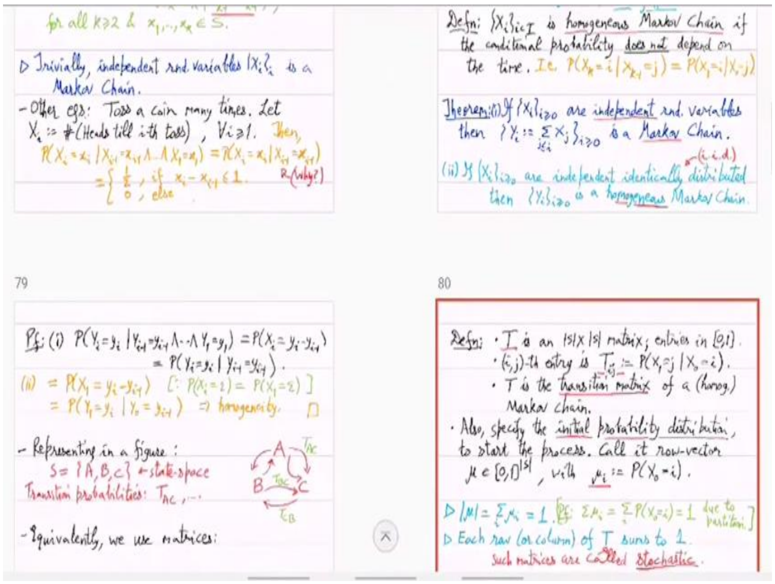
Probability for Computer Science
Prof. Nitin Saxena
Department of Computer Science and Engineering
Indian Institute of Technology – Kanpur

Lecture - 19
Drunkard's Walk, Evolution of Markov Chains

Last time we started discussing Markov chain which is a special very special kind of stochastic process.

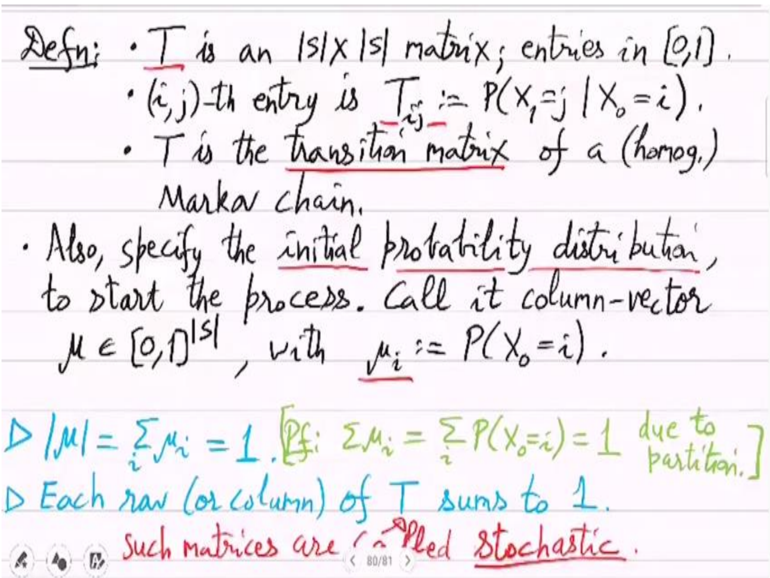
(Refer Slide Time: 00:24)



So, you are basically looking at a sequence of random variables which are in some sense memoryless. So, they do not remember a lot of the history but only the previous history just one step before. Moreover, we specialized Markov chains to homogeneous Markov chain where the conditional probability of k over k minus 1 is the same as what it was at any time before. So, in particular, 1 over X 1 over X 0.

That was the definition of homogeneous Markov chain. And we saw many examples. We saw this diagram on the state space. And then we started looking at a more systematic way to study it which is by transition matrix.

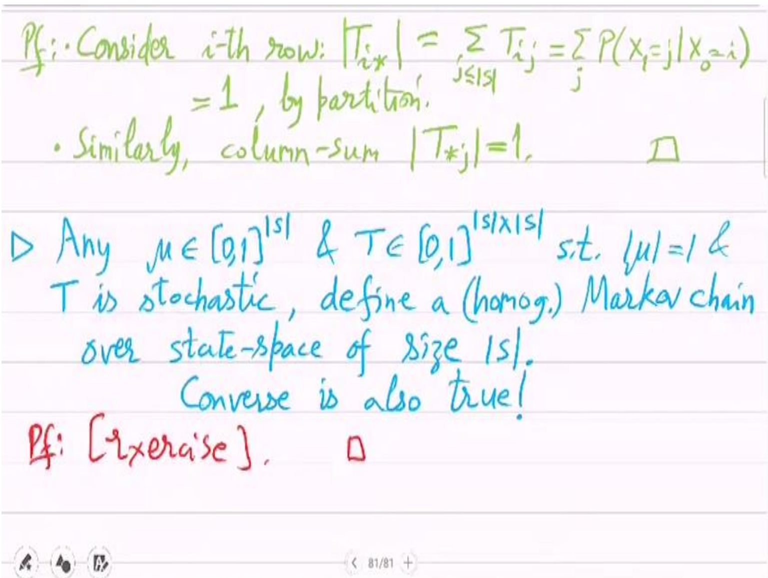
(Refer Slide Time: 01:18)



