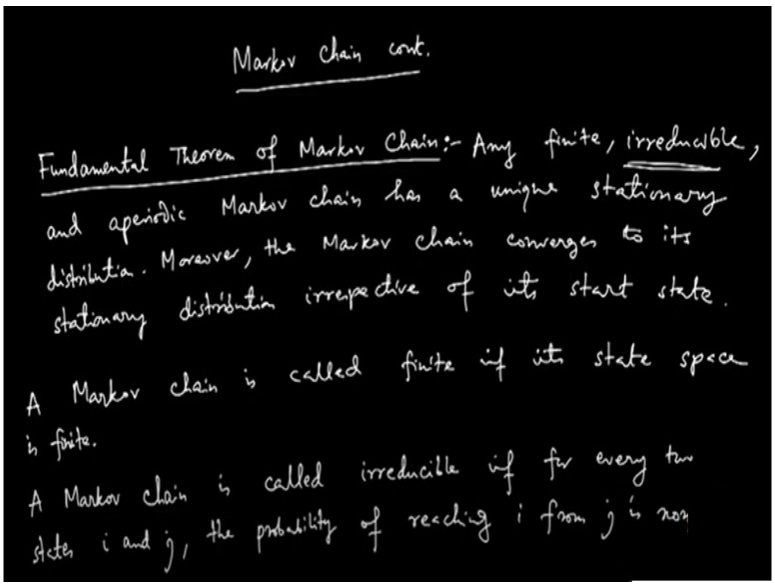
Selected Topics in Algorithm
Prof. Palash Dey
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Module No # 05
Lecture No # 22
Mixing Time, Reversible Markov Chain

Thank you welcome in the last class we have started discussing Markov Chain so in this class also we will continue our discussion on Markov Chain.

(Refer Slide Time: 00:36)

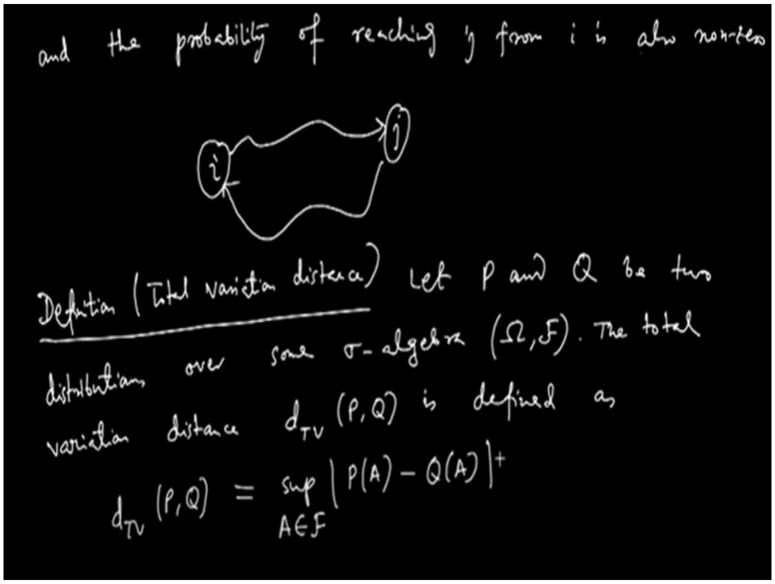


So Markov Chain continues, so now let me State the fundamental theorem of Markov Chain without proof fundamental. So it states that any finite irreducible and aperiodic Markov Chain has a unique stationary distribution. Moreover, the Markov Chain converges to its stationary distribution irrespective of its starts state. So in the last class we have looked at periodicity and aperiodic Markov Chains finite means the number of states is finite.

So a Markov Chain is called finite if its state space is finite. And just finite and approved city it ensures it guarantees stationary distribution but you know it does not guarantee unique stationary distribution. For that to see that you just take you know 2 disconnected copies of Markov Chain and it depends on the start state and there are various there are infinitely many stationary distributions and the convergence also depends on where it starts.

So what we need is connectivity that you know every state is reachable from every other state and that property is called irreducibility of the Markov Chain. So a Markov Chain is called irreducible if for every 2 states i and j the probability of reaching i from j is non-zero.

