Lecture 9:

- Markov Chain Monte Carlo
  - Sampling

    $\Rightarrow$ Counting
- Unbiased Estimators and Approximation Schemes.
  - Counting Colorings via Samplers
  - Sampling Colorings

Consider following problem (Counting DNF Assignments)

- Given a Boolean formula in $\boxed{\text{Disjunctive Normal Form}}$, how many satisfying assignments does it have?

$$DNF = (x_1 \wedge x_3 \wedge \overline{x_7}) \vee (x_2 \vee \overline{x_9} \vee x_{13}) \vee \cdots$$

↑ OR of Ands.

Note: if CNF, then this problem (counting # of sat Assignments) at least as hard as answering if $\# = 0$ or > 0.

$\Rightarrow$ at least as hard as SATisfiability.

But DNF is easy: if any clause is satisfiable $\Rightarrow$ φ is satisfiable.

- Here's a simple algorithm:—
  - Pick a random assignment in $\{T, F\}^n$

    Check if its is satisfying.

  Repeat this N times.

  Output $\hat{X} = \frac{1}{N}$ (# of sampled assignments that satisfied φ.)

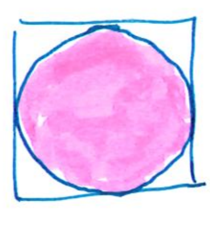
---

the "Monte Carlo" paradigm!

E.g. Estimate value of π.

- Pick a random $(x, y) \in [-1, 1]$ check if $x^2 + y^2 \leq 1$.]

$Pr(\text{point in circle}) = \frac{\pi}{4}$



- repeat N. times.

How many samples do we need?

Well: suppose we want

$$\Pr(|\text{Estimator} - \text{truth}| \le \varepsilon.\text{truth}) \ge 1-\delta$$

ε multiplicative error

whp.

Say correct estimate $= p. = \frac{|S|}{|U|}$.

$X_i = \mathbb{1}(\text{experiment } i \text{ succeeded}) \Rightarrow \mathbb{E}(X_i) = p.$

$Z = \sum_{i=1}^{N} X_i \Rightarrow \mathbb{E}[X] = Np.$

$\text{Estimator} = \frac{1}{N} Z = X$

$$\Pr(|\text{Est} - p| \ge \varepsilon p) = \Pr(|Z - pN| \ge \varepsilon p N)$$

$$\le 2.\exp\left(-\frac{(\varepsilon p N)^2}{2pN + \varepsilon p N}\right)$$

$$\le 2\exp\left(-\frac{\varepsilon^2 p N}{3}\right)$$

$$\le \delta$$

if $N \ge \frac{3}{p.\varepsilon^2} \log \frac{2}{\delta}$

Great! if $\#\text{samples} \ge \Omega\left(\frac{1}{p\varepsilon^2} \log \frac{2}{\delta}\right)$ we're set!

Great for estimating π: there $p = \pi/4 = \Theta(1)$.

Maybe not for DNF: $\#$sat assignments may be $\ll 2^n$.

$\Rightarrow p$ may be tiny! ☹

Needle in a Haystack ☹

Good feature:

can design our own haystack!

"Smart Monte Carlo"

"haystack"

Design a good universe $\mathcal{U}$ s.t.

(1) know the size of $\mathcal{U}$.

(2) can sample from $\mathcal{U}$ uniformly at random

(3) $S \subseteq \mathcal{U}$ and $\frac{|S|}{|\mathcal{U}|}$ is "large". (and can check if $x \in S$, naturally)

set we care about aka "needles"

General Algo:

- Sample from $\mathcal{U}$ to get x.
- output $|\mathcal{U}| \cdot \mathbb{1}(x \in S) = Y$

"unbiased estimator"

$\mathbb{E}[Y]$

$= |\mathcal{U}| \cdot \Pr(x \in S)$

$= |\mathcal{U}| \cdot \frac{|S|}{|\mathcal{U}|} = |S|.$

uniform sample

Generally: high variance

So sample $X_1 X_2 \dots X_N$ some N times

and output $X = \frac{1}{N} \sum_i X_i$

How to do this for DNF?