So, state space is S. Assume it to be finite. So, the transition matrix will be a matrix S cross S 0 1 entries. It will be stochastic matrix which means rows and sums will rows and columns will sum up to 1. That tells you the relationship between X k over X k minus 1. But then you also want to know you also need X 0. So, that is the initial distribution. So, we are calling it here mu which is a column vector. Think of it as a column vector.

Basically, it is just storing this specifying this probability. Probability of X 0 being i. That is in the ith entry. So, clearly this sum of mu i or this value of mu is 1 because of partition of probability of events. Next claim was that T is a stochastic matrix. So, why was why is that true? That is also for a simple reason.

(Refer Slide Time: 02:49)



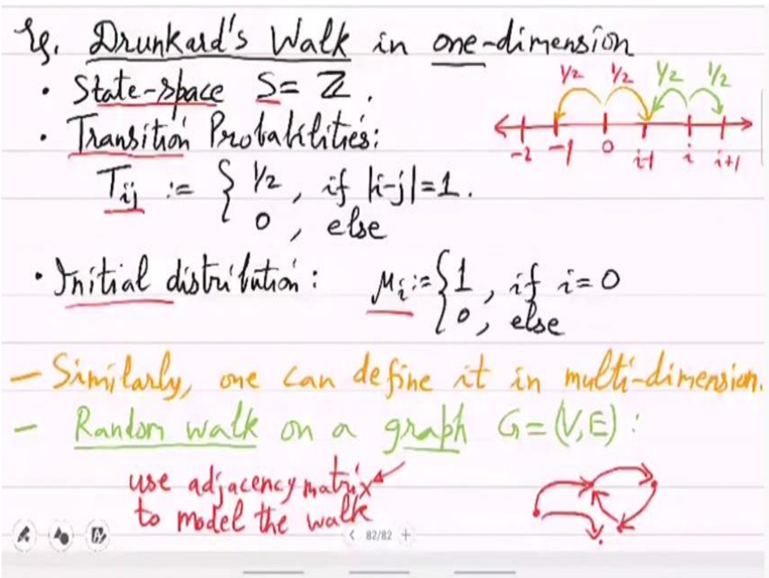
So, if you look at a row, consider ith row T i star. And look at the sum of that so which is nothing but T ij for all the j from 1 to S. So, this is the probability of going from value i to value j. And this can be in any time interval. In any point of time going from i to j is always the same probability. That is homogeneity. So, this is probability that X 1 is j given that X 0 is X 0 is i. What is the chance that X minus j and sum this up over all j?

So, again i is in this sum. i is fixed. It is only j that is changing and changing in all the ways all possible ways. So, this is again a partition. And so, you get equal to 1. That is the reason. So, you have seen that each row sum is 1 and similarly each column sum is also 1. In column, it will be j will be fixed and i's will be changing. But then it is changing in all possible ways. So, that again will be a partition of events. So, similarly, jth column.

So, those are useful properties. So, what you can show in general is that any vector column vector mu 0 1 entries well in the interval 0 1 entries and T S cross S matrix such that the value of mu is 1 and T is stochastic. Define a homogeneous Markov chain. So, you can start with any mu t value of mu 1 and stochastic matrix T. So, this will define a homogeneous Markov chain over state space of size S and vice versa. Converse is also true.

So, any homogeneous Markov chain can be written as mu and T and vice versa. So, it is a completely equivalent you can verify this by the definition. This is quite straight forward. Next, let us see an example.

(Refer Slide Time: 07:10)



So, which is drunkard’s walk in one dimension. So, as the name suggests a person who is drunk is trying to walk on a on the real line. And let us say started at T equal to 0. Now since the person is drunk he has very little memory. So, at any point of time, he can go either to the right side or to the left side and so on. So, from i the person can go to i minus 1 or i plus 1 and in a completely random way.

