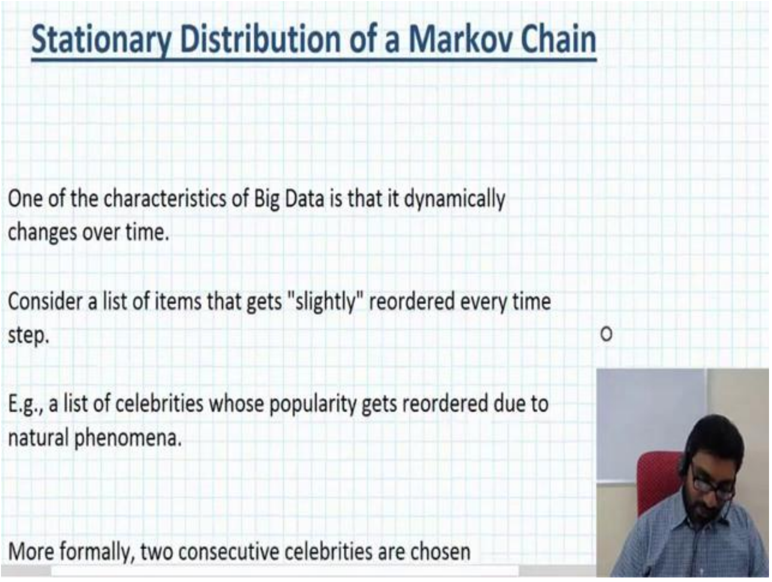
Algorithms for Big Data
Prof. John Ebenezer Augustine
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture – 17
Stationary Distribution of a Markov Chain

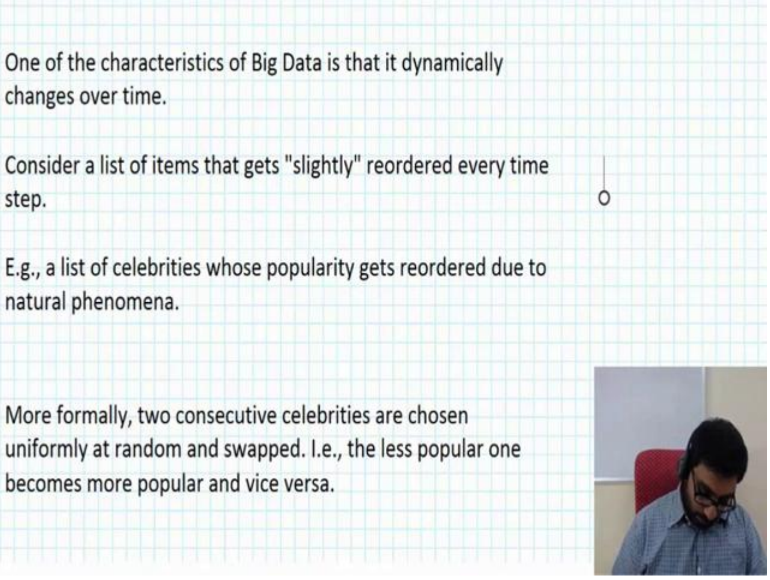
(Refer Slide Time: 00:14)



In the previous segments, we looked at the definition of Markov chains, and we looked at the notion of what makes some Markov chains more interesting than others, and we classified the states of a Markov chain, and generally built an understanding of Markov chains.

In today's lecture segment, we are going to little deeper, we are going to also look at an application; and inspired by that application we are going to study the notion of stationary distribution of a Markov chains. So, let us start by looking at this application. One of the key things about big data is that the data changes over time. the data is never static and it keeps changing over time in after in a fairly slow manner.