Say define A_i = assignments that satisfy clause i

$$|A_i| = 2^{n - \text{length}(\text{clause}_i)}.$$

Define $\mathcal{U} = \{(i,a) \mid a \in A_i,\ i \in [m]\}$. be disjoint union

so $|\mathcal{U}| = \sum_i |A_i|$

☹ An assignment a may appear multiple times as $(1,a)$, $(5,a)$ ....
if a satisfies clauses 1, 5...

So define canonical copy:

define $S' = \{(i,a) \mid a \in A_i$ but $a \notin A_j\ \forall j < i$, $i \in [m]\}$

Each satisfying assignment appears once now in S'.

$|S'| = \#$ A sat assignments.

Moreover $|S'| \geq |\mathcal{U}|/m$.

— x —

Good: we know $|\mathcal{U}|$.
know $|S'|/|\mathcal{U}|$ "large" and $S' \subseteq \mathcal{U}$.

How to sample from $\mathcal{U}$ uniformly?

$\boxed{\text{"Importance" Sampling:}}$

$$\begin{cases} \text{Pick } i \in [m] \text{ w.p. } \dfrac{|A_i|}{|\mathcal{U}|}. \\ \text{Pick } a \text{ uniformly at random from } A_i \end{cases}$$

$\Rightarrow$ all set. Get estimator for DNF counting!

Since prob of success $= p \geq 1/m$

Previous calculation shows that $\Omega\left(\frac{1}{p\varepsilon^2}\log\frac{1}{\delta}\right) = \left(\frac{m}{\varepsilon^2}\log\frac{1}{\delta}\right)$ samples suffice.

for $1 \pm \varepsilon$ estimate wp $1-\delta$.

—— x ——

In HW see another estimator, similar in spirit.

—— x ——

Btw! an (ε, δ) FPRAS "fully polynomial randomized approximation scheme"

for estimation is algo that outputs X s.t

$$\Pr(|X - p| \geq \varepsilon p) \leq \delta$$

And runtime is poly(instance size, $1/\varepsilon$, $\log 1/\delta$).

—— x ——

We just gave an (ε, δ) – FPRAS for DNF Counting. with $O\left(\frac{m}{\varepsilon^2}\log\frac{1}{\delta}\right)$ samples.

—— x ——

BTW: sps have $(\varepsilon, 1/4)$ – FPRAS for some problem. with N samples

$\Rightarrow$ Can get (ε, δ) – FPRAS with $O(N \log 1/\delta)$ samples

How?

Use median-of-means idea from HW2.

take $O(\log 1/\delta)$ samples from weak FPRAS

output the median.

We saw how:

sampling $\Rightarrow$ estimation of sizes (aka "counting").

<u>Let's see another use of this:</u> (non-trivial)

Suppose given a graph G, ~~we~~ and a number K, we can sample a coloring from all proper K-colorings of G uniformly at random.

<u>proper k-coloring</u>: is a map $\chi : V \to [K]$ such that $\chi(u) \neq \chi(v)$ $\forall$ edges $(u,v) \in E(G)$.

press button $\to$ $\boxed{A}$ $\to$ output $= \chi$

u.a.r from all proper K-colorings of G.

How can we use this sampler to <u>count</u> the # of K-colorings of G?

<u>Cute Idea:</u>

Let $\Omega(G)$ be # of K-colorings of G.

then suppose $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_m =$ empty graph

where $G_i = G_{i-1}$ minus one edge.

then $|\Omega(G)| = |\Omega(G_0)| = \dfrac{|\Omega(G_0)|}{|\Omega(G_1)|} \cdot \dfrac{|\Omega(G_1)|}{|\Omega(G_2)|} \cdots |\Omega(G_m)|$.

but

(a) $|\Omega(G_m)| = K^n$ because empty graph $\Rightarrow$ all colorings are proper.

Suppose could estimate

$\dfrac{|\Omega(G_i)|}{|\Omega(G_{i+1})|}$ up to $(1 \pm \varepsilon/m)$, then get $|\Omega(G)|$ up to $(1 \pm \varepsilon/m)^m$

$\cong (1 \pm \varepsilon)$.

