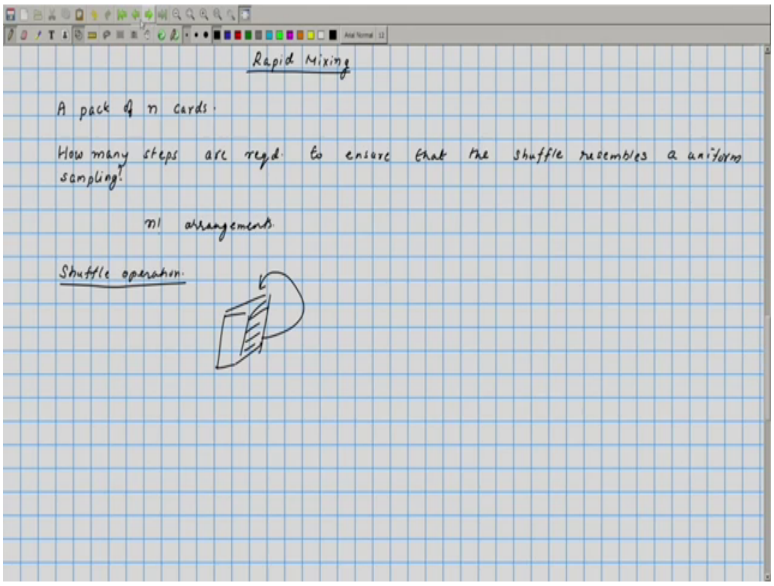
Randomized Algorithms
Prof. Benny George Kenkireth
Department of Computer Science & Engineering
Indian Institute of Technology, Guwahati

Lecture - 21
Rapid Mixing

So, in this lecture we will learn about Rapid Mixing.

(Refer Slide Time: 00:33)



Rapid Mixing

A pack of n cards.

How many steps are reqd. to ensure that the shuffle resembles a uniform sampling?

n! arrangements.

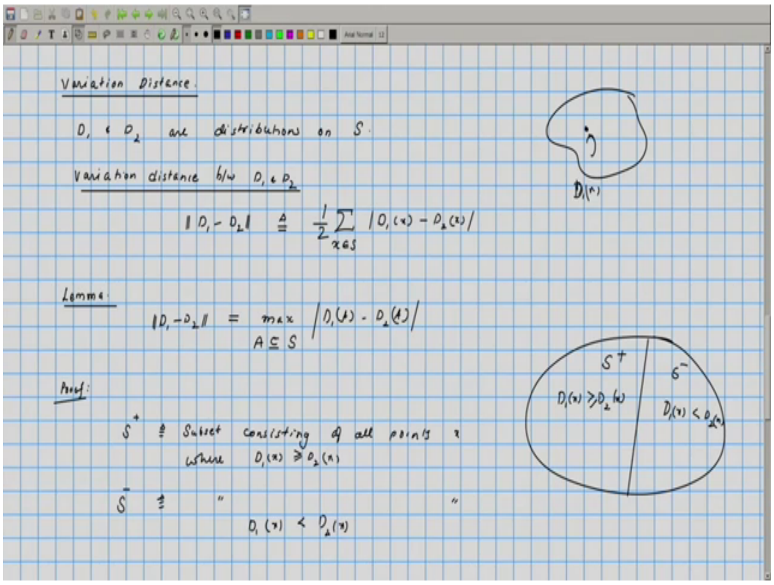
Shuffle operation.

So, the basic question that we want to know something of this kind. So, assume that we have n cards, how many times do you have to shuffle so that the shuffle becomes something like a uniform sampling. How many steps are required to ensure that the shuffle or the arrangement of cards resembles a uniform sampling? Its a little bit big question we will make it more precisely later on ok.

So, if we draw what we want to really have is let us imagine all the possible arrangements of cards say in factorial arrangements. When we sample after shuffling, how many times should we rearrange such that the arrangement that that at a particular time looks like one that is obtained while doing a uniform sampling from this n factorial arrangements.

So, we will describe what our shuffle operation is ok, suppose we have a deck of cards ok, we pick one card and then put it on top ok. So, amongst these n cards, we pick 1 at random uniformly at random, I will put it on the top. So, this is our shuffle operation. So, how many times do we have to do this shuffle operation so that the arrangement looks like the one obtained by uniform sampling. This is the first end that we want to answer. So, we were introduced a notion of how close it is to uniform sampling ok.

(Refer Slide Time: 02:45)



Variation Distance.

D_1 & D_2 are distributions on S.

Variation distance b/w D_1 & D_2

$$\| D_1 - D_2 \| \triangleq \frac{1}{2} \sum_{x \in S} | D_1(x) - D_2(x) |$$

Lemma:

$$\| D_1 - D_2 \| = \max_{A \subseteq S} | D_1(A) - D_2(A) |$$

Proof:

$S^+ \triangleq$ Subset consisting of all points x where $D_1(x) \geq D_2(x)$

$S^- \triangleq$ " " $D_1(x) < D_2(x)$

So, we will define something called as variation distance ok. So, let us imagine that D 1 and D 2 are distributions on a sample space ok. The variation distance between D 1 and D 2 so that is written as D 1 absolute value or D 1 minus D 2 ok. Norm of D 1 minus D 2 this is defined as sum over all the elements of the sample space the absolute value of D 1 x minus D 2 x this divided by 2.

