Approximate Counting

Decision problem in NP:

Π is in NP if for any YES instance I, $\exists$ a proof that I is a YES instance that can be verified in poly time.

Ex: "Is there a Hamiltonian cycle in the given graph?"

"Is there a tour of length $\leq k$ for a travelling Salesperson instance".

"Is the given Boolean formula satisfiable".

Corresponding to these decision problems, we can define counting problems.

Counting problem for a decision problem Π in NP:

Given an instance I of Π, produce as output a non-negative integer that is the no. of solutions (or proofs) for the instance I.

$$\#P = \left\{ F_n : \{0,1\}^{n^c} \to \mathbb{Z}_{\geq 0} \;\middle|\; F_n \text{ counts the no. of accepting paths in a NTM, and } n>0 \right\}$$

Ex: "No. of Hamiltonian cycles in the given graph"

"No of satisfying assignments to a Boolean formula"

"No of k-cliques in a given graph".

These are at least as hard as their decision versions

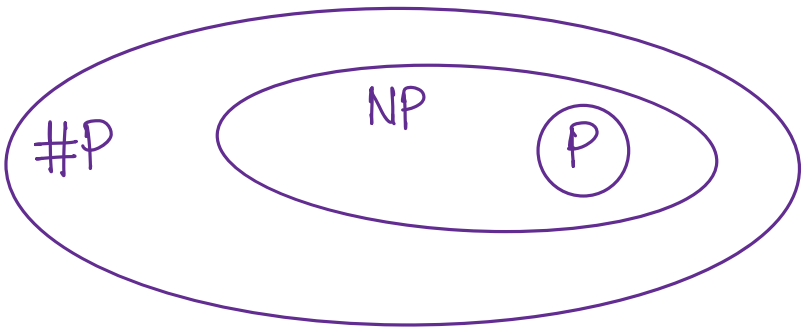


Figure: Hierarchy of functional Complexity classes.

Qn: Can we approximately count them?

- We do not know of det. approx. schemes
- what about randomized approx schemes?

Defn: A polynomial approximation scheme (PAS) for a counting problem P is a deterministic algo that takes as input an instance I and $\varepsilon \in \mathbb{R}_{>0}$, and in time polynomial in $n = |I|$ produces an output $A(I)$ such that

$$(1-\varepsilon)\,\#(I) \leq A(I) \leq (1+\varepsilon)\,\#(I).$$

A fully polynomial approximation scheme (FPAS) is a PAS that runs in time $\text{poly}(n, 1/\varepsilon)$.

Defn: A polynomial randomized approximation scheme (PRAS) for a counting problem P is a randomized algorithm A that takes as input an instance I and a real no. $\varepsilon > 0$, an in time poly(n) produces A(I) s.t

$$\Pr\left[(1-\varepsilon)\#(I) \leq A(I) \leq (1+\varepsilon)\#(I)\right] \geq \frac{3}{4}.$$

Fully polynomial randomized approximation scheme (FPRAS) is a PRAS that runs in time poly$(n, 1/\varepsilon)$.

With prob $< \frac{1}{4}$, we assume nothing about the dist of A(I) from #(I).

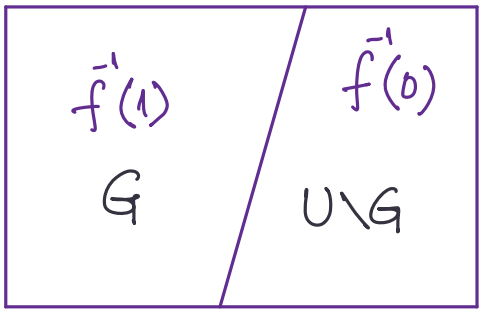
Defn: An (ε, δ)- FPRAS for a counting problem is a FPRAS that takes as input an instance I and computes an ε- approximation to #(I) w.p $\geq 1-\delta$ in time poly$(n, 1/\varepsilon, 1/\delta)$.

(n = $|I|$)

Abstract problem:

Let U be a finite set of known size. Let $f: U \to \{0,1\}$ be a Boolean function over U. Let $G = \{u \in U \mid f(u) = 1\}$.
Assumptions: u can be sampled uniformly and

randomly from U. Given $u \in U$, $f(u)$ can be computed "quickly".



Question: Estimate $|G|$.

Attempt 1:

Sample $u_1, \dots, u_N$ from U independently from U.

For all $i \in [N]$, let

$$Y_i = \begin{cases} 1 & \text{if } f(u_i) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

$Y_i = 1$ iff $u_i \in G$.

Let Z be a r.v s.t

$$Z = \frac{|U|}{} \cdot \sum_{i=1}^{N} \frac{Y_i}{N} .$$

We call it estimator random variable.