How to compute $\dfrac{|\Omega(G_i)|}{|\Omega(G_{i+1})|}$ (at least in some nice cases)

<u>Note</u>: $\Omega(G_i) \subseteq \Omega(G_{i+1})$. every proper coloring in G_i (more edges) is also proper in G_{i+1}

take any $\chi \in \Omega(G_{i+1}) \setminus \Omega(G_i)$. say $G_i = G_{i+1} + \{u,v\}$.

then must have $\chi(u) = \chi(v)$.

only problem that can arise.

$\boxed{\text{Suppose that } k \geq \Delta+1}$

max degree of G.

then define χ' new coloring for G_i

by choosing some free color for v and using $\chi'(v)$ = that color

$\chi'(w) = \chi(w)$ for all $w \neq v$.

<u>Claim</u>: can recover from χ' to χ.

How? just set $\chi(v) = \chi'(u)$ and $\chi(w) = \chi'(w)$.

So is 1-1 map from $\Omega(G_{i+1}) \setminus \Omega(G_i)$ to $\Omega(G_i)$ if $k \geq \Delta+1$.

$\Rightarrow$ <u>Claim</u>: $p_i := \dfrac{|\Omega(G_i)|}{|\Omega(G_{i+1})|} \geq 1/2$ if $k \geq \Delta+1$.

x

So sample colorings from $\Omega(G_{i+1})$ using the claimed sampling algo.

Check if it is valid for G_i. $\Rightarrow$ get unbiased estimator.

$\Rightarrow$ get $(1 \pm \varepsilon/m)$-apx for $\dfrac{|\Omega(G_i)|}{|\Omega(G_{i+1})|}$ wp $(1-\delta/m)$

using $O(\frac{m^2}{\varepsilon^2} \log \delta/m)$ samples. pair

$\boxed{\text{Chain these together to get } (1\pm\varepsilon)\text{-apx for } |\Omega(G)| \text{ wp } 1-\delta.}$

Note: Assumed that

given a graph H, we can sample from its K-colorings uniformly.

$\Rightarrow$ used to give

approximate counting algo ("how many k-colorings does graph G have"?")

$\boxed{\text{"reduce counting to sampling".}}$

Similar idea works for many "self reducible" problems. (See texts).

——— x ———

In fact can also get samples from counters

- see HW for examples.

——— x ———

- Is OK to have apx-uniform sampling too

→ i.e. sps can pick a random coloring from $\Omega(H)$ w.p.

$$(1 \pm \eta) . \Omega(H)$$

then can fold the error into the analysis if η is suitably small.

——— x ———

How can we get apx-uniform samplers?

Common technique is the Markov Chain Monte Carlo method

... after the break ....

INTERMISSION.

Markov Chain Monte Carlo & Mixing Times

Basic Idea is quite simple.

Suppose want to sample from some set $\mathcal{U}$ (uniformly, say).

I.e. want algo that outputs $x \in \mathcal{U}$ w.p. $1/|\mathcal{U}|$.

① Set up a Markov chain whose states are elements of $\mathcal{U}$.

<u>And</u> stationary distribution is uniform. $(\pi^*_x = 1/|\mathcal{U}| \ \forall x \in \mathcal{U})$

② Make sure MC is aperiodic and irreducible

So that $\pi^t \to \pi^*$ as $t \to \infty$

③ Simulate MC for some N steps.

~~"Hope" that X^N is "close enough" to uniform.~~

Prove that X^N is close to uniform.

i.e. $\Pr(X^N = x) \in \frac{(1 \pm \eta)}{|\mathcal{U}|} \quad \forall x \in \mathcal{U}$.

MC has "mixed" by N steps

How do we do this?

Let's do it by an example, and introduce necessary ideas along the way.

<u>Example:</u> K-colorings of graphs when $\underbrace{K \geq 2\Delta + 1}$

"lots of breathing room" allows mixing to happen.

A Markov Chain on Colorings (when $K \geq \Delta + 2$)

max degree

Given a coloring $\chi : V \to [K]$

Pick a vertex $v \in V$ unif at random

Pick a color $c \in [K]$ u.a.r.

if $\chi'(w) = \begin{cases} \chi(w) & \forall\, w \neq v \\ c & \text{if } w = v \end{cases}$ is a proper coloring,

move from $\chi \to \chi'$

else stay at χ.

Called the "Glauber dynamics" on colorings.

Lemma 1: there is a non-zero prob that $\chi \to \chi$.

