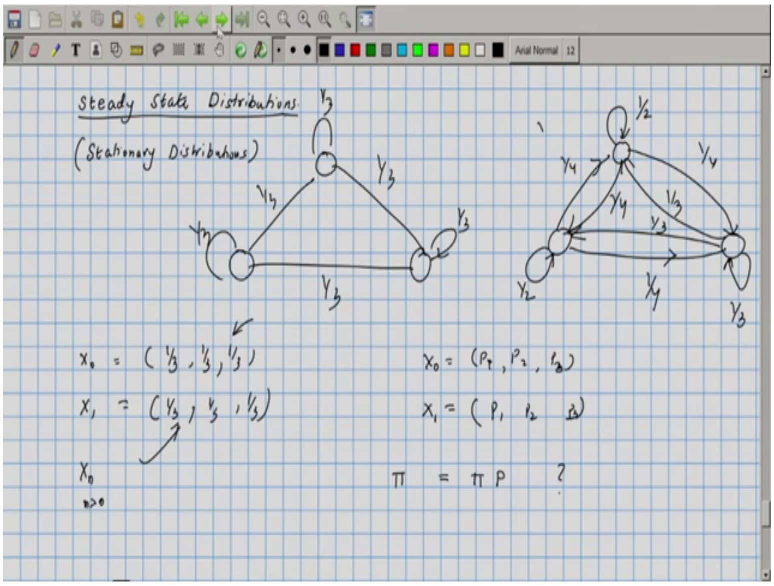


So, let us say X 0 X 1 X n and X m ok. So, if you are given the distribution for X n, how do you determine the distribution for X m ok. So, let us say this differences are, so, we want to compute the r step transition probabilities. So, if X n is equal to i or since it is time homogeneous, we could think of X 0 is equal to i and we want to know. So, probability; so, we are interested in probability that X r is equal to j given X 0 is equal to i ok. I can see that these probabilities can be obtained from the matrix P power r ok. So, we can just think of it for the 2, when r is 2. So, let us say this is P and this is another copy of P, so P times P, so i, so, we want probability that X 2 is equal to j given X 0 is equal to i ok. So, how old can X to be equal to j well X 0 is given to be i.

So, this probability we can write it as probability that X 2 is equal to j given X 1 equals k and X 0 is equal to i multiplied by probability that X 1 equals k given X 0 equals i summed up over all values of k that is the first step was to, was from i to k and the second step was from k to j and this summed up over all values gives the probability that X 2 equals j given X 0 equals i. So, this X 1 equals k given X 0 equals i is P i k and this quantity by Markov property this is going to be X 2 equals j given X 1 equals k. So, that is going to be equal to P k j. So, we can write this as P i k into P k j summed over all values of k that is just if you take the ith row and multiply it with the kth column sorry, ith row with the jth column you will get this expression ok. P 1 1 is going to be P i 1 is going to be this, P i 2 is going to be this and P i n is going to be this. They have to be multiplied with j, for a fixed j you are varying k. So, you will get the column.

So, if you take, if you take this product and sum it up over all values of k that is just the dot product of these two vectors; one row vector and one column vector and therefore, we can see that P 2, the 2 step transition probabilities is equal to P 1 times P 1. In fact, the m plus n transition probabilities by a similar reasoning is going to be equal to P m times P n, where P m denotes the n step transition probability and P n denotes the n step transition probability and they are in turn, going to be equal to, they are just going to be P raised to m the matrix raised to itself m times multiplied with itself m times. So, this is a, this is a brief introduction into what are Markov Chains or finite Markov Chains. Now, we will look at a certain property of Markov Chains called as stationarity or steady state distributions.

(Refer Slide Time: 24:37)



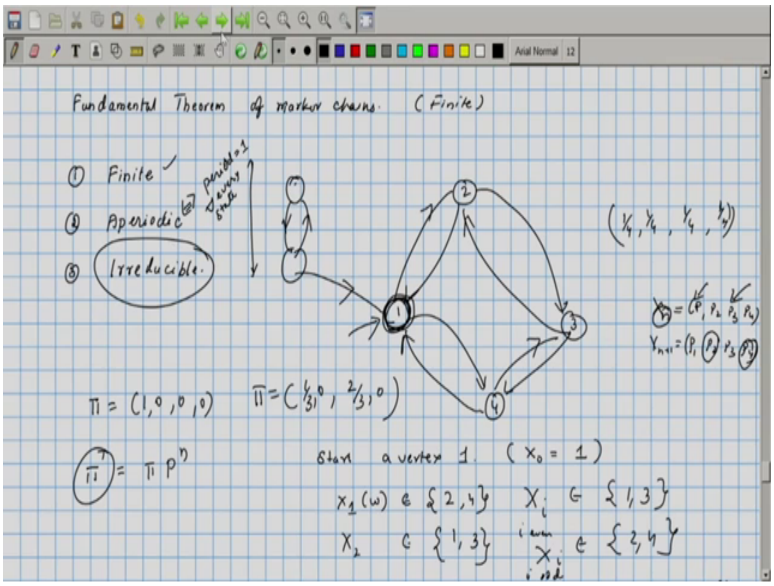
So, let us take an example that we are already familiar with. Let us say these were three states and you stay at each of them with one-third probability and transition to others with one-third probabilities. Now, if you start let us say your X 0, with all states being equally likely then what do you expect the distribution of X 1 to be? You can verify that it is going to be equal to 1 by 3 1 by 3 1 by 3 itself and further on for every X n. So, X n greater than 0 are all going to have this particular distribution.

