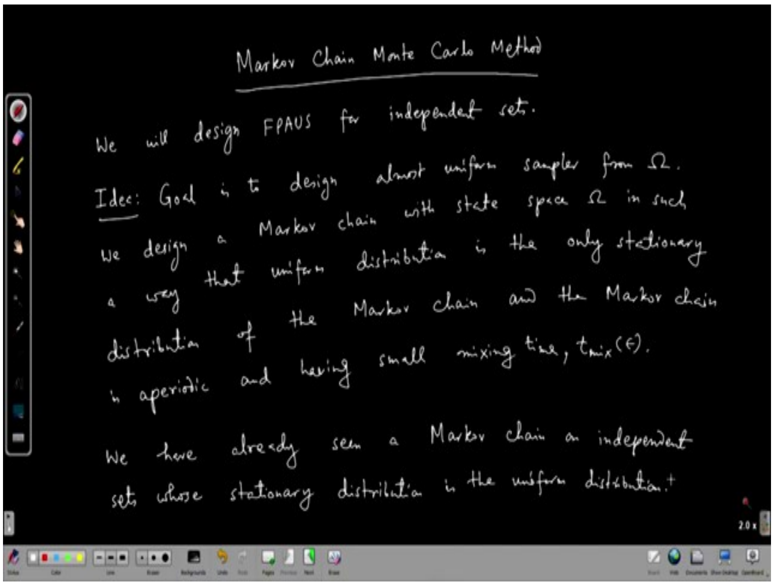
Selected Topics in Algorithm
Prof. Palash Dey
Department of Computer Science and Engineering
Indian Institute of Technology – Kharagpur

Lecture – 29
Counting Independent Sets of a Graph

Welcome, in the last place, we have started Markov Chain Monte Carlo method and we will see the demonstration of this method for designing uniform samplers or almost uniform samplers for independent sets.

(Refer Slide Time: 00:41)

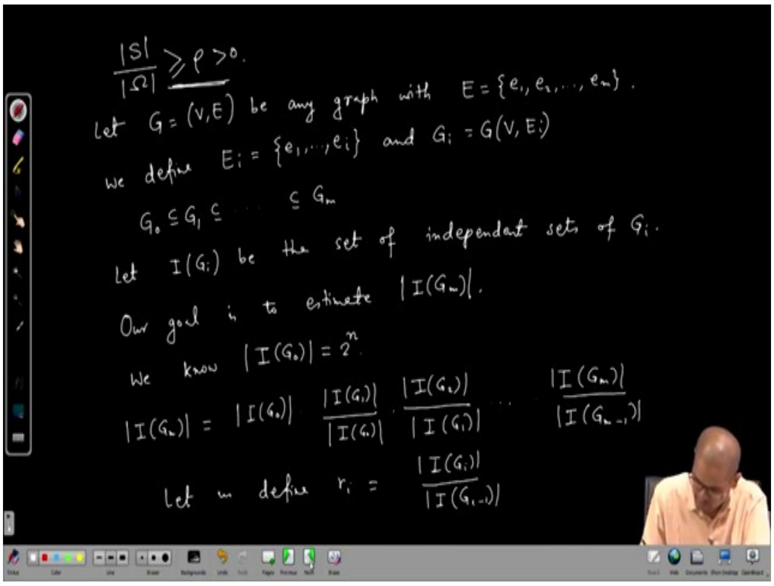


So, Markov Chain Monte Carlo and using that we will design and FPAUS, fully polynomial approximate uniform sampler for independent sets. But before that let us discuss where Markov Chain comes into picture. So, the idea so, the goal is to design a uniform sampler from U so, goal is to design almost uniform sampler from some underline set omega. So, for that we design a Markov Chain with state space omega in such a way that uniform distribution is the only stationary distribution of the Markov Chain.

And the Markov Chain is a periodic and having small mixing time set t mix epsilon. So, if this is the case then we run the Markov Chain for t mix epsilon time and the state, that x t will be distributed uniformly close to uniformly it is epsilon closed in total variation distance to uniform distribution across all state space. So, the algorithm or the sampler is simply run the Markov Chain for $t \geq t_{mix}(\epsilon)$ time and output the t-th step.

So, for independent sets we have already seen and Markov Chain on independent set, whose stationary distribution is the uniform distribution over independent sets. So, let us recall that so, we have already seen a Markov Chain on independent set whose stationary distribution is the uniform distribution.

(Refer Slide Time: 06:46)



So, if we run that Markov Chain for t step, the running time is dominated by t. So, we have a sampler we sample we uniform samples but again, so, using this Markov Chain assuming t mix is small enough. So, with t mix time we can draw an independent set uniformly at random and again apply the Monte Carlo method but what if or how do we show? That the number of independent sets is not too small compared to the sample space.