This position i the person can either go to i plus 1 or i minus 1. So, forward step or backward step with equal probability. That is a drunkard’s walk. So, here the state space S we are assuming it to be integers. So, these integer points is where the drunkard can move. One step forward, one step backward and both of them with equal probability. So, the transition if you look at the transition probability T ij, so Tij is half if so transitioning from position i to position j.

That is only possible if i and j are consecutive. So, either j is i plus 1 or i minus 1. Then movement is with probability half, otherwise 0. The drunkard cannot jump steps. That is the transition probability and finally the initial distribution. That is so mu i is 1 if i is equal to 0. And if for other positions it is 0 which means that it simply means that the drunkard started from position 0 at time t equal to 0. Position was exactly the origin.

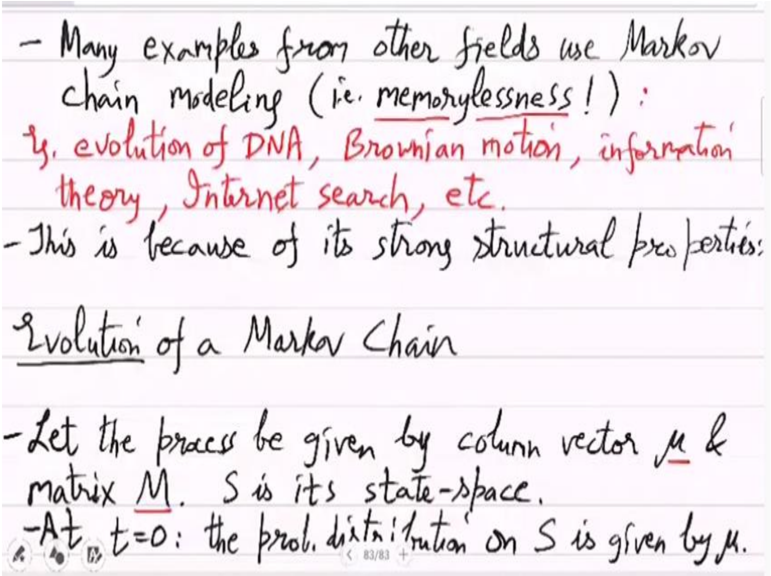
And then from here on either the drunkard moves to position 1 or to position minus 1 and so forth. So, this is drunkard’s walk. And you can see the matrix. It is actually infinite by infinite matrix with entries half or 0. It is a stochastic matrix. And in fact, you can imagine that in every row of the matrix there are only 2 places which have entries half. Others are all 0. So, every row is sparse, similarly, every column.

And in a similar way, so, similarly, one can define it in multi-dimensions. So, for example, if the drunkard is walking in a plane then you can draw a lattice and then you can give probabilities. There are 4 directions parallel to x and y axis. So, let us say to go anywhere chance is equal. So, it is one fourth or you can look at this imaginary walk of a drunkard man in a 3D space. Then there are 8 directions and so 1 by 8th each and so on so.

It gets more and more complicated. But it is possible. Another thing that you can do is you can define random walk on a graph. So, in a graph, it will be for example from this vertex to this vertex. One moves with some probability and so on. So, it can be quite general which again is similar to the diagram on the state space that we started with. But this can be a general graph with vertices and edges V and E.

In this case, we generally use the adjacency matrix of the graph. Adjacency matrix tells you whether i comma j is an edge. And now to get the transition matrix, you put the probability there. So, this is a very convenient walk to analyze and a very famous one as well.

(Refer Slide Time: 14:03)



So, many examples from other fields use Markov chain modeling which is basically transition that is memoryless. So, any phenomena ongoing phenomena that is memoryless all that it depends on is where you are now not on the history. That is a natural situation to use a Markov chain and homogeneous Markov chain. So, for example, evolution of DNA, or Brownian motion in physics.