(indeed if we choose color $c = \chi(v)$)

$\Rightarrow$ chain is aperiodic.

Lemma 2: Can get from any coloring χ to any other coloring χ^* using these moves. (if $K \geq \Delta + 2$)

Pf: Iteratively recolor vertices using colors of χ^*.

Sps vtx i is smallest vertex such that $\chi(v) \neq \chi^*(v)$.

i

Good $\{w : \chi(w) = \chi^*(w)\}$

Bad $\{w : \chi(w) = \chi^*(i)\}$

Rest $\{w : \chi(w) \neq \chi^*(i)$ $\neq \chi^*(w)\}$

Since $K \geq \Delta + 2$

$\forall w \in$ Bad set $\exists$ color not equal to $\chi^*(i)$ (1 such color)

not in its nbr hood ($\leq \Delta$ such colors)

"free color". So recolor w using such a color.

Do this until no bad neighbors of i

Now: color i with color $\chi^*(i)$

So finally get to χ^* using these vertex recoloring moves ▨

—— X ——

$\Rightarrow$ So the markov chain is (a) irreducible (strongly connected)

(b) aperiodic

So if we just keep walking randomly, then $\pi^{(t)} \longrightarrow \pi^*$ as $t \to \infty$.

But want quantitative bounds!

—— X ——

Oh, btw: what is the stationary distribution π^*?

Forgot to argue that it is Uniform.

Let's check: -

For any χ', it has a non zero prob to get to χ in one step only if $\chi' = \chi$

or if $\exists$ unique vertex v s.t. $\chi'(v) \neq \chi(v)$.

Basically: $P_{\chi',\chi} = \frac{1}{n.K}$ if $\exists v \in V$ $\exists c \in C$ $\neq \chi(v)$ $\chi'(v) = c$, $\chi'(w) = \chi(w)$ $\forall w \neq v$

$P_{\chi,\chi} = \frac{1}{K}$.

Now: Check that $\pi^* = \frac{1}{|\mathcal{U}|}$ $\qquad$ $\mathcal{U}$ = set of all ~~any~~ colorings

satisfies $\pi^*_x = \sum_{x'} \pi^*_{x'} \cdot P_{x'x}$

$$= \frac{1}{|\mathcal{U}|} \cdot \frac{1}{k} + n \cdot (k-1) \frac{1}{|\mathcal{U}|} \cdot \frac{1}{nk}$$

$$= \left(\frac{1}{k} + \frac{k-1}{k} \right) \frac{1}{|\mathcal{U}|} = \frac{1}{|\mathcal{U}|}.$$

☺

So stationary distrib is unif over all colorings!

——— x ———

Now the fun part. How to show mixing? "rapid mixing".

Thm: if #colors $\geq 4\Delta + 1$ $\Rightarrow$ very close to uniform in $O(n \lg n)$ moves!

how to measure?

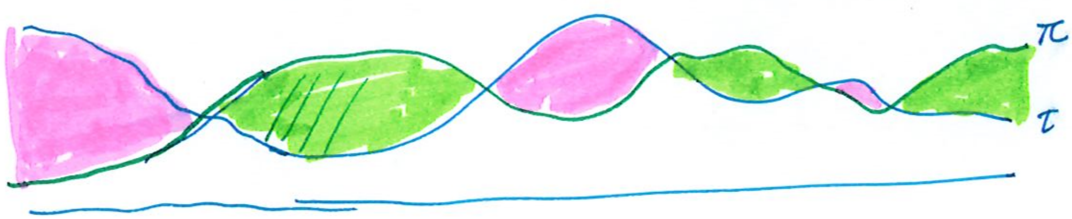
——— x ———

Given a space Ω and two probability distributions over Ω
π and τ

the total variation distance

$$\|\pi - \tau\|_{TV} = d_{TV}(\pi, \tau) := \frac{1}{2} \sum_{\omega \in \Omega} |\pi_\omega - \tau_\omega| \qquad = \frac{1}{2}(\text{red} + \text{green})$$

$$= \sum_{\omega \in \Omega} (\pi_\omega - \tau_\omega)^+ \qquad = \text{green} \ (= \text{red})$$



Fact: if d_{TV} dist $\leq$ epsilon $\Rightarrow$ for any event $\mathcal{E} \subseteq \Omega$,
$|Pr_\pi(\mathcal{E}) - Pr_\tau(\mathcal{E})| \leq$ epsilon.