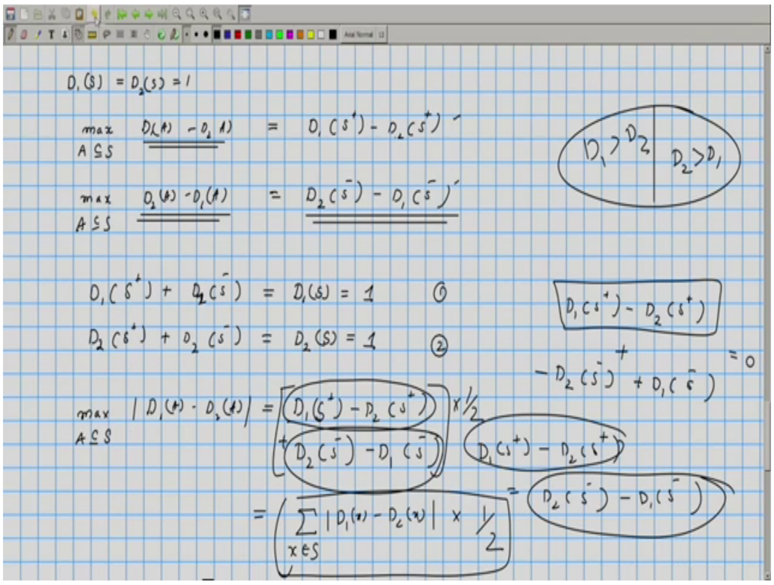
So, at each point in the sample space compute for each point compute D 1 x. D 1 x is the probability that is x is chosen if the underline this equation was D 1 and D 2 x is the probability if the underlined equation was D 2. Look at their difference; take its absolute value, sum of sum it up over all the points of the sample space and divide the sum by 2. What we get is called as variation distance ok.

Now, we will just relate this variation distance to some other quantity ok. Say if you look at D 1 minus D 2 the variation distance, this is equal to the maximum over all subsets of S of the sample space D 1 A minus D 2 A, it’s absolute value ok. So, look at all possible subsets of the sample space and look at the difference between the probabilities that these distributions assigned to the set or the event A I mean that will be equal to the variation distance; proof is very straight forward.

So, let us look at all these points and if you look at all points inside the sample space, there are some points some collection where D 1 x is greater than or equal to D 2 x which we will call as S plus and the others are where, D 1 x is less than D 2 x that we will call as S minus ok.

So, S plus is the subset consisting of all points, where D 1 x. So, all point x where D 1 x is greater than or equal to D 2 x and S minus is similarly the subset consisting of all points where D 1 x is less than D 2 x.

(Refer Slide Time: 06:19)



$D_1(S) = D_2(S) = 1$

$\max_{A \subseteq S} \underline{\underline{D_1(A) - D_2(A)}} = D_1(S^+) - D_2(S^+)$

$\max_{A \subseteq S} \underline{\underline{D_2(A) - D_1(A)}} = \underline{\underline{D_2(S^-) - D_1(S^-)}}$

$D_1(S^+) + D_1(S^-) = D_1(S) = 1$ ①

$D_2(S^+) + D_2(S^-) = D_2(S) = 1$ ②

$\max_{A \subseteq S} |D_1(A) - D_2(A)| = \left[D_1(S^+) - D_2(S^+) + D_2(S^-) - D_1(S^-) \right] \times \frac{1}{2}$

$= \left(\sum_{x \in S} |D_1(x) - D_2(x)| \times \frac{1}{2} \right)$

$\boxed{D_1(S^+) - D_2(S^+)} + -D_2(S^-) + D_1(S^-) = 0$

$D_1(S^+) - D_2(S^+) = D_2(S^-) - D_1(S^-)$

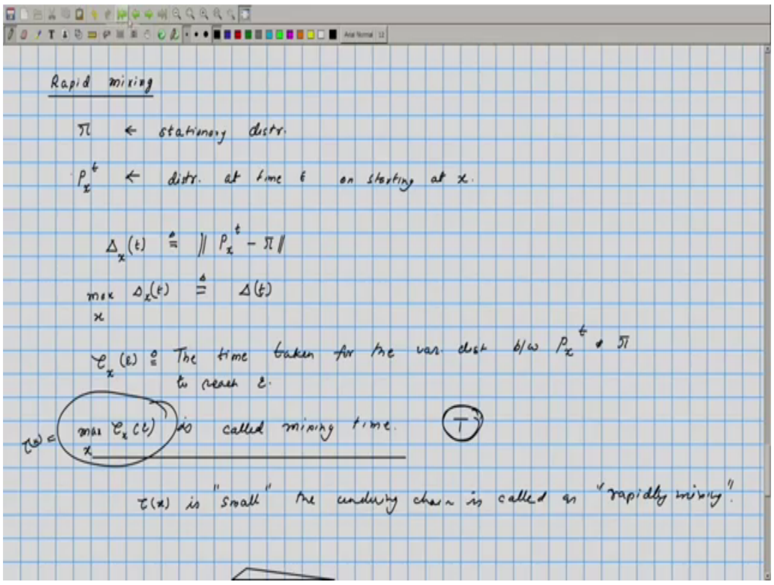
Now, note that if you take the entire sample space since this is the distribution D 1 S is equal to D 2 S and that is this is equal to 1. Now, let us look at this quantity, let they are interested in. Max over all subsets D 1 A minus D 2 A, it is absolute value. Clearly if you take S plus that will be the A which maximizes this quantity. So, if you take D 1 A minus D 2 A, this is maximized when A is S. So, this is equal to D 1 S plus minus D 2 S plus and if you look at D 2 A minus D 1 A and if you try to maximize that quantity that is maximize when A becomes S minus.