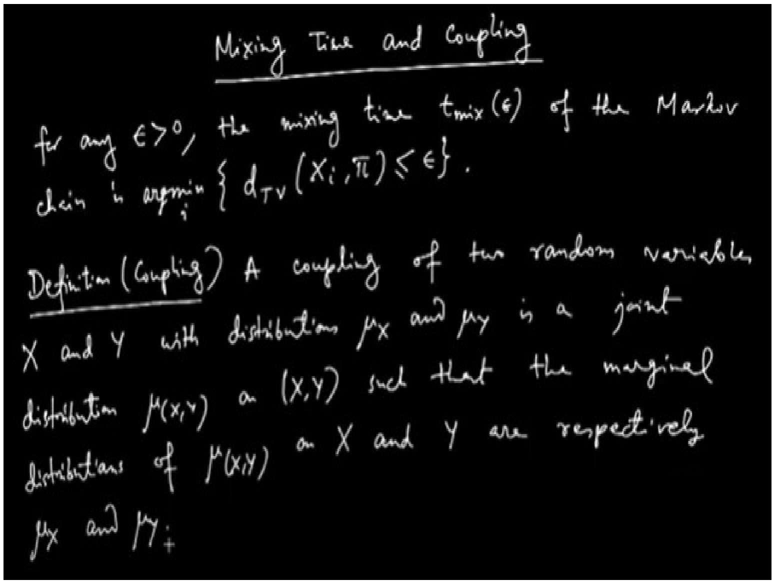
(Refer Slide Time: 06:01)



And the probability of reaching j from i is also non-zero that means basically from i there is a path to j and from j also there is a path to i. So we will not prove this prove this result because that proof is not very much with the test of the course but it is available in any standard textbook on Markov Chain. So use this result and now let us define some more concepts. So here is another definition which defines a distance metric on probability distributions.

So it is called total variation distance. So let P and Q be 2 distributions over some common space which formally is called sigma algebra under Sigma algebra consists of a set of outcomes and set of events so a f is the set of events. Then the total variation distance $d_{TV}(P,Q)$ is defined as $d_{TV}(P,Q)$ is supreme cover all events $S \in F$ $|P(A) - Q(A)|$. It is across all events the difference that these 2 probability distributions P and Q give to these 2 events. So this is needed for our next topic which is called mixing time and coupling.

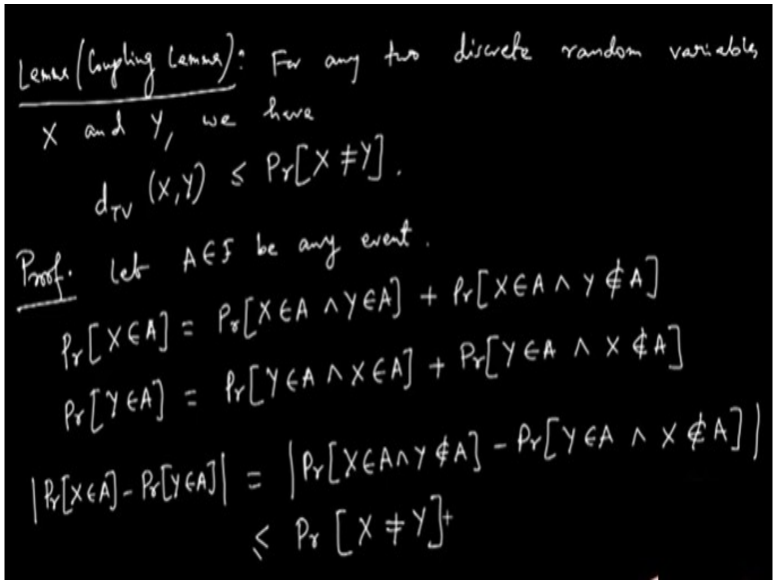
(Refer Slide Time: 09:37)



So what is mixing time? So for any $\epsilon > 0$ the mixing time t_ϵ or let me write this a t-mix mixing time of the Markov Chain is minimum over i such that total variational distance between X_i and π is less than equal to ϵ is arraigning. So beyond i the take the distribution of X_i and π they differ by at most ϵ the total variation distance at most ϵ. So this is called The Mixing time and now what is coupling?

A coupling of 2 random variables X and Y with distributions μ_X and μ_Y a coupling of X and Y with distributions μ_X and μ_Y is joint distribution $\mu_{X,Y}$. Such that the marginal distributions of $\mu_{X,Y}$ on X and Y are respectively μ_X and μ_Y. So it is a joint distribution were which respects the individual marginal distributions that is the cup that is called coupling