So, the distribution of the Xi is are not going to change, if you start from this particular distribution. Now is this, because we had constructed, this particular example or is it a more general trend. So, suppose I had had this transition with probability half and this would I just say half and this would let us say one third, you go here with one fourth, here with one fourth, this with one fourth, come back here with one fourth, go here with one third, come here with one third.

Now, if you take this particular Markov Chain, it is not symmetric, but is there some distribution in which you could start the Markov Chain; that means, X 0 is equal to P 0 P or P 1 P 2 P 3 and then X 1 will exactly have the same distribution P 1 P 2 P 3 can there be such a distribution. In other words can there be a distribution Pi such that Pi is equal to Pi P ok. If there was such a distribution then that kind of a distribution, this is called as a steady state distribution.

The question that will bother us for the, for the next few lectures is what are the steady state distributions of Markov Chains, which kind of Markov Chains can have unique steady state distributions or stationary distributions. So, steady state distributions will also be called as stationary distributions. When will the stationary distributions be unique, when will there exist a stationary distribution, how can we compute the stationary distribution?

(Refer Slide Time: 28:04)



Fundamental theorem of Markov Chains tells us the answer to this question. It characterizes or tells us what are the Markov Chains, which can have unique stationary distributions. So, the requirements, so we are focused on finite Markov Chain, so the theorem states that the conditions required are 1, finite second is a periodic and the third condition is irreducible ok. If these three conditions are met then there the Markov Chains will have the, have these three properties. It will have a unique stationary distribution and it can be computed by just simulating the Markov Chains or can be approximated by just simulating the Markov Chains ok. So, we let us understand, what these properties are; finite we already know, the number of states should be finite.

Now, let us look at a periodic. So, let us look at this particular Markov Chain, where the states are; there are four states ok. So, in this by symmetry we can guess that the steady state distribution is going to be one fourth one fourth one fourth one fourth on each state ok, but it does not. So, there is one undesirable property at this particular Markov Chain, if you keep on running this Markov Chain that is if you keep on simulating the Markov Chain. You will not hit the steady state distribution, if the starting distribution was not properly chosen. For example, let me say that I start at vertex one with probability, I mean let us say I start with this, but at this vertex with probability 1 now. So, this means the random variable X 0 is equal to 1. Now, X 1 the random variable X 1 can take values.

So, X 1 of omega. So, the value of the random variable can be only 2 or 4. So, this must belong to the set 2 comma 4, it can go only 2 and 4. Now, X 2 the random variable X 2, the values it can take can be only 1 comma 3 and in fact, X i for any i which is even, can belong to 1 comma 3 and x i i odd can belong to 2 comma 4.

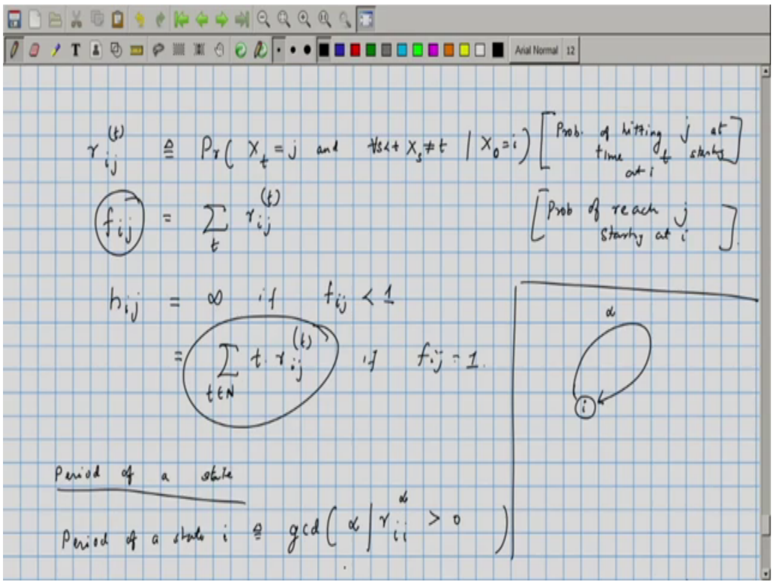
So, the probability is that you are in state 2 and 4 at even times is 0 and at odd times is 1 and therefore, there can if you just start at this state, I mean at at this configuration Pi, where Pi is given by 1 0 0 0, there cannot be a distribution, I mean you cannot mean keep on simulating this Markov Chain and reach a configuration Pi bar such that this will be close to steady state distribution, because steady state distribution means the probabilities should not change after I mean suppose X n has this distribution P 1 P 2 P 3 P 4 X n plus 1 also should have P 1 P 2 P 3 P 4.