So, this is equal to D 2 S minus minus D 1 S minus ok. S minus consists of all points where D 2 dominates. So, on that subset D 2 S minus minus D 1 S minus will be maximized ok. If we look at D 1 S plus plus D 1 S minus that is two disjoint parts of the sample spaces. So, this is equal to D 1 S and is equal to 1. Similarly D 2 S plus plus D 2 S minus will be equal to D 2 S and that will also be equal to 1. So, max overall subsets absolute value of D 1 A minus D 2 A this is equal to both these quantities and they those quantities will essentially be equal. If we take this is equation 1 and equation 2, they subtract them what we will get is D 1 S plus minus D 2 S plus plus minus of D 2 S minus plus D 1 S minus and at the right hand sides are both being equal to 1 they subtract out and gives is 0.

So, D 1 S plus minus D 2 S plus is equal to D 1 S plus minus D 2 S plus is equal to D 2 S minus minus D 1 S minus. So, these quantities are the same. So, therefore, D 1 A minus D 2 A is maximized at this point and D 2 A minus D 1 A is maximized at S minus and they are equal. So, the maximum of the absolute value is just D 1 S plus minus D 2 S plus plus D 2 S minus minus D 1 S minus the whole divided by half. Now if you take D 1 S plus minus D 2 S plus this consist of one region which consists of all points, where D 1 is greater than D 2 and the second portion is all those points were D 2 is greater than D 1.

So, we can write this entire expression as sum over the absolute value of D 1 x minus D 2 x, where x belongs to S and this into half. So, this is nothing but the variation distance. So, variation distance can also be seen as the maximum possible difference that can be there for the probabilities of events when you look at those events under distribution D 1 and D 2 ok. Now, that we understand what is variation distance.

(Refer Slide Time: 11:25)



Rapid mixing

$\pi \leftarrow$ stationary distr.

$P_x^t \leftarrow$ distr. at time t on starting at x.

$$\Delta_x(t) \overset{\Delta}{=} \| P_x^t - \pi \|$$

$$\max_x \Delta_x(t) \overset{\Delta}{=} \Delta(t)$$

$\tau_x(\varepsilon)$: The time taken for the var. dist b/w P_x^t & π to reach ε.

$\tau(\varepsilon) = \max_x \tau_x(\varepsilon)$ is called mixing time.

$\tau(\varepsilon)$ is "small" the underlying chain is called as "rapidly mixing".

We will want to look at the number of steps ok; we need to define what is called as Rapid Mixing. We will do this by defining rapid mixing in terms of variation distance ok. So, let us say so, intuitively what rapid mixing means is the Markov chain quickly moves towards it is stationary distribution. So, let us say pi is what we will call as the stationary distribution and let P x t denote the distribution at time t on starting at x.

So let be x any point in the state space of the Markov chain and let the Markov chain evolve for t seconds or t times steps. How close is it to the stationary distribution? If it approaches the stationary distribution quickly, then we will say it is the rapidly mixing Markov chain. So, we will define certain terms. So, let us say delta x t ok. So, this is the variation distance this is the distribution P x t is the distribution and pi is the distribution.

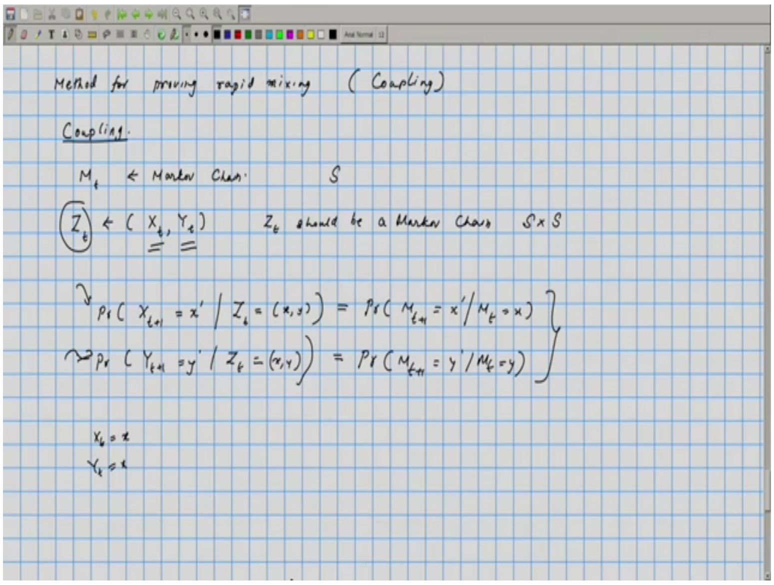
So, let us look at the variation distance between P x t and pi ok. So, this we can think of as a difference between the distribution, if you allow the Markov chain to start at state x and evolve it for t time units. Depending upon the value of x, the this number could be different and the maximum overall choices of x delta x t is what we will call as delta t.

We will define some more quantities say tau x epsilon, we will define as the time taken for the variation distance between P x t and pi to reach epsilon ok. So, if you had evolved this Markov chains by starting at x how much time will it take before the variation distance becomes less than epsilon. So, that is denoted by tau x t and the maximum value of tau x epsilon when x ranges over all possible states, this is defined as the mixing time ok. So, the mixing time is small.