Mixing Time (in TV-distance)

The ε-mixing time of a Markov chain is at most $\tau = \tau(\varepsilon)$ if starting from any starting distribution $\pi^{(0)}$, the distribution $\|\pi^{(\tau)} - \pi^*\|_{TV} \le \varepsilon$.

N.b. can also define ℓ_2-mixing time

or other notions (KL divergence ....)

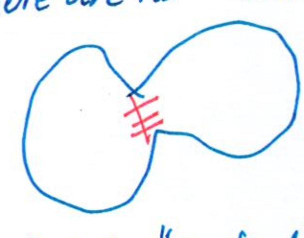
But this is a clean measure.

—x—

OK: so we want to see how long to run MC so that $\|\pi^{(t)} - \pi^*\|_{TV} \le \varepsilon$?

Several techniques

① Show that there are no "bottlenecks" in the MC so that starting anywhere gets us everywhere.



related to "expansion"

(a) Show that the "spectral* gap" is large (see later lecture)

(b) Show that $\exists$ a good flow (dual idea)

② [Today] A coupling based idea.

A coupling of a Markov Chain M_t with state space Ω is another markov chain $Z_t = (X_t, Y_t)$ on state space $\Omega \times \Omega$ s.t.

① $Pr(X_{t+1} = x' \mid Z_t = (x,y)) = Pr(M_{t+1} = x' \mid M_t = x)$

② $Pr(Y_{t+1} = y' \mid Z_t = (x,y)) = Pr(M_{t+1} = y' \mid M_t = x)$.

Both coordinates are faithful copies of the orig. MC.

but they could be arbitrarily correlated.

—— x ——

Idea: design coupling s.t.

(a) X_{t+1}, Y_{t+1} come "closer" to eachother than X_t, Y_t were

(b) $X_{t+1} = Y_{t+1}$ if $X_t = Y_t$.

Why?

Thm: (Coupling Thm) Sps $Z_t = (X_t, Y_t)$ is a coupling for MC. M_t on Ω.

Sps $\exists$ time $T = T_\varepsilon$ s.t. "coupling time".

$$Pr(X_T \neq Y_t \mid X_0 = x, Y_0 = y) \leq \varepsilon \qquad \forall x, y \in \Omega.$$

then $\tau(\varepsilon) \leq T_\varepsilon$ ← coupling time

mixing time

Intuitive why this is true — but let's see a proof....

Pf: Suppose $Y_0 \sim \pi^*$ and X_0 = arbitrary.

Then for $T = T_\varepsilon$ and any event $A \subseteq \Omega$

$$\Pr(X_T \in A) \geq \Pr(X_T = Y_T \wedge Y_T \in A)$$

$$= 1 - \left(\Pr\left((X_t \neq Y_t) \vee Y_t \notin A\right)\right)$$

$$\geq 1 - \Pr(X_t \neq Y_t) - \Pr(Y_t \notin A)$$

$$= \Pr(Y_t \in A) - \Pr(X_t \neq Y_t) = \Pr_{\pi^*}(A) - \underbrace{\Pr(X_t \neq Y_t)}_{\varepsilon}$$

$$\Rightarrow \quad \Pr_{\pi_T}(A) - \Pr_{\pi^*}(A) \geq -\varepsilon$$

similarly $\quad \Pr_{\pi_t}(A) - \Pr_{\pi^*}(A) \leq \varepsilon$

$$\Rightarrow \quad |\Pr_{\pi_t}(A) - \Pr_{\pi^*}(A)| \leq \varepsilon \quad \Rightarrow \text{TV dist} \leq \varepsilon$$

X

So the goal is: design a coupling so that $\Pr(X_t = Y_t)$ is high pretty fast!

E.g. Suppose we were choosing random point in hypercube $\{0,1\}^n$

- MC ("Glauber dynamics")

  Pick location $i \in [n]$ u.a.r.

  Pick bit $b \in \{0,1\}$ u.a.r. set $X_i = b$.

- How fast does it mix? Let's design a coupling.

  Given (X_t, Y_t), do the following.