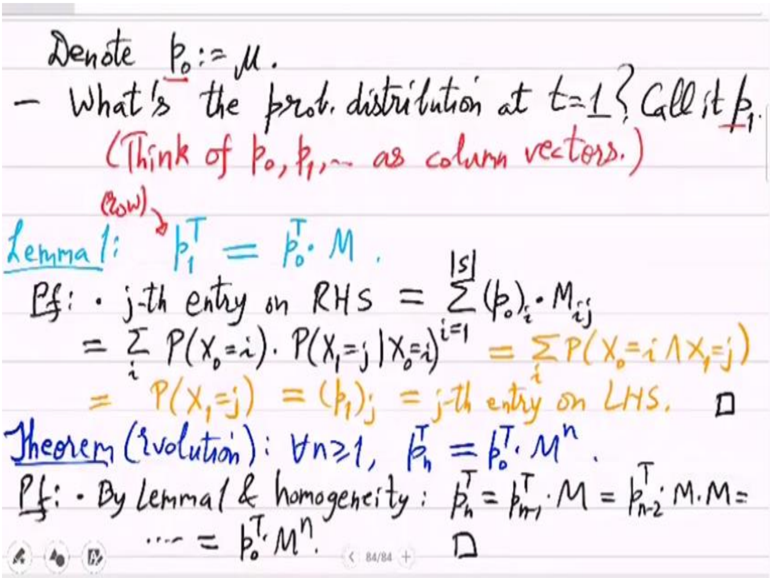
So, in Brownian motion, again these particles, atoms or molecules, they move but they do not seem to have a memory of the past. So, wherever they are based on that they can take certain steps. This is what has been seen in experiments in information theory. One example that we will see in more detail later is google search or let us say internet search etcetera. So, there are a variety a diverse set of examples that utilize Markov chain modeling.

And the advantage of this modeling is that you can use linear algebra and learn and predict a lot about the system. So, this is because of its strong structural properties. So, the first property that we will see is how does a Markov chain evolve. What are the equations? So, these random variables that the sequence which is changing with time. How is it changing? What is the evolution? And, how can you predict the future in this process?

So, let the process be given by column vector mu and matrix M. So, goes without saying that the column vector is the initial probability distribution and matrix is the transition matrix. So, it is stochastic in particular. So, now at t equal to 0, the probability distribution on S, so, that is the state space. Probability distribution on S is given by mu. So, basically the how did the process start?

What was the probability of being in some state ith state? So, that is t equal to 0 which we will denote by p 0.

(Refer Slide Time: 19:34)



That is t equal to 0, situation. Now, what is the probability distribution at t equal to 1? Call it p 1. So, think of p 0, p 1 as column vectors. So, p 0 is what the process started with that is the probability distribution. Now, in one step, how did the Markov chain evolve? What can you say about the probabilities on the state space now? So, notice that this is the understanding of Markov chain that we are taking.

That at any given point of time, there is a probability distribution on the state space. So, what is it at time t equal to n? So, let us write that precisely. So, the beauty is that this vector p 1 is related to p 0 as follows. It is simply a product. So, p 1 transpose is that is a row. So, column becomes a row. So, row times the matrix, so, which will again be a row. So, this is how the probability distribution is evolving.

So, that was the advantage of using matrices and vectors. You get this equation. And, how do you show this? Why is this true? So, let us look at the jth entry in the right hand side. So, what is that? So, that will be the ith entry of p 0. I mean essentially the you are you want to take the inner product of p 0 row with the jth column of M. Go over all i's which is what let us see it as probabilities. So, this is probability that X 0 is i.

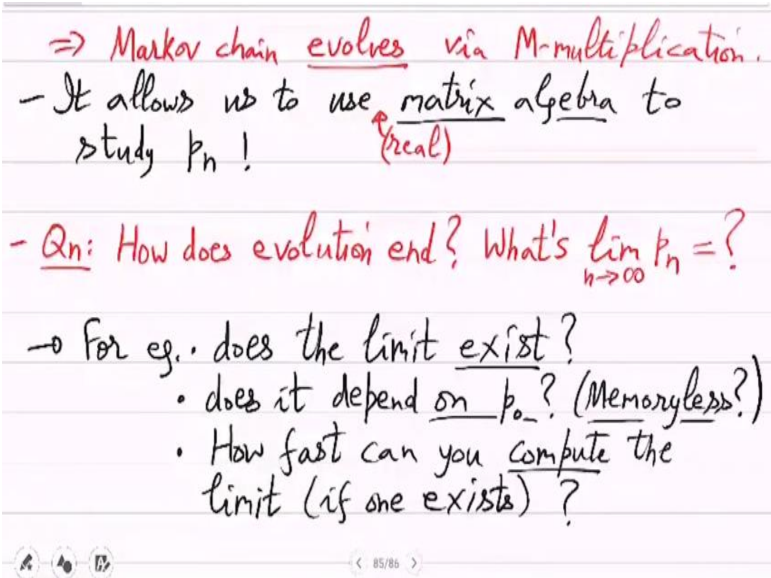
Since, this does not depend on time, so, we will just look at the transition actually we are looking at transition from X 0 to X 1. So, the random variable X 0 took value i and then the transition probability from i to j is X 1 is j given X 0 is i. And, what is that? So, this is probability of X 0 i and X 1 j. This is the intersection probability going over all i. So, this is the sum is then by partition formula just probability of X 1 j which is exactly the jth entry of p 1.