What it means is for every state, if you begin a Markov chain at that particular state, the number of steps taken for the variation distance to be smaller than epsilon is at most this value ok. Suppose the mixing time is t, then in t steps we can guarantee that the variation distance between the distribution and the study state distribution will be no more than epsilon ok.

So, if tau x epsilon the maximum or all x is called as tau x and tau x is what we call as a mixing time; this tau x is small. We will later on described as to what it means to say that the tau x is small, but when it is small then the underlying chain is called as rapidly mixing ok. So, in a nut shell what this means is we now have a definition for telling that Markov chain mixes quickly. Now, we want to know or when we say that it mixes quickly means that it reaches its steady state distribution quickly or we want to know how can be proved that is certain Markov chains are rapidly mixing.

(Refer Slide Time: 16:29)



Method for proving rapid mixing (Coupling)

Coupling.

$M_t \leftarrow$ Markov Chain. S

$Z_t \leftarrow (\underline{\underline{X_t}}, \underline{\underline{Y_t}})$ Z_t should be a Markov Chain $S \times S$

$$\left.\begin{aligned} &Pr(X_{t+1} = x' \,/\, Z_t = (x,y)) = Pr(M_{t+1} = x' / M_t = x) \\ &Pr(Y_{t+1} = y' \,/\, Z_t = (x,y)) = Pr(M_{t+1} = y' / M_t = y) \end{aligned}\right\}$$

$X_t = x$

$Y_t = x$

So, methods for proving rapid mixing ok, we will see once such method; the method is called as coupling ok. So, we will define what is a coupling we will require another definition ok. So, let M t be a Markov chain and let this Markov chain beyond some stage space S. Now let us take another Markov chain Z t which consists of two parts ok. So, X t and Y t. So, it consists of two processes X t and Y t and we say so Z t also we wanted to be a Markov chain and the Z t is a Markov chain on the cross product space on S cross S.

Now we can think of many such Markov's chain for example, if you take a simple random walk on a graph, now if we take a copy of it and do these things together; then that is a that can be something like Z t ok. So, now, when do we say that this is a coupling of the chain M t ok?

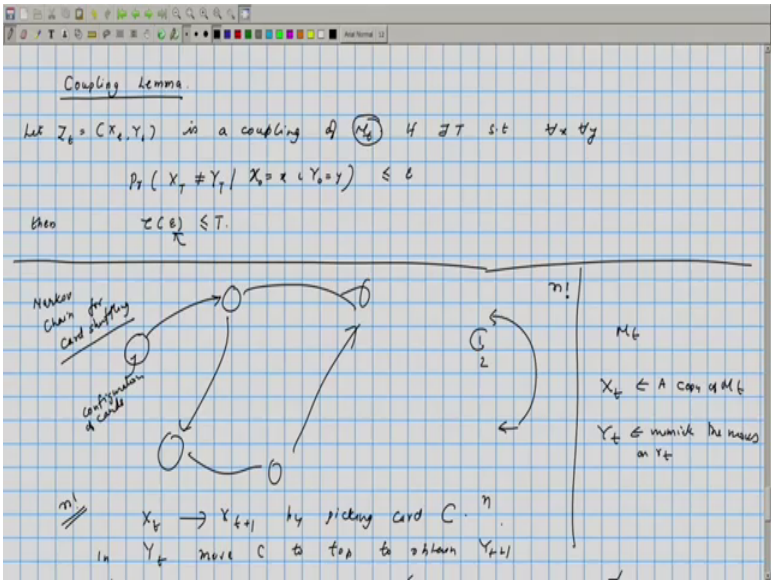
So, if you look at the marginals since Z t consists of two parts X t and Y t if we look at the marginals, they should behave exactly like the Markov chain M t. In other words, probability that X t plus 1 is equal to x prime given Z t is equal to x comma y. This probability should be same as probability that M t plus 1 is equal to x prime given M t is equal to x.

And similarly, the other marginal also should behave similarly that is Y t plus 1 is equal to y bar given Z t is equal to x, y should be equal to the probability that M t plus 1 is equal to y bar given M t is equal to y. If these conditions are met, then we will say that this is a coupling ok. Z t is the coupling of the Markov chain M t.

So, one way to obtain coupling would be to take identical copies which are completely independent of each other ok, but that is not the kind of coupling that we will be needing in order to prove rapid mixing. What we will look at is a coupling where in these X t and Y t, they are in they are own rights Markov chain, but once they have a tendency we will try to make couplings in such a way that this X t and Y t has a tendency to move towards a common state and once they reach the common state they continue in the that particular state ok. So, when they reach a common state that is if X t is equal to x and Y t equals x ok, when such a situation has been reached we will say that the chains have coupled ok.

So, once again we first define what is a coupling? Coupling just means that the marginals should behave like the underlying Markov chain ok. We say that they are coupled if the individual Markov chains X t and Y t have coalesced into a particular state. From there onwards they stay in that state even if they continue in the same state, we know that these equations will be satisfied because both of them will have identical marginals ok. So, now, we have seen what is coupling and what is a coupled Markov chain?

(Refer Slide Time: 20:47)



Coupling Lemma.

Let $Z_t = (X_t, Y_t)$ is a coupling of M_t if $\exists T$ s.t $\forall x \; \forall y$

$$Pr(X_T \neq Y_T \mid X_0 = x \;\&\; Y_0 = y) \leq \epsilon$$

then $\tau(\epsilon) \leq T$.

