Monte Carlo Methods

collection of tools for estimating values thru sampling & estimation

Today
- more approx counting
- MCMC
- coupling

(ε, δ) Approximation

A randomized alg gives an (ε, δ) approx for value V if the output X of the alg satisfies

$$\Pr(|X - V| > \varepsilon |V|) \leq \delta$$

Monte Carlo Thm

Let $X_1, X_2, \ldots, X_m$ iid. Bernoulli with $E(X_i) = \mu$

If $m \geq \frac{3 \ln(\frac{2}{\delta})}{\varepsilon^2 \mu}$ then

$$\Pr\left(\left| \frac{1}{m} \sum_{i=1}^{m} X_i - \mu \right| \geq \varepsilon \mu \right) \leq \delta$$

Pf: Chernoff bounds

$\#P$ complexity class associated with counting solns to problems in NP

$\#P$ complete problems:

- # indep sets in a graph
- # satisfying assignments to DNF formula
- # perfect matchings in a bipartite graph

Fully polynomial randomized approx scheme FPRAS

a randomized alg for which, given an input x and any parameters ε & δ with $0 < \varepsilon, \delta < 1$, the alg outputs an (ε, δ) approx to $V(x)$ in time poly in $\frac{1}{\varepsilon}$, $\ln \frac{1}{\delta}$ & size of input

x DNF formula — $V(x)$ # satisfying assignments

x graph. — $V(x)$ # indep sets in graph.

Note: suffices to take $\delta = \frac{1}{4}$

because easy to boost error prob:

Run $k = 16 \log\left(\frac{2}{\delta}\right)$ trials w. error prob $\frac{1}{4}$ $\Rightarrow$ $y_1, y_2, \ldots, y_k$

Let $m = \text{median}(y_1, \ldots, y_k)$

Then $\Pr\left[m \notin (1 \pm \varepsilon) V(x)\right] \leq \delta$ by Chernoff

Want to sample from some set (e.g. ISs of graph G)

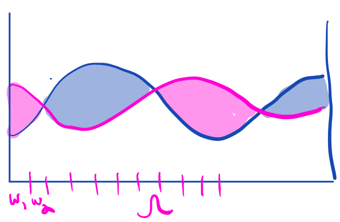
where π is desired distribution $\pi_I = \Pr(\text{output IS } I)$

(often π uniform distn)

Definition

total variation distance between 2 distns on some sample space Ω

$\|D_1, D_2\|_{TV}$ $\quad \|D_1 - D_2\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |D_1(x) - D_2(x)| = \max_{A \subseteq \Omega} |D_1(A) - D_2(A)|$



$\|D_1 - D_2\|_{TV}$ = pink area = blue area

For sampling problems, we seek

$\boxed{\text{FPAUS - fully polynomial almost uniform sampler}}$

Input: x, ε $\longrightarrow$ $\boxed{\text{FPAUS: Fully polynomial almost uniform sampler}}$ $\longrightarrow$ outputs samples from $\Omega(x)$ sampled from distn μ

graph

runs in time polynomial in size of x & $\ln\left(\frac{1}{\varepsilon}\right)$

s.t. $d_{TV}(\mu, \pi) \leq \varepsilon$

output distn.

$\Omega(x)$ set of indep sets in G

π desired distn uniform

$$\max_S \left| \Pr_\mu(\text{sample in } S) - \underbrace{\pi(S)}_{\frac{|S|}{|\Omega(x)|}} \right| \leq \varepsilon$$

Approximate counting $\Longleftrightarrow$ Approximate Sampling
FPRAS FPAUS

self reducible problems
randomized reduction from problems of size k to problems of size $k-1$

Lemma Given FPAUS for sampling indep sets of graph G, we can construct FPRAS for estimating $|\Omega(G)|$ # ISs.