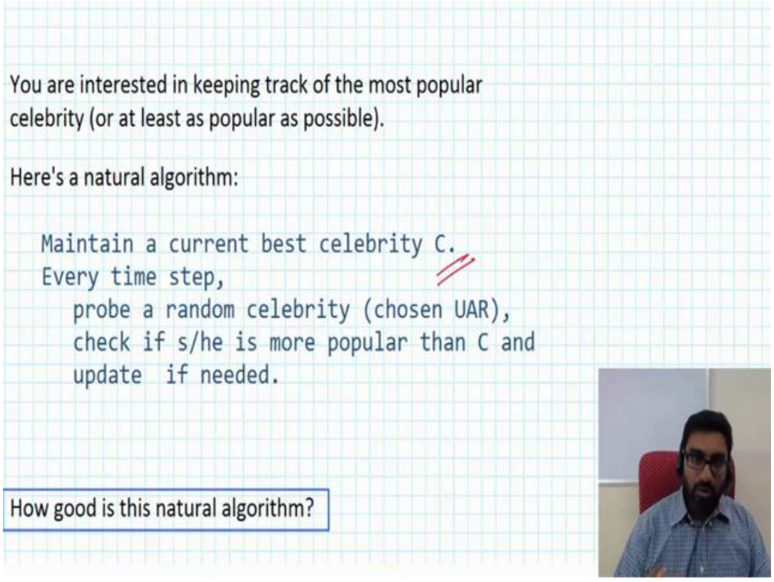
(Refer Slide Time: 01:07)



And let us look at a concrete example. Let us look at the list of celebrities, and this list is ordered by popularity. So, number 1 is the most popular celebrity, number 2 is the second most popular celebrity and so on and so forth. And this list is not static, I mean this list has I mean at any point in time such a list exists, but obviously the list way back and say 1950, it is going to be very different from the list that we have in 2016. So, this the list keeps changing over time.

So, for our purposes, we are going to look at this list in this formal sense, we have a list of celebrities ordered by their popularities, and some phenomenon - natural phenomenon takes place may be a hit movie or hit song or scandal or whatever changes the ordering of the popularity of the celebrities. And how do we model that we allow the basically a random consecutive pair is chosen uniformly of random, and the more popular of the two will be made the less popular one and the less popular would be made the more popular, so basically swapping their popularities. And this phenomenon can so this process itself is a random process and overtime the list of the popularity or a list keeps changing.

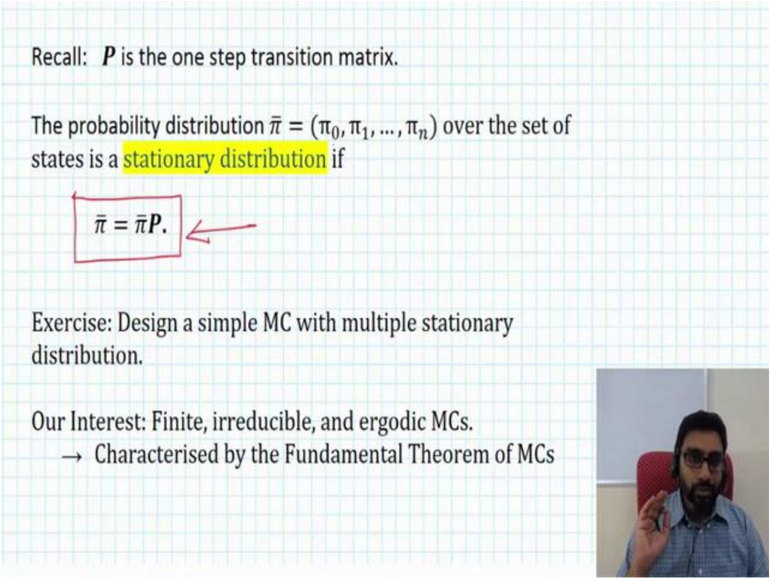
(Refer Slide Time: 02:56)



And we given such a changing list we are interested in maintaining the most popular celebrity and we want to do this without extending too much effort into it. So, here is very simple algorithm for it. And any point in time just maintain the most popular celebrity that you have currently seen so far, and you randomly probe a new celebrity, and check if this new celebrity that you have probed randomly is more popular than the celebrity C or not. And if the new probe celebrity is more popular then you update C that is it.

So, very simple natural algorithm, it is a perfect algorithm that can just run in the background. And eventually we hope to be able to report a celebrity that is fairly popular I mean the celebrity this due to the dynamic nature, we may not always report the most popular celebrity, because the moment you the most the current celebrity see that you have even if he or she is the most popular one due to dynamism can move down the list and by the time you correct yourself it is going to take some time. So, we cannot guarantee that we are always going to be reporting the most popular celebrity or we would not be at able to report a fairly high popular highly popular celebrity. So, we want to characterize how good this algorithm in that sense.