$X_t \leftarrow$ A copy of M_t

$Y_t \leftarrow$ mimick the moves on Y_t

$X_t \rightarrow Y_{t+1}$ by picking card C.

In Y_t move C to top to obtain Y_{t+1}

Now, we will have a lemma called as coupling lemma, it is known as coupling lemma and this is the key tool that we will use to show that certain Markov chains are rapidly mixing. While we were thinking about rapidly mixing, we mention that the Markov chains should quickly move to its stationary distribution, but we did not say that what exactly is quickly ok. We will do that when we are looking at the examples of using coupling lemma ok.

So, let us take coupling lemma suppose Z t is equal to X t, Y t is a coupling of some Markov chain M t. If there exists a T such that probability that X t not equal to Y t given x 0 is equal to x and y 0 equals y. If this is less than epsilon, then tau epsilon is less than or equal to t ok. Let us see this lemma closely. So, what we have is, we have coupling of Markov chains.

So, X t and Y t are the marginal. Now suppose somebody could guarantee that the probability that these do not become this X t and Y t probability that they do not become equal a small irrespective of this starting state ok. No matter which state you begin; the probability that X t and Y t are different.

If that probability is bounded by epsilon that means, the probabilities at most epsilon, then the mixing time for parameter epsilon is going to be less than T. In other words, in time T they would have mixed to the extent of variation distance being no more than epsilon ok. We will see an application of this lemma. So, let us we will solve the card shuffling problem using this ok. So, in our card shuffling problem, we will think of a Markov chain in the following way ok.

So, individual states are a is one particular configuration of cards ok. So, the total sample space is the total stable space of the Markov chain is of size n factorial. Each possible arrangement is one of the states and from each state what we will do is if the arrangement was say 1 to up to n, we choose one position at random and then put it on the top of the pile ok. So, at each node what we are doing is and whenever we are in a particular state, what we are doing is we will choose one of these cards at random and put it at the top of the pile and we keep on doing this.

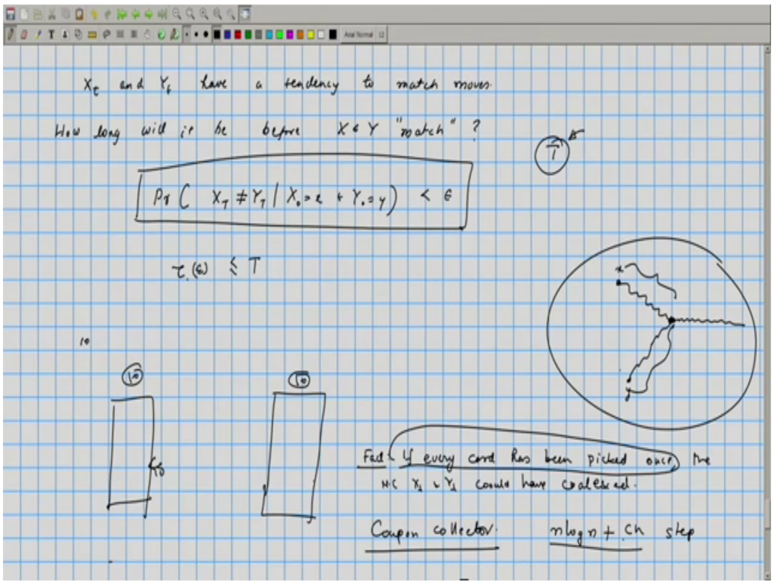
So, this is the M t for us ok, the underlined Markov chain or the Markov process is M t, wherein the states are configurations and from each state goes to one of the neighboring states. So, you can note that the sample space is of the state space is of size n factorial, but from each node there are exactly n minus 1 transactions to make we can also just redundantly say that the first when we are choosing one at random, we choose the first node also with the equal probability in which case we just stay in that state.

So, we can say that they are at most or exactly n neighbors for each particular sample state in this or state space. So, we have one such Markov's chain and what we will do in order to construct the coupling is the following X t is just a copy of M t and Y t is going to be; so, Y t is going to mimic the moves on X t ok. What does it mean to mimic the moves on X t? So, X t is basically moving from one particular configuration into another by choosing one particular card ok.

So, the state at any time t is fully captured by the arrangement of the deck. Now, suppose we moved from X t to X t plus 1 by picking a card say C. In Y t what we will do is in Y t move C to top to obtain Y t plus 1; that means, if card number 10 was bought to the top to obtain the next state in X t, in Y t also you just pick the card 10 wherever it is and put it on the top of the file ok.

So, of course starting configurations could be different for these two different chains, but X t and Y t there is some kind of relationship between them ok. What is interesting is these chains have a tendency to move towards a configuration such that the states match and if you have these states match, then from that point onwards they will remain in the particular config; they will they will have x the subsequent states being equal.

(Refer Slide Time: 27:17)



X_t and Y_t have a tendency to match moves.

How long will it be before X & Y "match" ?

$$\boxed{Pr(X_T \neq Y_T \mid X_0 = x \,\&\, Y_0 = y) < \epsilon}$$

$$\tau(\epsilon) \leq T$$

Fact: If every card has been picked once, the MC X_t & Y_t could have coalesced.

Coupon collector. $n \log n + cn$ step

So, X t and Y t have tendency to match moves ok, what we need to know is how long we will it take for X t and Y t to be identical and how do we use coupling lemma? So, coupling lemma states that if the probability that they do not know match if that is small after a particular time t, then the time t is going to be a bound on the mixing time ok.