Pf Want estimate $\bar{\Omega}$ of $|\Omega(G)|$

$$\Pr\left(\left|\bar{\Omega} - |\Omega(G)|\right| \geq \varepsilon |\Omega(G)| \right) \leq \delta$$

$G=(V,E)$ $e_1, e_2, \ldots, e_m$ arbitrary ordering of edges

$E_i = \{e_1, e_2, \ldots, e_i\}$ $G_i = (V, E_i)$ $G_m = G$ $G_0 = (V, \phi)$

$\Omega(G_i)$: set of ISs in G_i

$$|\Omega(G)| = \frac{|\Omega(G_m)|}{|\Omega(G_{m-1})|} \times \frac{|\Omega(G_{m-1})|}{|\Omega(G_{m-2})|} \times \cdots \times \frac{|\Omega(G_1)|}{|\Omega(G_0)|} \times \underbrace{|\Omega(G_0)|}_{2^n}$$

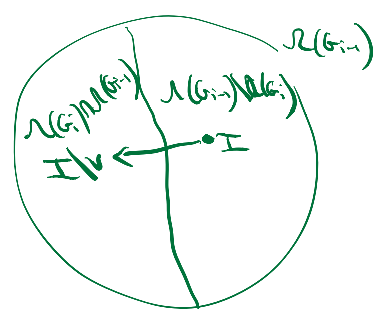
Let $r_i = \dfrac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$

$$\underline{|\Omega(G)| = \prod_{i=1}^{m} r_i \, 2^n}$$

Observation: $\boxed{\frac{1}{2} \leq r_i \leq 1}$

G_i has one extra edge, say (u,v)

$\Omega(G_i) \subseteq \Omega(G_{i-1})$



Can't sample from uniform distn on $\Omega(G_{i-1})$
but can use FPAUS

Estimate r_i (call estimate $\tilde{r}_i$) Output as my estimate for $|\Omega(G)| = \prod_{i=1}^{m} \tilde{r}_i \; 2^n$

Two errors we need to bound: $\left| \overbrace{Pr(\text{sample in } \Omega(G_i))}^{E(\tilde{r}_i)} - \underbrace{\frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}}_{r_i} \right| \leq \frac{\varepsilon}{6m}$

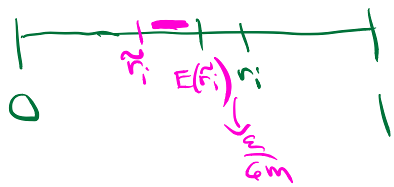
(1) FPAUS $\neq$ exact sampler so exp value $\neq r_i$

But using, say, $\boxed{\frac{\varepsilon}{6m}}$ sampler $|E(\tilde{r}_i) - r_i| \leq \frac{\varepsilon}{6m}$

(2) With samples we get approx to $E(\tilde{r}_i)$

since r_i big $(\geq \frac{1}{2})$, $E(\tilde{r}_i)$ big $\Rightarrow$

Monte Carlo Thm $\Rightarrow$ $O\left(\frac{m^2}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right)\right)$ samples suffice



From these bound $R = \prod_{i=1}^{m} \frac{\tilde{r}_i}{r_i}$

Claim: If $\tilde{r}_i$ is $\left(\frac{\varepsilon}{2m}, \frac{\delta}{m}\right)$ approx to r_i $\forall i$
then X is (ε, δ) approx to $|\Omega(G)|$

$X = \prod_{i=1}^{m} \tilde{r}_i \; 2^n$

$|\Omega(G)| = \prod_{i=1}^{m} r_i \; 2^n$

More details

Use FPAUS with $\eta = \frac{\varepsilon}{12m}$

$$\Rightarrow \left| \underbrace{\Pr(\text{sample in } \Omega(G_i))}_{E(\tilde{r}_i)} - \underbrace{\frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}}_{r_i} \right| \le \eta$$

Use FPAUS $s = O\left(\left(\frac{m}{\varepsilon}\right)^2 \lg\left(\frac{2m}{\delta}\right)\right)$ times. on G_{i-1}

since $E(\tilde{r}_i) \ge \frac{1}{2} - \eta$

By Chernoff this guarantees)

$$\Pr\left(|\tilde{r}_i - E(\tilde{r}_i)| \ge E(\tilde{r}_i)\frac{\varepsilon}{12m} \right) \le \frac{\delta}{m}$$

where $\tilde{r}_i$ = fraction of FPAUS samples in $\Omega(G_i)$

$$\Rightarrow E(\tilde{r}_i)\left(1 - \frac{\varepsilon}{12m}\right) \le \tilde{r}_i \le \underbrace{E(\tilde{r}_i)\left(1 + \frac{\varepsilon}{12m}\right)} \quad \text{w.p.} \ge 1 - \frac{\delta}{m}$$