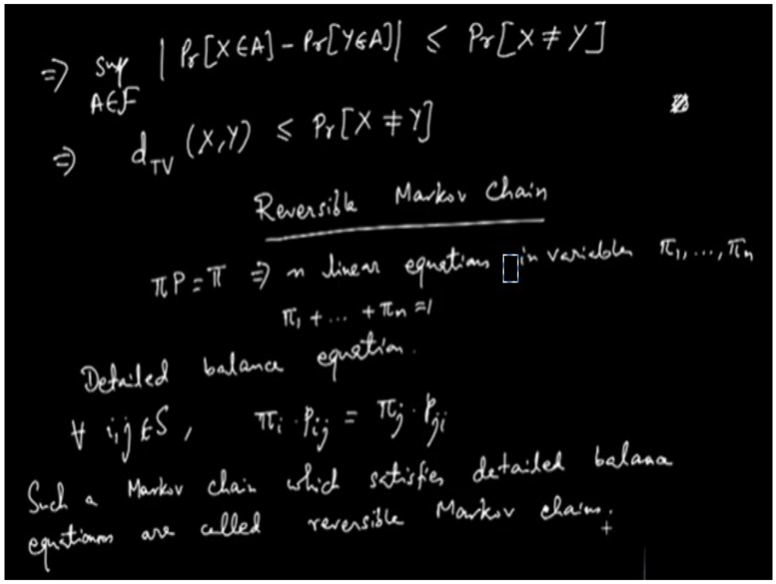
(Refer Slide Time: 14:01)



And what is coupling Lemma? This is called coupling Lemma. So for any 2 distributions for any 2 discrete random variables X and Y we have total variational distance between X and Y is less than equal to probability that X is not equal to y. So let us prove this Lemma proof. So let A be any event then we have the following probability that X takes value in A is probability that by law of total probability X takes value in A and Y takes value in A + probability that X takes value in A and Y does not take value in A similarly probability that y belongs to A is probability that Y takes value A.

And probability that X takes value A + probability that Y takes value A and X does not take value A. so if we subtract this and take mod then mod of probability X in A - probability Y in a this is mod of probability X takes value in A and Y does not take value in A - probability Y takes value in A and X does not take value A. So this is for one A so this is less than the probability that X not equal to Y and this holds for every event A.

(Refer Slide Time: 17:53)



And hence supreme of over A in if probability that X takes value in A - probability that y takes value in A this is less than equal to probability X not equal to Y. Left hand side is total variational total variation distance d_{TV} between X and Y this is less than equal to probability X not equal to 1. Which concludes the proof of this Lemma? So next we define some more concept of Markov Chain and there is a class of Markov Chains which are called reversible Markov Chain which is very common.

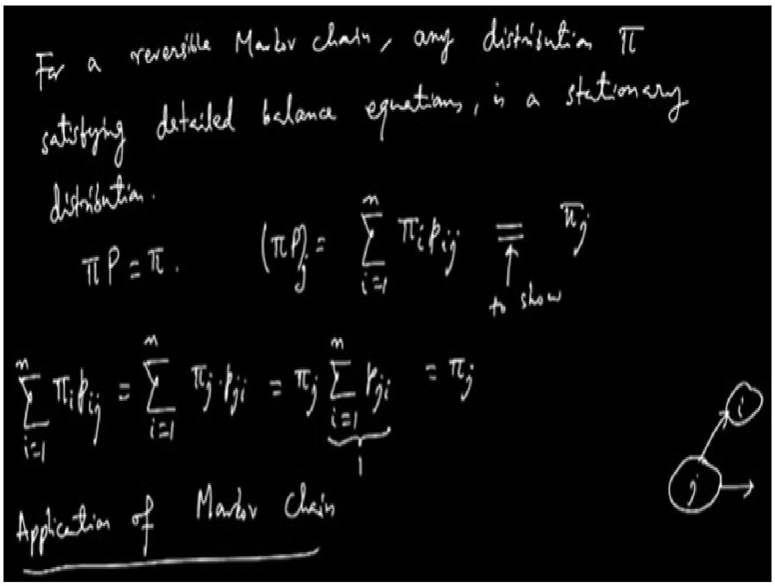
So let us define reversible Mark Chain. So Markov Chain is called reversible. So, what is the motivation? So the fundamental theorem of Markov Chain states that you know for every finite aperiodic and irreducible Markov Chain there exists a unique stationary distribution but how to compute the distribution? Recall the condition of stationary distribution is $\pi P = \pi$.

So it boils down to n linear equation solving n linear equations. So this is equivalent to n linear equations in variables $\pi_1, \pi_2, \ldots, \pi_n$ and of course these are distribution. So we have another equation $\pi_1 + \pi_2 + \ldots + \pi_n = 1$ so if we solve this n + 1 equations, We get a distribution we get our unique stationary distribution basically says under those conditions this state of this set of linear equations will have a unique solution.

Now it turns out that you know for some Markov Chains satisfy some extra condition which is called detailed balance equations. If a Markov Chain satisfies detailed balance equation it will be for this Markov Chain computing stationary distributions is much more straightforward. Now what is little balance equation it says that for every, i j any 2 state for every, i j in states is $\pi_i P_{ij} = \pi_j P_{ji}$ and these this holds for all pair of sets i j and if and these are the detailed balance equations and any Markov Chain.