So, jth entry on left hand side. So, that finishes the proof. Entry by entry you can compare that the 2 rows left hand side right hand side are equal. This is what you have to remember that evolution is by matrix multiplication. So, let us write down now the bigger theorem which immediately follows actually that for all n, t equal to n. So, the first step the first let us say time then second t equal to 2 and so on till t equal to n.

If you look at the probability distribution p n transpose as row, then it is matrix multiplication n times on the initial distribution. What is the proof of this? So, this actually follows by Lemma 1 and homogeneity. So, p n transpose is just by homogeneity. The change from p n minus 1 to p n will be the same as that from p 0 to p 1 which was just 1 multiplication by M. And now, you can go to p n minus 2 and so on.

So, ultimately, you get p 0 transpose M to the n by induction. So, that is evolution by multiplication.

(Refer Slide Time: 26:17)



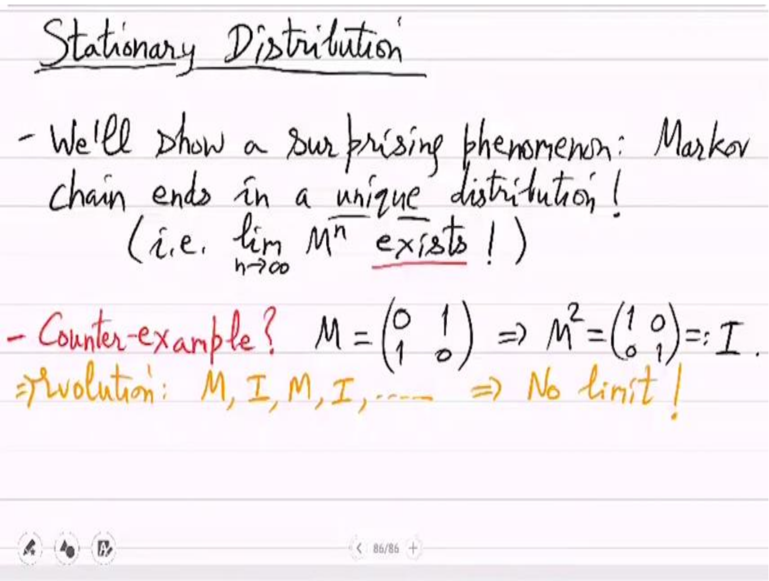
So, Markov chain evolves via M multiplication with matrix multiplication. And this is very nice because it allows us to use matrix algebra to study p n. So, we do not have to use very difficult math but just matrix algebra will do. And in fact, these are real matrices. So, real matrix algebra will suffice for the analysis. So, the question now that we will ask is what happens to this evolution in the end. So, evolution has started but how will it end.

That is, what is the limit of this probability distribution as n tends to infinity? That is the next question. This is what you want to predict. And in fact, completely understand that in the end if you keep waiting then in the end, what is the probability distribution? Does it really depend on the matrix? Does it depend on initial distribution? And, can you compute it? So, there are a lot of questions here. And so, we come to the topic of stationary distribution.

So, for example, does it exist? Can ask this question, does the limit exist? It, does it depend on? It should depend on the matrix, the transition probabilities of course. But, does it depend on where you were in the beginning? Or, is the beginning forgotten? Because, it is a memoryless process. So, why should the beginning matter, only the transitions should matter? And, how fast can you compute the limit if it exists?

It is the question of computation. You have this you have the transition probabilities. Now, can you compute the limit? Because n you cannot wait for n infinity. That is forever. So, is there a fast algorithm to compute it? So, these are the questions and we will answer them very convincingly in the next topic which is stationary distribution.

(Refer Slide Time: 30:33)



So, as the name suggests most probably the limit does exist. This is what we will show. So, that is a surprising thing that Markov chain ends in a unique distribution. In fact, so, unique meaning that it does not depend on the start the initial probability distribution. So, we are actually saying that the limit of the matrix itself matrix powers exists. This is what we are aiming for.

That you are given a stochastic matrix and then you keep taking powers ultimately it will converge to some nice looking matrix. It seems to be an amazing property if true. But I can give you a immediate or you can think of an immediate counter example which is you take this matrix 0 1 1 0 2 by 2 matrix. You square it. You will get identity. And then you cube it, you will get M back.

So, the evolution that you are seeing is M, identity then again M, identity and so on. So, this is actually limit does not exist in this case. So, what is going wrong? So, next, we will see how to correct this.