To apply estimator theorem recall that we suppose we are estimating size of S and we are sampling from omega. Then this should be bounded away from 0 and hence for Monte Carlo method to be successful this favourable outcome should be considerably large than the sample space. So, you cannot just simply pick arbitrary sub set of vertices and check whether it is independent set or not.

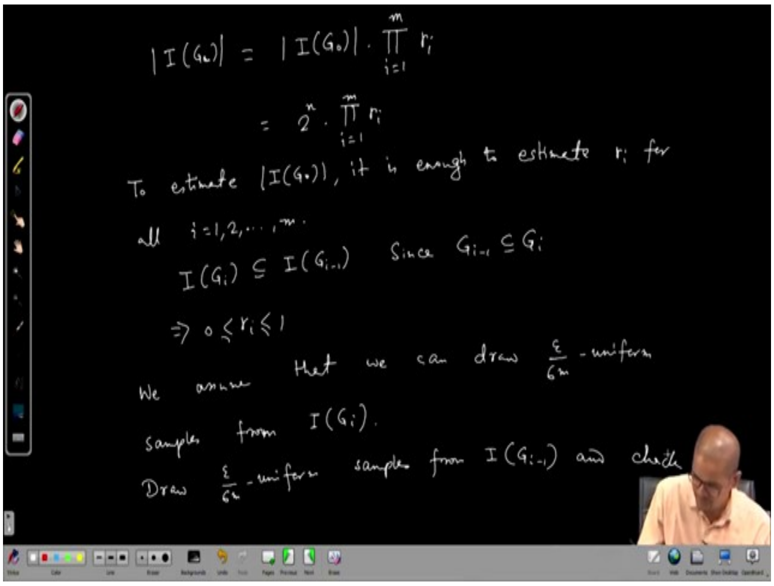
So, to tackle that what we do is this approach. So, let so, in DNF counting also we face this problem. And what we did is that instead of sampling any arbitrary assignment of Boolean variables, we sampled from a suitable sample space suitable omega which contains the solution. The same approach is followed here but here the idea is little more non trivial. So, let us see that so, let $G=(V,E)$ be any graph with each set $E=\{e_1,e_2,\ldots,e_m\}$.

What we do is that we first we define E_i to be the first i edges $\{e_1, \ldots, e_i\}$. First, in with respect to this numbering and G_i is the induced graph V and with this edges E_i. So, the all vertices remain but only edges E_i is there so, for using this. We construct this sequence of graphs $G_0 \subset G_1 \subset \ldots \subset G_m$ simply because $E_0 \subset E_1 \subset \ldots \subset E_m$.

All graphs have the same set of vertices but the edges are getting added one by one. Let $I(G_i)$ be the set of independent sets of G_i. And our goal is to estimate cardinality $I(G_m)$. And what we know? We know $I(G_0)$, in G_0 there is no edges. So, all subset of vertices is an independent set. This is equal to 2^n. Now, we write cardinality $I(G_m)$ as cardinality $I(G_0)$ times cardinality $I(G_1)$ by cardinality $I(G_0)$ times cardinality $I(G_2)$ by cardinality $I(G_1)$ cardinality $I(G_m)$ by cardinality $I(G_{m-1})$.

And let us call these numbers, let us define r_i to be cardinality $I(G_i)$ by cardinality $I(G_{i-1})$.

(Refer Slide Time: 13:08)