(Refer Slide Time: 04:30)

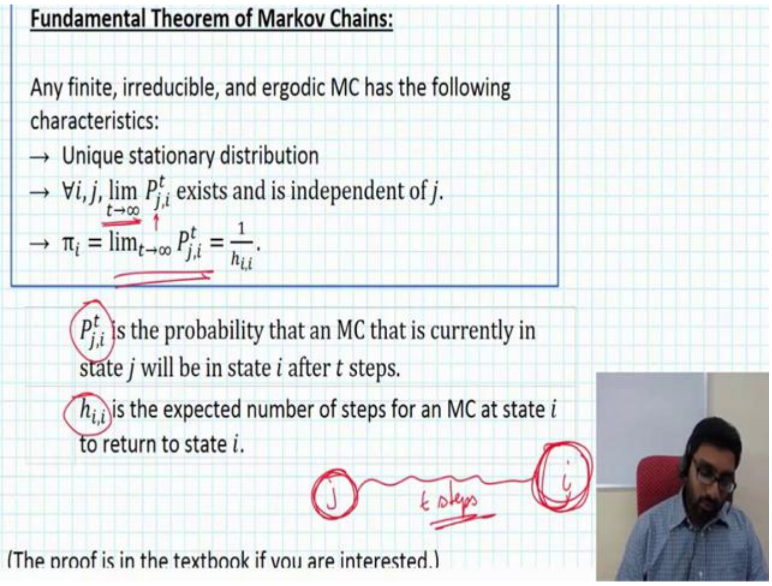


Well in order to do that we need to build an understanding of the theory of stationary distribution so that is what we are going to do now. Recall that we used to prop the notation P to denote the one-step transition matrix. And what is a stationary distribution, it is basically a probability distribution given by pi 0, pi 1 and so on up to pi n over the set of states in the Markov chain that obeys a condition.

What is that condition pi equals pi times P? So, notice you can you can think of this a distribution as a state in which the Markov chain will be infinitely far out into the future in some once (Refer Time: 05:33) some steady state. So, if the Markov chain is going to be in its state, so in one of the state 0 to n in accordance to this probability distribution then it will continue to be in one of the states according to this probability distribution. So, at some sort of a steady state behavior of the Markov chain and that is usually what we are interested in the context of Markov chains; we are interested in the steady state behavior of the Markov chains.

Here is a point in time we should take a few moments to design a simple Markov chain with multiple stationary distributions. And if it means that you need to pause for a few minutes please do that, but think about it carefully and design Markov chain with multiple stationary distributions. But eventually our interest is going to be finite irreducible and ergodic Markov chains, but do not worry about that you just need to design a Markov chain that is has that has multiple stationary distributions.

(Refer Slide Time: 06:49)