Which satisfies detailed balance equation is called a reversible Markov Chain. So such a Markov Chain which satisfies detailed balance equations is called reversible Markov Chain. So let us see and for reversible Markov Chains you know this π's will be the stationary distribution.

(Refer Slide Time: 23:13)

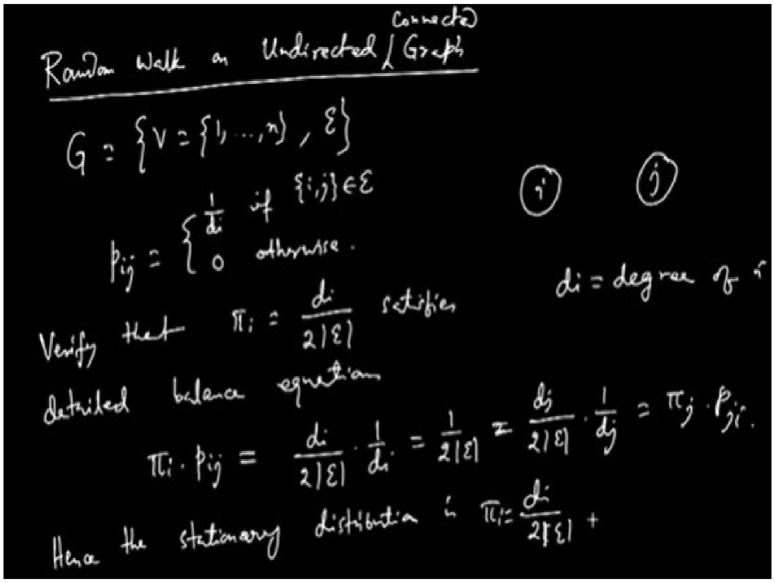


So for see that you know let us see why this is easy for a reversible Markov Chain let us show that you know any set of any distribution π which satisfies detailed balance equation is a stationary distribution for reversible Markov Chain any distribution pi satisfying detailed balance equations is stationary distribution. So let us prove this so what stationary distribution that means it satisfies $\pi P = \pi$ that means what is πP?

And this I need to show this is to show this should be equal to π_j for stationary distribution. So suppose π is a distribution which satisfies detailed balance condition so what is this the jth coordinate of πP? $\sum_{i=1}^{n} \pi_i P_{ij}$ now apply detailed balance conditions $\pi_i P_{ij}$ is same as $\pi_j P_{ji}$.

Now π_j is independent of i so it comes out π_j 1 to, n P_{ji} now what is summation of P_{ji}? So here is the state j and these are the probabilities this is i. So this is the sum of the probabilities of outgoing edges which sums to one because it is a stochastic Matrix. So this sum is one this is equal to π_j. So this shows that for a reversible Markov Chain any distribution which satisfies detailed balance conditions must be a stationary distribution. Now let us see how this can be used. So now we see some applications our first application is application of Markov Chain.

(Refer Slide Time: 27:47)



The most important application is what we will see we will study at depth is random walk on graphs on undirected graphs. So suppose G be a graph on vertex it is 1 to n and edge set is E and at every vertex it picks one of its outgoing neighbors outgoing edges uniformly at random. So P_{ij} here is i here is j if there is an edge from i to j it picks that edge uniformly randomly among all its neighbors so suppose d i is the degree of i.

So p_{ij} is $\frac{1}{d_i}$ if there is an edge from i to j. So undirected graph so there is an edge $\{i, j\} \in E$ and 0 otherwise now let us find out what is the stationary distribution assume connected underrated connected graph suppose it is connected. So there is no isolated vertex so that means this is defined degree is never 0. So, we now you know verify that $\pi_i = \frac{d_i}{2|E|}$ satisfies detailed balanced equations.

So what is retail balance equation let us recall $\pi_i p_{ij}$ should be equal to $\pi_j p_{ji}$ now if there is an edge between if there is no edge between i and j then both p_{ij} and p_{ji} are 0. So both $\pi_i p_{ij}$ and

$\pi_j p_{ji}$ are 0. So if there is no edge between me and j then this equation is satisfied so if there is an edge so let us see what is um how this equation is satisfied then $\pi_i = \frac{d_i}{2|E|}$.

But $p_{ij} = \frac{1}{d_i}$ this is $\frac{1}{2|E|}$ which can also be written as $\frac{d_i}{2|E|} \frac{1}{d_j} = \pi_j p_{ji}$. So it satisfies the detailed balance equations hence the stationary distribution is $\pi_i = \frac{d_i}{2|E|}$ this is one of the stationary distribution and see how it is how easy it is to compute because it is because it is a reversible Markov Chain. It satisfies a detailed balance equation. So we will continue from here in the next class.