So, let us ask this question; how long will it be before X and Y match ok. In other words, what we want to know is probability that X T equal to or let us look at the complimentary event not equal to Y T given X 0 equals X and Y 0 equals Y ok. We want this probability to be less than epsilon can be find the time t such that this statement is true. If we could find such a time, then we can use the coupling lemma and say that the mixing time for the variation distance to be smaller than epsilon or mixing time for parameter epsilon will surely be less than T ok. So, how good or how small a T can be; find out how much. So, X t and Y t are to be different ok.

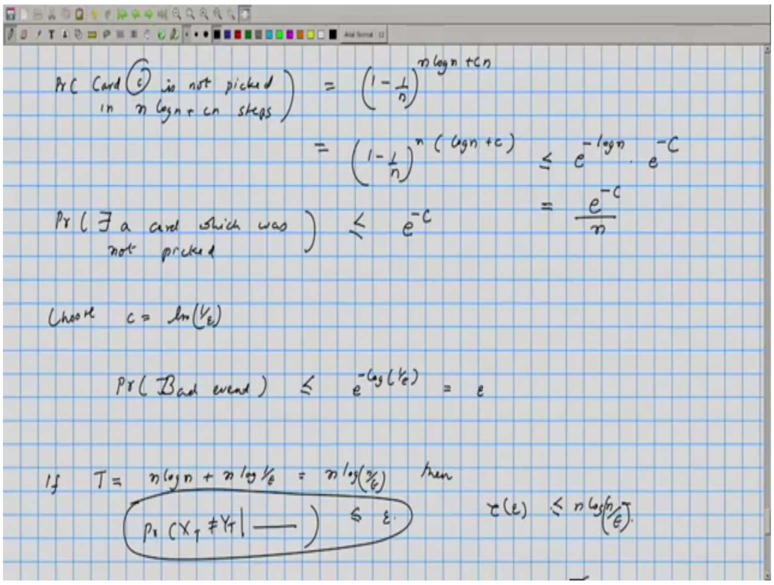
So, when can we say that X t and Y t are same? X t and Y t are same means their stage space have coalesced ok. So, if you think of this entire thing as a starting your random walk at some particular point x and the other random process Y we will start at some particular small y; they will move around this is also move around and finally, they will reach a point and after that they will continue on the same path ok. We want to know how long will it take before this happens ok.

Since the deck is completely characterized by the arrangement of cards and we know that if some card let us say if card 10 has a ever been picked, their position in both cards we will then always be same ok. So, this is some particular card, if card 10 was picked and then it was bought to the top, then in the other deck also the card 10 would have been bought to the top from wherever it was.

So, once a card is picked, later on arrangements we will not change the position of 10 that holds true for any other card that you pick. So, if every card has been picked once, then X t and Y t would coalesced ok, so, that we will write it as a fact. If every card has been picked once, the Markov's chain X t and Y t would have coalesced ok. So, now, what we need to tell is how long will it take for every card to be picked at least once ok, that is just coupon collector ok.

So, let us say that we keep on picking the cards for n log n plus c n steps; c we will choose later on. So, if we keep on stimulating the Markov's chains for n log plus c steps each step every card has a probability of being picked.

(Refer Slide Time: 31:55)



$$\Pr\left(\begin{array}{c}\text{Card } \textcircled{i} \text{ is not picked} \\ \text{in } n\log n + cn \text{ steps}\end{array}\right) = \left(1-\frac{1}{n}\right)^{n\log n + cn}$$

$$= \left(1-\frac{1}{n}\right)^{n(\log n + c)} \leq e^{-\log n} \cdot e^{-C}$$

$$= \frac{e^{-C}}{n}$$

$$\Pr\left(\begin{array}{c}\exists \text{ a card which was} \\ \text{not picked}\end{array}\right) \leq e^{-C}$$

Choose $c = \ln(1/\varepsilon)$

$$\Pr(\text{Bad event}) \leq e^{-\log(1/\varepsilon)} = \varepsilon$$

If $T = n\log n + n\log 1/\varepsilon = n\log(n/\varepsilon)$ then

$$\Pr(X_T \neq Y_T \mid \text{——}) \leq \varepsilon.$$

$$\tau(\varepsilon) \leq n\log\left(\frac{n}{\varepsilon}\right)$$

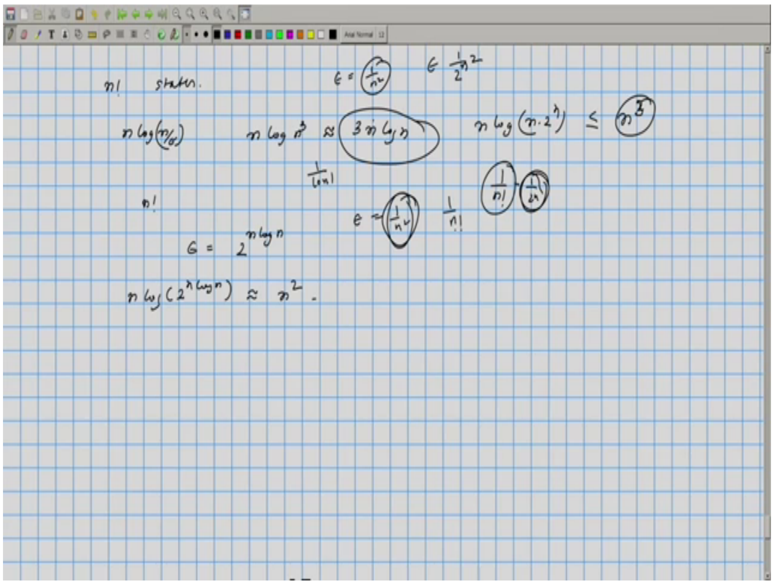
So, let us look at some particular card. Let us say probability that card i is not picked in n log n plus c n steps ok. That would mean that it should not have been picked in the first step, not in the second, not in the third and so on. So, 1 minus 1 minus is the probability of not being picked in the first step and then, it is not picked for n log n steps that would be the product. So, 1 minus 1 by n whole raised to n log n plus c n ok. So, this is going to be 1 minus 1 by n the whole raised to n times log n plus c ok. So, that is going to be less than 1 minus 1 by n raised to n that is e raised to minus 1 ok. So, this is e raised to log n into e raised to minus c. So, that is equal to e raised to minus c by e raised to log n is just one by e raised to minus log n is by n ok.