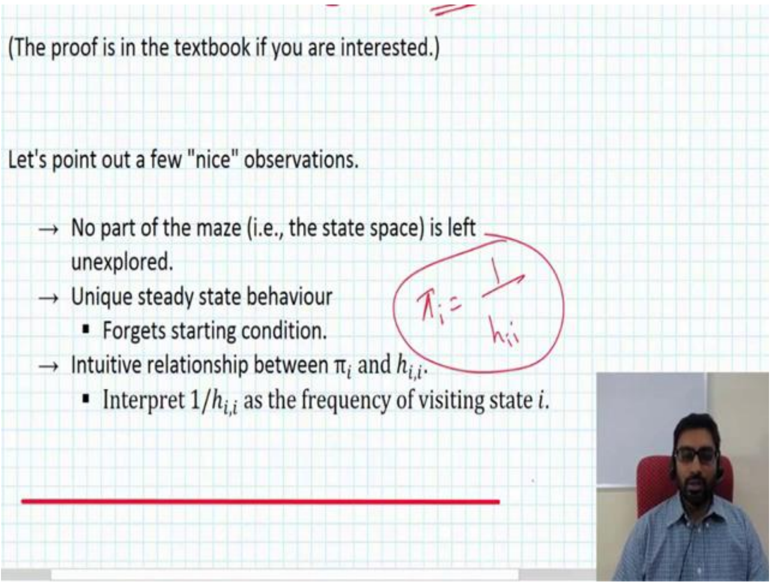


Let us now look at the fundamental theorem of Markov chains. And since we are interested in finite irreducible an ergodic Markov change that is what this theorem is going to talk about. And such Markov chains have a unique stationary distribution. So, if it has to be unique that means essentially the history is forgotten whatever be the way the Markov chain starts eventually it is going to reach a steady state, where the distribution is going to be the same, it cannot lead to multiple different distributions. It can only lead to one unique stationary distribution, so that is very important.

And. Secondly, the theorem states that well let us for remind ourselves what a piece of subscript j comma i superscript t is. It is the probability that a Markov chain that is currently in state j will be in state i after t steps. So, it is currently at state j, and after these t steps, the probability that the Markov chain will be in state i is this p j comma i superscript t and we are interested in that probability with t tending to infinity. And this theorem states that the probability exists and moreover it is independent of j, so this little j really does not matter regardless the way you start as t tends to infinity this probability is going to converge to same value.

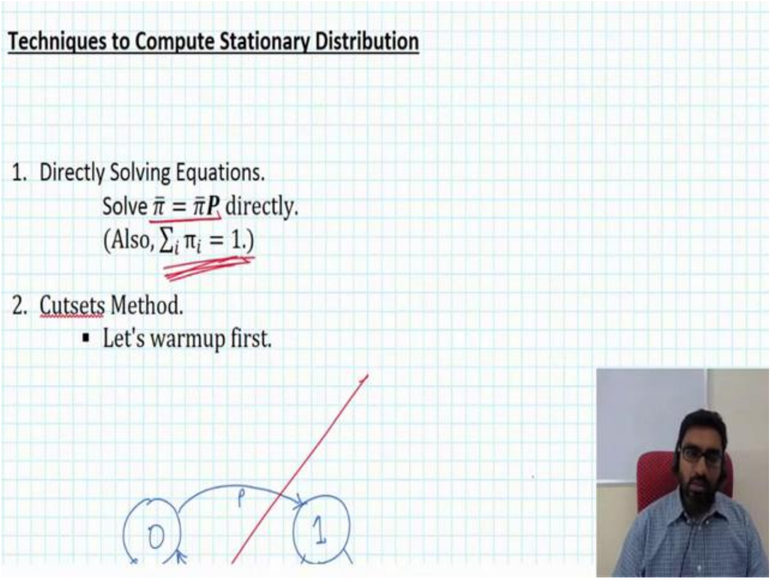
And finally, lets remind ourselves what h i i is, h i i is the expected number of steps for a Markov chain at state i to return to state i. So, your state i, how long does it take to you know take a walk, but then return back to i. And that is if you would look at 1 over h i i, it gives you a sense of how frequently you will be coming back to state i. And quite naturally that frequency is exactly equal to this limiting probability as t tends to infinity that p j comma i superscript t and that is exactly our pi i, the probability with which at steady state the Markov chain will be at state i. So, this is our fundamental theorem of Markov chains, take a few minutes to pause and make sure you understand the statements. The proof was also present in the textbook, but we are going to skip it for the purpose of our lecture.

(Refer Slide Time: 10:02)