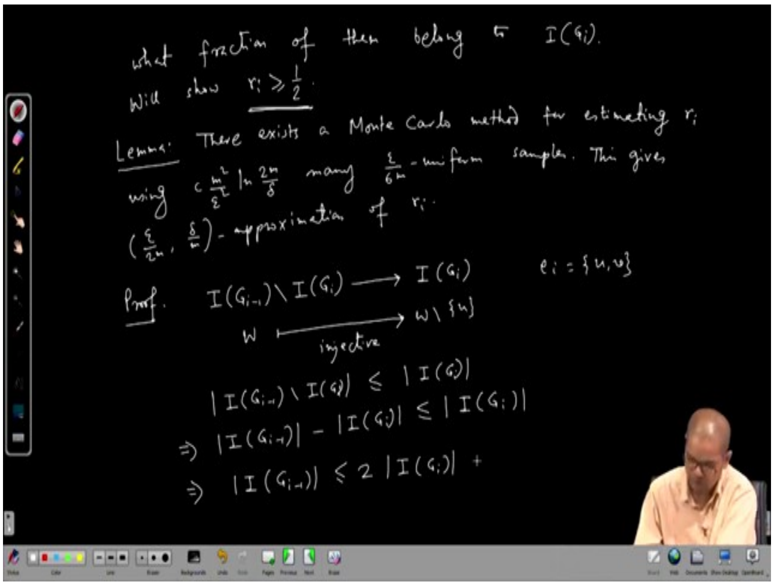


So, cardinality $I(G_m)$ is cardinality $I(G_0)$ times $\prod_{i=1}^{m} r_i$ and cardinality $I(G_0)$ is $2^n \prod_{i=1}^{m} r_i$. So, to estimate $I(G_m)$ cardinality $I(G_m)$ it is enough to estimate r_i for all i = 1, 2 to m. Also observe that each $I(G_i)$ is a subset of $I(G_{i-1})$. This is because G_{i-1} is a sub graph of G_i. So, this implies that each r_i is in between 0 and 1 these are fraction.

So, to approximately estimate cardinality $I(G_m)$ because if we can exactly compute what is cardinality $I(G_m)$. Then we can solve the independent set problem which is quite unlikely to be solved in polynomial time. So, what we do is that if we do an approximate counter, so, we assume that we can draw approximately uniform samples epsilon by $6m$ uniform samples from $I(G_i)$.

So, the idea what is the idea? The idea is to draw epsilon by $6m$ uniform samples from $I(G_{i-1})$.

(Refer Slide Time: 17:30)



And check what fraction of them belong to $I(G_i)$. So, draw any almost uniform sample or approximately uniform sample from approximately uniform independent set of G_{i-1}. And check whether it is an independent set of G_i or not now and this approach will work because will show that we will show each r_i is bounded away. That is again important, so, we will show this. So, for that now let write this formally in a Lemma.

There exist so, suppose, assuming or almost uniform sample or approximately uniform samples from $I(G_i)$ for each i. There exists Markov Chain Monte Carlo method or using Markov Chain will do the sample sampling and once we have samples there exist a Monte Carlo method for estimating r_i using $\frac{cm^2}{\epsilon^2}\ln\left(2\frac{m}{\delta}\right)$ many $\frac{\epsilon}{6m}$ uniform samples.

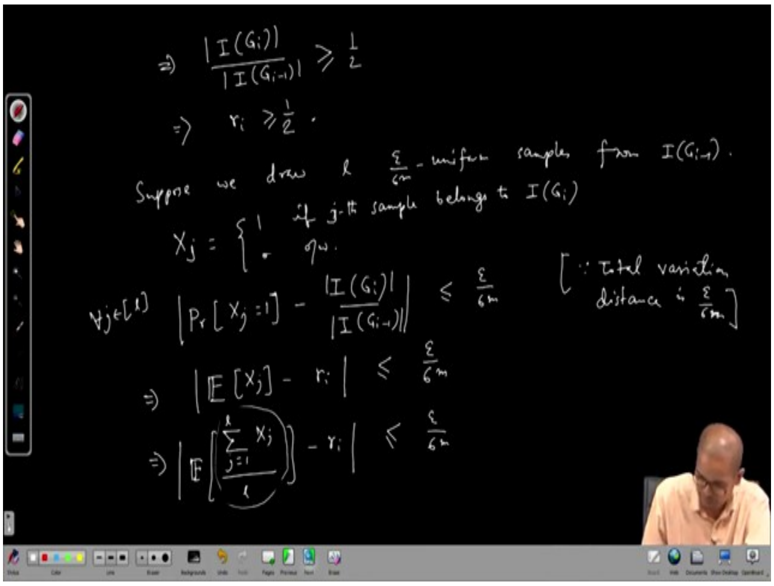
So, there is no, there exists a Monte Carlo method for estimating r_i using this many uniform samples using and this gives $(\frac{\epsilon}{6m}, \frac{\delta}{m})$ approximation of r_i. Proof, so, again we will first begin with showing r_i is greater than equal to half. So, for that we observe that look at this set

$I(G_{i-1}) \backslash I(G_i)$. So, this is the set of all independent sets of G_{i-1} which is not an independent set of G_i so, this to $I(G_i)$.

So, there is a difference on only this edge, so, suppose e_i is $\{u, v\}$. So, if i take an independent set of G_{i-1} which is not an independent set in G_i then both u and v must have been picked. So, an independent set W we can map it to an independent set $W \backslash \{u\}$ because this sets this picks both u and v. So, this is an injective map. Injective mapping or one to one mapping that means cardinality $I(G_{i-1}) \backslash I(G_i)$ is less than equal to cardinality $I(G_i)$.