So, a card i is not picked with probability e raised to minus c. So, probability that there exists a card which was not picked ok. So, this probability is going to be less than or equal to we can apply union bound there are n possibilities. So, that is going to be e raised to minus c ok. So, if we choose c to be say log 1 by epsilon and probability that there exists a card which is not picked or we will call that is a probability of the bad event is going to be less than e raised to log 1 by epsilon and negative of this.

So, that is going to be epsilon ok. So, probability of bad event is going to be epsilon. So we can just write this as if T is equal to n log n plus n log 1 by epsilon that is equal to n log n by epsilon, then the probability that X T not equal to Y T given a starting configuration that is going to be less than epsilon.

Now, once we have this we know that the mixing time is going to be less than n log n by epsilon ok. So, in n log n by epsilon time, the Markov chain that we had constructed will reach stationary very close to stationary distribution. In the sense the variation distance between the distribution that has been reached and the stationary distribution is going to be less than epsilon ok.

(Refer Slide Time: 35:21)



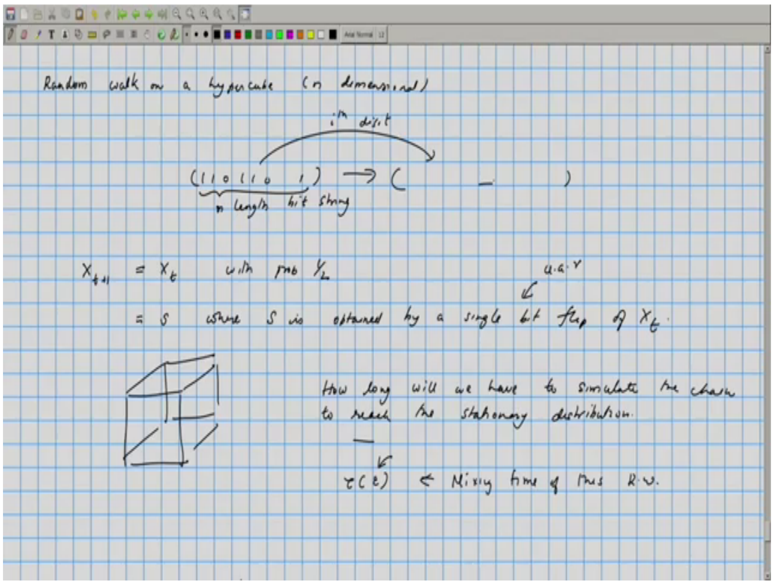
So, look at the Markov chain, it had n factorial states ok. Although it had n factorial states, in n log n by epsilon moves, the number of the fraction of states that would have been explored is very small that is a tiny fraction and if let us say epsilon is 1 by n square ok. The number of states that has been explored is n log n cube which is approximately 3 log 3 n log n ok. So, although you have only stimulated this the Markov chain for 3 n log n steps, the Markov chain is now equally likely to be in any one of the n factorial states ok; almost equally likely.

So, since there are n factorial state, since the Markov chain has an the one which we constructed is a periodic and if it is all nice properties. We can argue that its study state distribution is going to be the uniform distribution ok. We could take a epsilon to be much bigger than this, we could take epsilon to be let us say 1 by 2 raised to n; even then we are running it for only if epsilon is a 1 by 2 raised to n, then n log n times two raised to n is going to be less than n square or let us say certain less than n cube. So, even if you run the Markov chain for n cubes steps, we are going to get some distribution on the states which is very close to the uniform distribution.

So, 1 by n factorial minus 1 by 2 raised to n or something like that ok. So, you could have taken this to be larger than 1 by 2 raised to n is larger quantity than 1 by n factorial, but we could have taken this to be say n square and even then, we do not I mean this not much harm done in the calculations ok. So, if epsilon is 1 by 10 n factorial which is 2 raised to n log n ok, then n log 2 raised to n log n is approximately n square ok.

So, if you take this to be n cube, then we know that we get very close to the uniform distribution ok. If you taken n to b epsilon to be 1 by n square that is not good enough because this probability is significantly higher than 1 by n factorial ok.

(Refer Slide Time: 38:17)



Random walk on a hypercube (n dimensional)

i^{th} digit

$(\underbrace{1\,1\,0\,1\,1\,0\quad 1}_{n \text{ length bit string}}) \longrightarrow (\qquad - \qquad)$

$X_{t+1} = X_t$ with prob $1/2$

u.a.r

$= S$ where S is obtained by a single bit flip of X_t.