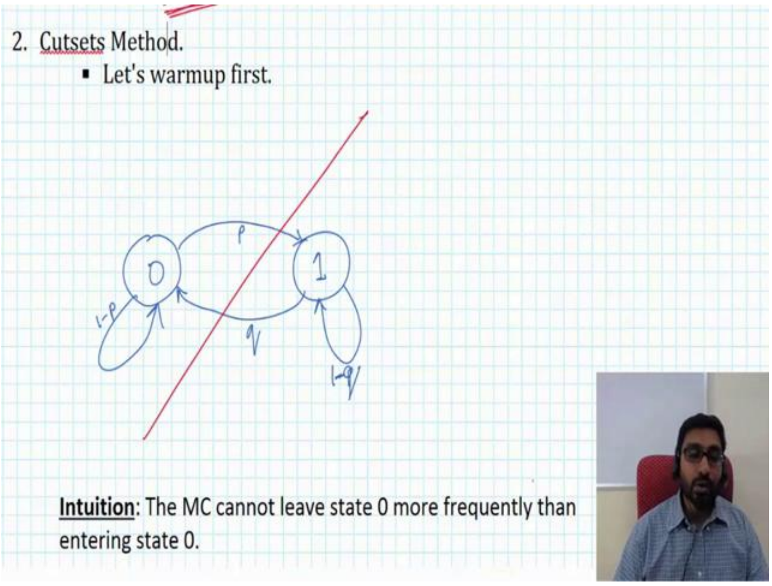
And just to reinforce our intuition, let us to make a few nice observations. Notice first that is the Markov chain for a finite irreducible and ergodic Markov chain, the stationary distribution is going to be such that no part of the maze is going to be left unexplored, every state is going be explored. So, all the pi i(s) are going to be strictly positive. And there is going be a unique steady state behavior initial conditions are forgotten. So, it does not matter with how you start eventually the Markov chains going to end up in a steady state behavior. And there is a nice clean intuitive relationship between a pi i and h i i, so that is pi i equals 1 over h i i.

(Refer Slide Time: 11:04)



Now, that we know what a steady state or what a stationary distribution is, and we are now interested in computing the stationary distribution, and so we have several ways to do that. The most obvious way is to simply solve a system of linear equations directly falling out of the definition of stationary distribution. So, a particular vector pi would be the stationary distribution if and only if it obeys this equality pi equals pi times p. And you have also one more equation since this pi is a probability distribution if we add up the components it has to add up to 1, so you get that is 1. So, that is just solving this system of linear equation, which is going to give you a way to solve value for pi, but this is not very feasible if the number of states in the Markov chain is very large.

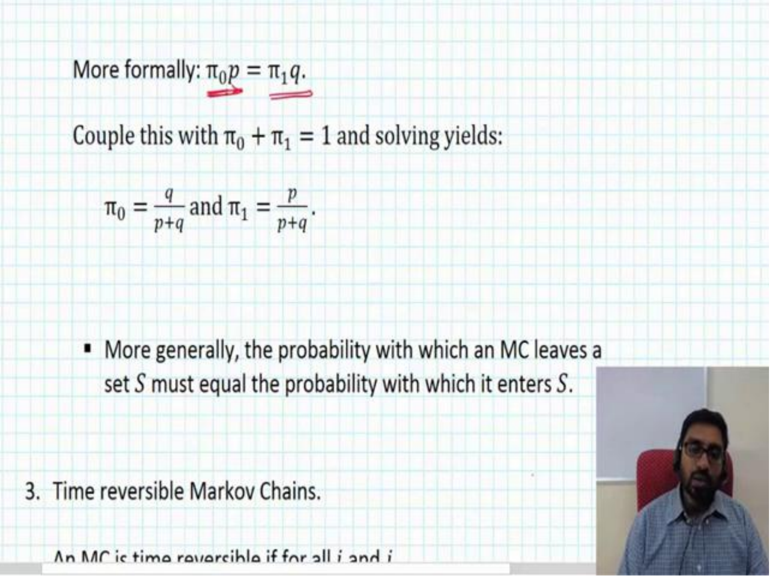
(Refer Slide Time: 12:11)



So, there is probably more intuitive ways to do that and one of them is the Cutsets method. So, let us warm up first, consider a very, very simple, but nevertheless very useful Markov chain. So, it simply has just two states 0 and 1. And with probability you p go from 0 to 1, and with probability q you go from 1 to 0. And these Markov chains these simple Markov chain can be used to moral a variety of things, where there is two states say good behavior and a bad behavior what is the probabilities. So, you may be interested in understanding what the probability is that you are going to be in the good state versus the bad state and these as you can see naturally leads us to the notion of stationary distribution. So, we are interested in understanding the stationary distribution or the steady state behavior of this Markov chain. So, there is a very clean intuition behind this Markov chain.