Now, here if n was let us say odd then P 1 would have been 0 and P 3 would have been 0 and in the next day, if this was steady state then P 1 should have been 0 and P 3 should have been 0, but we know that is not going to be the case, because P 2 and P 4 are going to be 0 ok. So, by starting at this configuration, we cannot really hope to reach closer to the steady state distribution, this is not a problem with starting at a deterministic configuration. We could also have taken let us say Pi is equal to 1 by 3 0 2 by 3 0 ok.

So, one third probability here and my two third probability here and we know that similar thing would have been (Refer Time: 33:43) ok. So, the reason why you cannot start at these kinds of configurations is essentially the periodic behaviour of this, if you start at 1, you can reach back 1 only at even instances. Now, this is also true for let us say, I mean if you had a slightly more complex random walk or a Markov Chain.

So, suppose you started from I mean. So, this is the Markov Chain, you will reach this at some stage and then after that you can come back here only at even multiples ok. So, let us define periodicity. We will define it in terms of the period, I mean what is it? When do we call that a Markov Chain is a periodic ok.

(Refer Slide Time: 34:41)



So, let us first define the notion of a period ok. So, let us define before we do this let us define some other probabilities. So, first we will have this probability called r i j t ok. So, this is defined as probability of X n being equal to j. So, at time t you are at state j and X s is not equal to j for all s less than t given X 0 is equal to i that as r i j t is a probability that you are at state j at time t and you are not at state j during any at any smaller time given that you had started off at state i. So, the probability of reaching j starting at i for the first time at time t that is captured by the probability r i j and f i j is equal to sum over t r i j t.

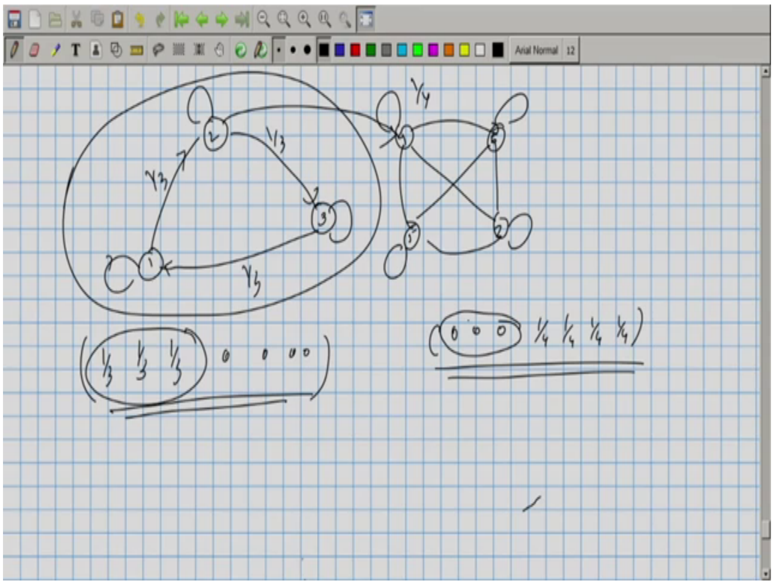
So, f i j denotes the probability that you will reach j if you start at i at some time. So, r i j we will call as the probability of hitting j at time t and this is the probability of reaching j all this starting it i. So, f i j is a probability and we can look at the expected time. So, h i j we can denote as the time taken to reach j starting at i. So, maybe if you are lucky you could reach in one step, if you are unlucky you could reach in 50000 steps and so on.

So, the expected value of the time taken to reach j starting at i is denoted by h i j and this we will say that this is equal to this expectation will be equal to infinity, if f i j is less than 1, if f i j is equal to 1 then this is equal to some t times r i j t, t belonging to naturals. So, if f i j is equal to 1 then this is the value of h i j, otherwise it is infinity.