How long will we have to simulate the chain to reach the stationary distribution.

$\tau(\varepsilon) \leftarrow$ Mixing time of this R.W.

Now, we will see at another application of this coupling method, wherein we will look at random walk on a hyper cube ok. So, hyper cube we had seen before its a graph ok. Let us say n dimensional hyper cube ok. It is a graph, where every bits string of length n is present ok. So, these are the vertices. Every n length bit string is a vertex and the neighbors are all those vertices which can be obtained from this by flipping 1 digit.

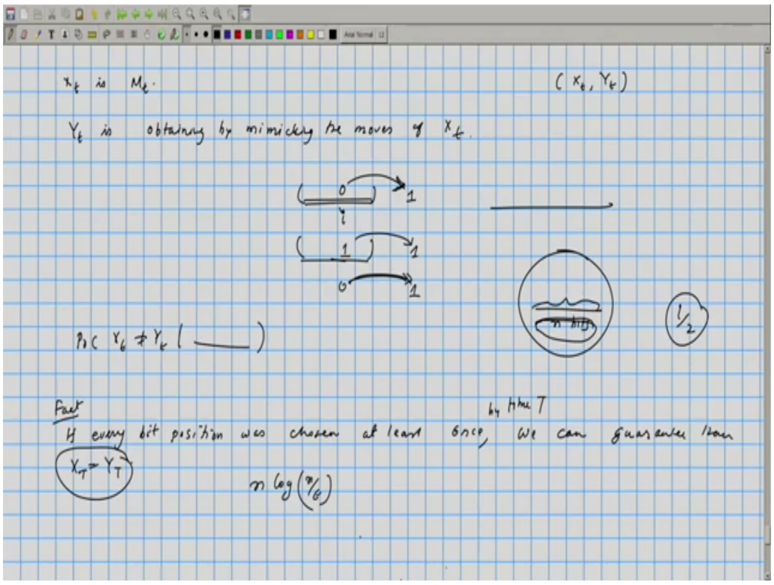
So, if you flip the ith digit whatever you get that is going to be one of the neighbors and you can flip any of the n digits ok. Now, in this random walk we will assume that we toss a coin and if the coin is heads, we stay there itself; otherwise we will transition by choosing one random position to flip ok.

So, X t plus 1 is equal to X t with probability half and this is equal to say S, where S is obtained by a single bit flip of x t. So, take X t and flip one digit, one bit and that bit is chosen uniformly at random. So, it is a ran position all of those things are happening with probability 1 by 2 n ok. So, this is the random walk on the hyper cube that we are interested in. Now what we want to know is how long will it be before we will get uniform distribution or all most uniform distribution on the hyper cube.

So, how long will we have to stimulate the chain to reach the study state? You can convenes yourself that there is only one stationary distribution for that this and it is the uniform distribution ok.

So, in order to answer this question, we are interested in is we want to say if it is to be away from the uniform distribution by an amount epsilon. Then, how much time we will it take? In other words, we want to find out the mixing time of this random walk, we can do very similar thing as the previous example.

(Refer Slide Time: 41:29)



x_t is M_t.

(X_t, Y_t)

Y_t is obtaining by mimicking the moves of X_t.

$Pr(Y_t \neq Y_t(\_\_\_\_))$

Fact

If every bit position was chosen at least once, by time T we can guarantee that $X_T = Y_T$

$n \log(n/\epsilon)$

So, can X t is going to be the same random walk. So, X t plus 1 or the random walk X t is M t and Y t is obtained by mimicking the moves of X t ok. Mimicking the moves means if X t had changed the ith bit and Y t will also change the ith bit ok. So, we have to be careful. If the ith position, if you look at a state that consist of a bit string, if the ith position was flipped from 0 to 1 in a X t and if Y t was in a state, where this is already 1; then you just do nothing. If it was 0, then you change it to 1.

So, when you mimic, it essentially means that after the move both X t and Y t will agree on the position, where the change has been made. Note that if you look at these two Markov's chain jointly, then you can see that the marginals will exactly be the empty itself. Because there was some probability of this bit flipping, if it was flipping then it can flip to can either flip or stay with some probability and the same thing happens in Y t as well ok. So, X t and Y t has the tendency to couple and once they have coupled it remains the marginal's anyways are same as empty once they have coupled, they remain in that particular state.

So, what we want to know is probability that X t not equal to Y t given some initial condition. So, whatever is the initial condition, if we have identified all the even if you have chosen at random, all the positions in the state space that is I mean you look at any state each state will consist of n bits; if each of these bits were chosen even once, then the chains would have coalesced.

So, we will just write this is a fact. If every bit position was chosen at least once, we can guarantee that X T would be equal to Y T ok. So, that is if every bit position was chosen at least once by time T, then we can guarantee that X T equals Y T and from there onwards they will remains ok.

So, now, how do we upper bound the time taken for X T equals Y T. Again, it is a same as a previous analysis. Each of these positions had equal probability of being sampled. There is a factor half which is with the probability half you remain in the same stage itself, but if you ignore that part and if you just imagine this to be a chain which is running two times slower than the earlier one.

We can say that this is similar to coupon collector problem. How much time we will take for you to collect n coupons and that is again going to be something like n log n and if you run the chain for n log n by e steps and by epsilon steps, you can argue that all positions would have been chosen at least once ok.

So, this concludes the week on Markov's chains and the mixing properties of Markov's chains. We will again see the mixing properties of Markov chains, when we look at approximate counting.