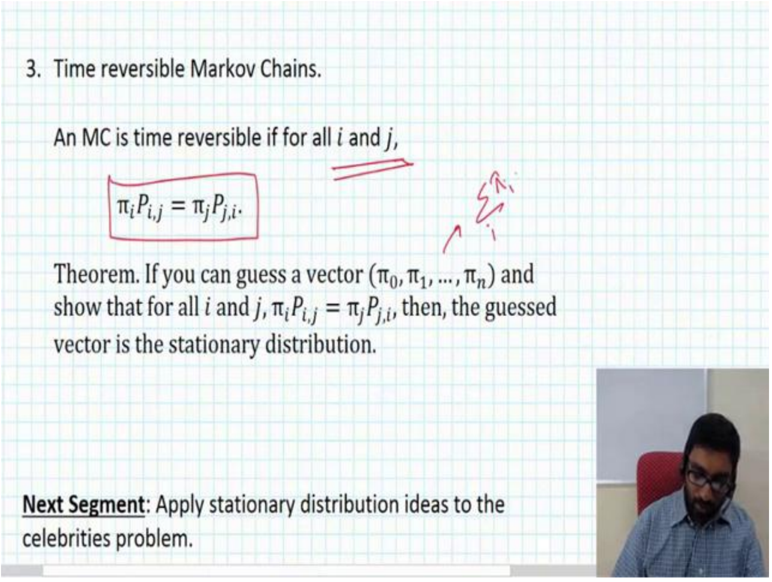
So, let us just consider the cut dividing state 0 and state 1. The frequency with which the stationary distribution moves from state 0 to state 1 must be equal to the state at the frequency with which it moves from 1 to 0. And this is some sort of a law of conservation of Markov chains, if you will making that up, but gives you some notion that you know you really cannot lose the Markov chain when it goes from one side to the other. So, if goes from one side to the other it has to come back from the other side back to the former side and so on and so forth. And we can use that to actually compute the stationary distribution. So, here is how to do that.

(Refer Slide Time: 14:02)



This notion of the frequency with which the Markov chain goes from left to right, and right to left is can be captured by pi 0 times p. Pi 0 is the probability with which the Markov chain is going to be at state 0. And given there is a state 0, it will actually move to the other side of the cut with probability p and vice versa pi 1 times q and these 2 are equal according to our intuition, so great. So, we have one equation. And the other equation of course, we can always resort to summing up the probabilities in the distribution it has to add up to over 1. And if we solve them we get pi 0 is q over p plus q and pi 1 is p over p plus q. So, as you can see we now have a handle on the stationary distribution of this simple, but nevertheless is very useful Markov chain.

(Refer Slide Time: 15:09)



And this notion can be generalized to Markov chains with a more complex structure. We will see one such Markov chain in the subsequent segment. The third technique to compute the stationary distribution applies only to time reversible Markov chains. A Markov chain is said to be time reversible, if for every pair of states i and j, this equation holds pi i times the probability of transition from i to j equals pi j times their probability of transition from j to i. And why this is called time reversible, well we are going to have an exercise problem dedicated to gaining and understanding of why this is called time reversible. So, for now let us just keep this definition did not mind.

And let us look at a theorem now the theorem does not depend as far as the statement goes does not depend on time reversibility, but as you can see this theorem will only be applicable when the Markov chain happens to be time reversible. If you can guess a vector, it is pi 0 pi 1 up to pi and so and you want to your hope to this guess is going to be the standard deviation, sorry this guess is going to be the stationary distribution. And if you can furthermore show that for all pairs i and j pi i times p i j equals pi j times p j i then the guesses in fact, a stationary distribution.

And of course, you also want the summation overall i pi i(s) equal to 1. And if this is sort of a guess and check technique for computing the stationary distribution of a time reversible Markov chain. So, we would now built sufficient ideas about stationary distribution in Markov chains and so on and so forth.

What we are going to do in the next segment is we are going to go back to the problem that we talked about at the start of the segment. And we are going to revisit that problem and we are going to gain an understanding of how good a celebrity, we will be or that simple algorithm will be outputting. And we will in that process exercise several notions of Markov chains in stationary distributions.