What does period of a state ok. So, let us say you are at one particular node you want to; so, let us say that node is i, you want to come back to this state i ok. How much time does it take? Let us denote that by alpha and look at all such alpha the g c d of all those alpha's is what we will call as the period. So, we can write this as state i, this is g c d of a collection of numbers alpha such that r i i that is a start at i and come back to i in alpha steps, if you could start at i and come back to it in alpha step collect all those alphas. So, this is should be greater than 0.

So, look at all those alphas take their g c d that is the period and an a periodic Markov Chain, is a chain whose period is equal to what. So, this means period equals 1 for every state. So, look at every node in the Markov Chain or look at when if you had represented it in this particular fashion, which you can do for a finite Markov Chain. The period of every state should be 1 then we call it as an a periodic Markov Chain. We will need to look at the third requirement inside in, in the fundamental theorem, which is irreducibility ok.

(Refer Slide Time: 40:28)

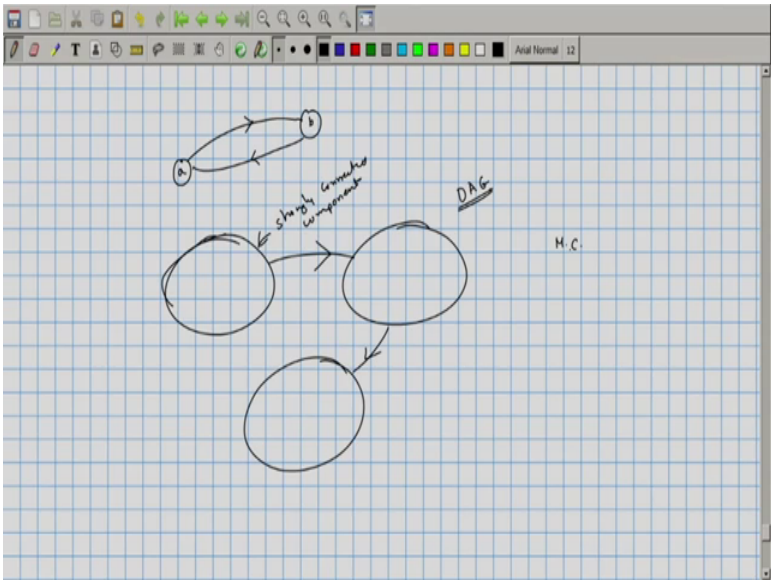


So, let us look at the following example ok. So, let us say this is a Markov Chain, where all the transition happens at one third probability and this is a Markov Chain in which all these happen with probability one fourth and we are taking the union of these. So, we have in total seven states.

Now, if you start in let us say 1 by 3 1 by 3 1 by 3 and 0 for the other four states then we know that we are just going to remain in this portion of the Markov Chain and therefore, the steady state distribution is again going to be 1 I mean; this is going to be a steady state distribution 1 by 3 1 by 3 1 by 3 by a similar logic 0 0 0 1 by 4 1 by 4 1 by 4 1 by 4. This is also going to be a steady state distribution.

So, if you take the union of these two, that is going to be a Markov Chain, which will not have a unique stationary distribution, if you had let us say a transition from here to here, the probability suitably adjusted, then you can see that this is not going to be a steady state distribution. This is going to be a step, this is going to be the only steady state distribution ok, but there are some states with 0 probabilities. So, we do not want to have these kinds of steady state distributions. We want steady state distributions in which every state has a non-zero likelihood of I mean every state has a non-zero probability. So, that is the essence that we want to capture by means of irreducibility.

(Refer Slide Time: 42:42)



So, we can formally define it as let us just look at the Markov Chain and we will say that two states are in the same connected component, if they can be reachable from each other. If there is a path from a to b and the path from b to a then you say that these two are in the same connected component. So, that so, look at a set of vertices such that they are all connected to each other, that will be called as a strongly connected component and there are no, so these are the vertices which are all connected to I mean the you can go from 1 vertex to another.

So, any graph can be broken down into strongly connected components where you can have edges between the strongly connected components, but it cannot be the case that I mean. So, when you break it down in the strongly connected component, it breaks down into a directed acyclic graph ok. There cannot be a cycle inside that, because if there is a cycle then 1 or more components might fuse together to give a even larger strongly connected component. So, you can break it down into distinct, you can break down any directed graph into strongly connected components and if we look at the strongly connected component for the Markov Chains that we were looking at, if it as precisely one strongly connected component then those kind of graphs are called as irreducible graphs or irreducible Markov Chains.

So, we have understood the three components of the fundamental theorem. First of all it should be the Markov Chain should be finite, second it should be a periodic and third it should be irreducible, under these conditions the Markov Chains will have a unique stationary distribution and that how we can compute that also we will see, we will see that in the next lecture.