  Pick same location & same bit in both.

  brings things closer, once they merge stay together.

So how long to meet whp $\geq 1-\varepsilon$?

Want to hit all coords at least once, whp. $1-\varepsilon$.

- Exercise: E[time to hit all coords] $= O(n \log n)$. In fact $n \cdot H_n = \mu$

  $\Rightarrow \Pr(\text{not hit all coords in } 2\mu \text{ steps}) \leq 1/2$.

  Now imagine this is start of another round.

  (where merge in round whp 1/2).

  $\Rightarrow \Pr(\text{Not Merged in } k \text{ rounds}) \leq (1/2)^k$

  $\Rightarrow$ Coupling time $\leq O(n \lg n \lg 1/\varepsilon)$.

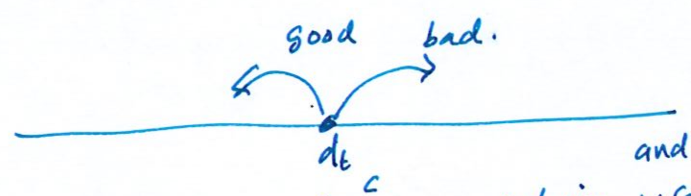
- Exercise: Improve to $O(n \log n/\varepsilon)$.

—— X ——

Back to our Coloring Problem:

Same coupling — both X_t, Y_t pick same vertex v, same color c.

define d_t = # vtxs where X_t, Y_t disagree.



Again: good move if color c not in v's neighborhood in either chain and v disagrees in X_t, Y_t

$\Rightarrow$ at least $K-2\Delta \geq 2\Delta+1$ colors since $K \geq 4\Delta+1$.

$$\Rightarrow \Pr(\text{good}) = \frac{d_t}{Kn} \cdot (K-2\Delta) \geq \frac{2\Delta+1}{Kn} d_t$$

Bad move: v agrees but picks color where can move in one chain but not in other.

Hmm: d "disagreeing" vertices.

- v in nbrhood of such vtx $\Rightarrow$ $\leq d_t \Delta$ choices for v.
- color chosen must be color of that disagreeing vtx! (in either X_t or Y_t)

$\Rightarrow$ 2 choices of color.

$\Rightarrow$ $\Pr(\text{bad move}) \leq \frac{2d_t\Delta}{kn}$.

$\Rightarrow$ drift towards negative

$\Rightarrow$ hit 0 w.p. $\geq 1-\varepsilon$ in time $O(kn \ln(n/\varepsilon))$ ☺

(Formal Proof on next page)

Great: Can sample k-colorings of graphs (approximately) uniformly

when $k \geq 4\Delta+1$

(Better results known, best result has $k \geq$)

$\Rightarrow$ Can use with "self reduction" idea to get

$(1+\varepsilon)$-apx estimate of # colorings in $\text{poly}(n/\varepsilon)$ time.

[Called "Potts model" in Physics, people interested in these questions]

"Hard-Core model" — sample independent sets.

"Ising model" — sample cuts. etc.

Formal Proof:

$$E[d_{t+1} - d_t \mid (x_t, y_t)] \leq -\frac{(2\Delta+1)}{kn} d_t + \frac{2\Delta}{kn} \cdot d_t = -\frac{d_t}{kn}$$

$$\Rightarrow E[d_{t+1} \mid (x_t, y_t)] \leq d_t\left(1 - \frac{1}{kn}\right)$$

$$\Rightarrow E[d_T \mid (x_0, y_0)] \leq n\left(1 - \frac{1}{kn}\right)^T$$

Now: $\Pr(d_t > 0 \mid (x_0, y_0)]$

$$= \Pr(d_t \geq 1 \mid (x_0, y_0)] \leq E[d_t \mid (x_0, y_0)] \leq \varepsilon$$

first moment method.

if $T = kn \ln n/\varepsilon$.

Great:

Saw general intro on unbiased estimators

- FPRAS – approx some value p to within $\pm\varepsilon p$ w.p. $1-\delta$.
- FPRAS for #DNF solutions,

  for k-colorings in graphs where $K \geq$ ~~2Δ~~ $4\Delta+1$

  via sampling

MCMC for sampling colorings

- vast area, see notes by Eric Vigoda or Mark Jerrum or ....
- also active area!