But because $I(G_{i-1})$ is a subset of $I(G_i)$ we have this is cardinality $I(G_{i-1})$ – cardinality $I(G_i)$ less than equal to cardinality $I(G_i)$. This is cardinality $I(G_{i-1})$ is less than equal to twice cardinality $I(G_i)$.

(Refer Slide Time: 23:24)



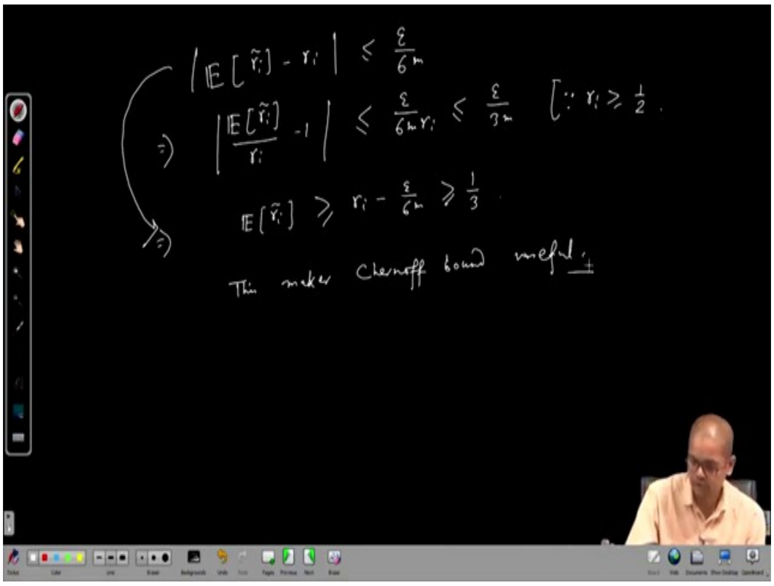
That this implies that cardinality $I(G_i)$ by cardinality $I(G_{i-1})$ this is greater than equal to half but left hand side is nothing but r_i means r_i is greater than equal to half. So, next we show that if we can estimate r_i within or almost uniformly from $I(G_i)$ then we can sample or we can estimate $I(G_i)$ we can estimate r_i. So, suppose we draw one $\frac{\epsilon}{6m}$ uniform samples from $I(G_{i-1})$. And again, what is the Monte Carlo method?

Again, simple see what fraction of them belong to $I(G_i)$ and that is my estimate. That is r if k of them belongs to ij then k by l is my r_i. So, let us define for that X_i is 1 if ith sample belongs to i or let me write j here is jth sample belongs to $I(G_i)$ and 0 otherwise. So, what we have is because it is an almost uniform sample probability that $X_j=1$. This is the probability that a jth sample belongs to $I(G_i)$.

And jth sample is an $\frac{\epsilon}{6m}$ uniform from $I(G_{i-1})$. So, this – $I(G_i)$ by $I(G_{i-1})$ this is less than equal to $\frac{\epsilon}{6m}$. Because since total variation distance is $\frac{\epsilon}{6m}$ that means that is why the probability of any event differs from the actual probability by $\frac{\epsilon}{6m}$. But this implies that probability of $X_j=1$ is nothing but expectation of X_j – its cardinalities and this is r_i this is less than equal to $\frac{\epsilon}{6m}$.

But this is nothing but summation j = 1 to l you see this holds true for all j this is for all j in 1 to l. This is $\frac{X_j}{l}-r_i$, this is $\frac{\epsilon}{6m}$. And this is our estimate this is r_i tilde. This is the output of the algorithm,

(Refer Slide Time: 28:14)



So, expectation of $\widetilde{r_i}-r_i$ mod this is less than equal to $\frac{\epsilon}{6m}$. So that means that expectation of $\frac{\widetilde{r_i}}{r_i}-1$ this is less than equal to $\frac{\epsilon}{6m}r_i$ but this is less than equal to $\frac{\epsilon}{3m}$. Since r_i is greater than equal to half. So, this implies that expectation of $\widetilde{r_i}$ is greater than equal to $r_i-\frac{\epsilon}{6m}$. This follows from here that if expectation of $\widetilde{r_i}$ deviates from r_i in absolute value within $\frac{\epsilon}{6m}$.

That means expectation of $\widetilde{r_i}$ is at least $r_i - \frac{\epsilon}{6m}$ which is for large enough m this is greater than equal to one third. So, for Chernoff Bound to be applicable to be useful expectation also should be bounded away from from 0. So, this makes hence this makes Chernoff bound useful. So, next, in the next class we will use. This will continue this proof and we will see how using this we can design an approximate counter for counting the independent sets of a graph. Thank you.