$$\le (r_i + \eta)\left(1 + \frac{\varepsilon}{12m}\right)$$

$$\le r_i\left(1 + \frac{\eta}{r_i}\right)\left(1 + \frac{\varepsilon}{12m}\right)$$

$$\le r_i(1 + 2\eta)\left(1 + \frac{\varepsilon}{12m}\right)$$

$$\le r_i\left(1 + \frac{\varepsilon}{6m}\right)\left(1 + \frac{\varepsilon}{12m}\right)$$

$$r_i\left(1 - \frac{\varepsilon}{3m}\right) \le \tilde{r}_i \le r_i\left(1 + \frac{\varepsilon}{3m}\right) \quad \text{w.p.} \ge 1 - \frac{\delta}{m}$$

$$\Rightarrow \prod r_i \left(1 - \frac{\varepsilon}{3m}\right)^m \le \prod \tilde{r}_i \le \prod r_i \left(1 + \frac{\varepsilon}{3m}\right)^m$$

w.p. $\ge 1 - \delta$ (by union bound)

$$\Rightarrow \prod r_i (1 - \varepsilon) \le \prod \tilde{r}_i \le \prod r_i (1 + \varepsilon) \quad \text{w.p.} \ge 1 - \delta$$

✓

Altogether $O^*(m^3)$ samples.

More careful analysis $\Rightarrow$ $O^*(m^2)$ samples

How to sample elts from a universe $\boxed{\Omega}$ with $|U|=n$ according to distn $\pi=(\pi_1,\dots,\pi_n)$?

→ ISs in a graph

→ desired distn that I want to sample

Cool idea: design MC whose state space is Ω that has stationary distn π

- simulate MC until it "mixes"
- use state at that time as sample

Ω want to sample fm Ω s.t. $\Pr(\text{sample is } w_i) = \pi_{w_i}$

$p_0 P^t \to \pi$

$\Pr(X_t = i) \approx \pi_i \quad \forall i$

2 key questions:

1. how to design chain w/ right π?
2. how to bound mixing time?

Example: sampling indep sets uniformly from $G=(V,E)$

states of MC: indep sets $(I_1, I_2, \dots, I_R)$ designing

X_t : some indep set $\qquad X_t$

MC:

choose vertex v u.a.r. from V

if $v \in X_t$, then $X_{t+1} := X_t \setminus v$

if $v \notin X_t$ and can be added without violating independence then $X_{t+1} := X_t \cup v$

otherwise $X_{t+1} := X_t$

$I \quad (v_1, \dots, v_k) \qquad I' \quad (w_1, \dots, w_r)$

I_u

Claim:

- MC irreducible
- if $\exists$ edge (u,v) then aperiodic
- stationary distn uniform (chain doubly stochastic)

$P_{I,I'} = P_{I',I} \qquad \left[\frac{1}{n} \text{ or } 0\right]$

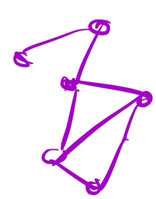
General technique: given Ω and a connected graph on Ω define transition prob s so that will have stationary distn π

Metropolis Algorithm

Input: Ω ; $\underline{\text{connected graph } G=(\Omega, E)}$, π s.t. $\sum_{i\in\Omega} \pi_i = 1$, $\pi_i > 0$

Let Δ max degree in G

$$P_{xy} = \begin{cases} \frac{1}{2\Delta}\min\left(1, \frac{\pi_y}{\pi_x}\right) & x \neq y \;\; y \in N(x) \\ 0 & x \neq y \;\; y \notin N(x) \\ 1 - \sum_{y \neq x} P_{xy} & x = y \end{cases}$$



Claim $\quad \pi_x P_{xy} = \pi_y P_{yx} \quad \forall\, x \neq y$

wlog $\pi_x < \pi_y$

$$\pi_x P_{xy} = \frac{\pi_x}{2\Delta}$$

$$\pi_y P_{yx} = \pi_y \frac{1}{2\Delta} \frac{\pi_x}{\pi_y} = \frac{\pi_x}{2\Delta}$$

Example application

Suppose I want to sample ISs according to $\pi(I) = \frac{\lambda^{|I|}}{Z}$ $\quad Z = \sum_I \lambda^{|I|}$

$\lambda > 1$

use graph above.

Bounding mixing time

(1) Spectral gap, conductance, expansion

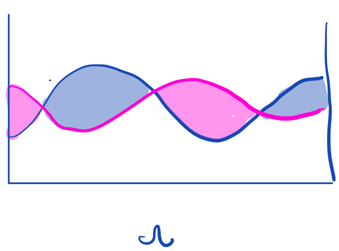
(2) coupling

⋮

$\boxed{\text{Coupling}}$

total variation distance between 2 distns on some sample space Ω

$$\|D_1 - D_2\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |D_1(x) - D_2(x)| = \max_{A \subseteq \Omega} |D_1(A) - D_2(A)|$$



$\|D_1 - D_2\|_{TV} =$ pink area $=$ blue area

Common defn of mixing time $\tau(\varepsilon)$

$$\tau(\varepsilon) = \min \{ t \mid \|p_0^t - \pi\|_{TV} \le \varepsilon \}$$

say MC is rapidly mixing if $\tau(\varepsilon)$ polynomial in $\log|\Omega|$ and $\log\left(\frac{1}{\varepsilon}\right)$

(we know this is related to spectral gap)

$\boxed{\text{Coupling}}$ simple & elegant approach to bounding mixing time

Given a MC P on Ω, a coupling is a MC on $\Omega \times \Omega$ defining stochastic process (X_t, Y_t) s.t.

① each of X_t & Y_t in isolation is faithful copy of MC

$$\Pr\left(X_{t+1} = z \mid (X_t, Y_t) = (x,y)\right) = P_{xz}$$

$$\Pr\left(Y_{t+1} = w \mid (X_t, Y_t) = (x,y)\right) = P_{yw}$$

② If $X_t = Y_t$ then $X_{t+1} = Y_{t+1}$

$\boxed{\text{Coupling Lemma}}$ Let (X_t, Y_t) be a coupling

Suppose $\exists\, T$ s.t. $\forall x, y$

$$\Pr\left(X_T \neq Y_T \mid X_0 = x, Y_0 = y\right) \leq \varepsilon$$

Then $\tau(\varepsilon) \leq T$

$\tau(\varepsilon) = \min\{t \mid \|p_0^t - \pi\|_{TV} \leq \varepsilon\}$

Pf Pick coupling where $Y_0 \sim \pi$

$\forall A \subseteq \Omega$ $\Pr(X_T \in A) \geq \Pr(X_T = Y_T \cap Y_T \in A)$

$\sum_{w \in A} \Pr(X_T = w)$

$$= 1 - \Pr(X_T \neq Y_T \cup Y_T \notin A)$$

$$\geq \underline{1} - \Pr(X_T \neq Y_T) - \underbrace{\Pr(Y_T \notin A)}_{\pi_{\bar{A}}}$$

$$\geq \pi_A - \varepsilon$$

same arg $\Pr(X_T \notin A) \geq \pi_{\bar{A}} - \varepsilon$

$$\Pr(X_T \in A) \leq \pi_A + \varepsilon$$

$$\left|\Pr(X_T \in A) - \pi_A\right| \leq \varepsilon$$

$$\|\text{dist of } X_T, \pi\|_{TV} \leq \varepsilon$$

Examples:

(1) Random walk on hypercube $N = 2^n$ nodes

(001001)

101001 011001 ....

in each step, choose random coordinate i, random bit $b \in \{0,1\}$

change i^{th} bit to b

x_0 (001001) $y_0 =$ (110101)

001011 110111

001011 100111

Find T s.t. $\Pr(X_T \neq Y_T) \leq \varepsilon$

$E(\#\text{steps until } X_T = Y_T)$

$= n \log n.$

$\Pr($after $\underbrace{n\log n + cn}$ steps i hasn't been sampled$)$

$$= \left(1 - \frac{1}{n}\right)^{n \ln n + cn}$$

$$\leq e^{-(\ln n + c)} = \frac{e^{-c}}{n}$$

$$c = \ln\left(\frac{1}{\varepsilon}\right) \qquad \Rightarrow \frac{\varepsilon}{n}$$

$\Pr(\exists i$ that hasn't been sampled$) \leq \varepsilon,$

$n \log n + n \ln\left(\frac{1}{\varepsilon}\right) = n \ln\left(\frac{n}{\varepsilon}\right)$ steps