Claim: $\mathbb{E}[Z] = |G|$. ← $\mathbb{E}[\sum Y_i] \cdot \frac{|U|}{N} = \sum_{i=1}^{N} \mathbb{E}[Y_i] \cdot \frac{|U|}{N}$

$= \frac{|U|}{N} \sum_{i=1}^{N} \Pr[Y_i = 1]$

$= \Pr[Y_i = 1] \hookrightarrow |G|/|U|$

[Estimator theorem]

Theorem: Let $\rho = \frac{|G|}{|U|}$. Then the Monte Carlo method yields an ε-approximation to $|G|$ w. prob $\geq 1 - \delta$ provided

$$N \geq \frac{4}{\varepsilon^2 \rho} \cdot \ln \frac{2}{\delta} .$$

Proof: Note that $\Pr[Y_i = 1] = \frac{|G|}{|U|}$. Let $Y = \sum_{i=1}^{N} Y_i$.

$$Z = \frac{|U| \cdot Y}{N}$$

$$\Pr\left[(1-\varepsilon)|G| \le Z \le (1+\varepsilon)|G|\right]$$

$$\Pr\left[(1-\varepsilon)|G| \le \frac{|U| \cdot Y}{N} \le (1+\varepsilon)|G|\right] = \Pr\left[(1-\varepsilon) \cdot \frac{|G|}{|U|} \cdot N \le \frac{|U| \cdot Y}{N} \cdot \frac{N}{|U|} \le (1+\varepsilon) \cdot \frac{|G|}{|U|} \cdot N\right]$$

$$= \Pr\left[(1-\varepsilon) N\rho \le Y \le (1+\varepsilon) \cdot N\rho\right] \quad \left(\frac{|G|}{|U|} \to \rho\right)$$

$$\rho = \frac{|G|}{|U|}.$$

$$\ge 1 - \Pr[Y > (1+\varepsilon) N\rho] - \Pr[Y < (1-\varepsilon) N\rho]$$

$$= 1 - 2 \cdot e^{-N\rho\varepsilon^2/4} \Rightarrow \frac{\delta}{2} \ge e^{-\frac{N\rho\varepsilon^2}{4}}$$

$$2^n$$

$$\ge 1 - \delta$$

$$\delta \ge 2 \cdot e^{-N\rho \cdot \varepsilon^2/4} \Rightarrow \frac{N\rho\varepsilon^2}{4} \ge \ln\frac{2}{\delta}$$

$$N = \text{poly}\left(\frac{1}{\varepsilon}, \ln\frac{1}{\delta}\right) \left(\frac{1}{\rho}\right)$$

$$\Rightarrow N \ge \frac{4}{\rho\varepsilon^2} \ln\frac{2}{\delta}.$$

Question: But, N is dependent on $\rho = \frac{|G|}{|U|}$.
Problem is not solved yet.

↳ N could be exponential.

Ex: F is a Boolean formula in Disjunctive Normal Form. That is OR of ANDs.

$$F = T_1 \vee T_2 \vee \cdots \vee T_m \quad \text{and} \quad T_i = L_1 \wedge L_2 \cdots \wedge L_k$$

($L_1 \wedge L_2 \cdots \wedge L_k$: over n variables)

$T_1 \vee T_2 \vee \cdots \vee T_m$: OR of terms = Conjunctive clauses

$L_1 \wedge L_2 \cdots \wedge L_k$: AND of literals.

$U = \{T,F\}^n$

Conjunction = $\wedge$, disjunction = $\vee$.

#F := No. of satisfying solutions of F.

$G = \{u \in U \mid F(u) = T\}$.

$\frac{|G|}{|U|}$ could be inverse exp.

Fix: Coverage argument // Biased sampling.

$H_1, \dots, H_m \subseteq V$ ← space of assignments.

H_i = subset of assignments that satisfy term T_i.

$\left| \bigcup_{i=1}^{m} H_i \right|$ is exactly the count we are interested in → H.

$|H_i| = 2^{n-r_i}$ $\quad x_1 \wedge x_2 \cdots \wedge x_{r_i}$ $\quad T_i = L_{i_1} \wedge \cdots \wedge L_{i_{r_i}}$

$|H_i|$: T_i has r_i literals.

$U = H_1 \uplus H_2 \uplus \cdots \uplus H_m$ } multiset union.

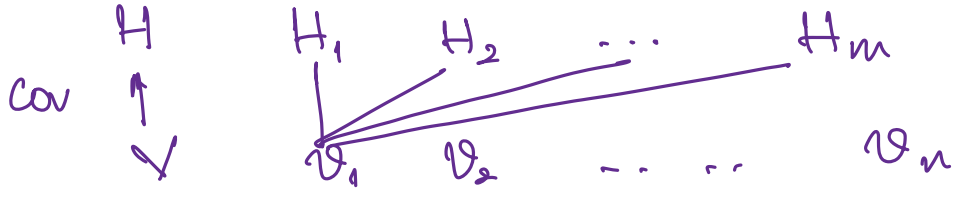
$|U| = \sum_{i=1}^{m} |H_i| \geq |H|$ where $H = \bigcup_{i=1}^{m} H_i$

$U = \{(v, i) \mid v \in H_i\}$

v $T_1, T_2, T_3, \dots$
$(v,1)$ $(v,2)$ $(v,3)$

Coverage of an assignment v:

$$cov(v) = \{ (v,i) \mid (v,i) \in U \} \;//\; \{ i \mid (v,i) \in U \}$$



$-|cov(v)| \leq m$. and $U = \bigcup_{v \in H} cov(v)$

$$|U| = \sum_{v \in H} |cov(v)|.$$

Defn: $f : U \to \{0,1\}$ be defined as follows:

$$f((v,i)) = \begin{cases} 1 & \text{if } i = \min\{j \mid (v,j) \in U\} \\ 0 \end{cases}$$

and $G = \{(v,j) \mid f((v,j)) = 1\}$.

Obs: $|G|$ is exactly $|H|$

It is easy to compute H_i and its cardinality.

{Similarly, $|H|$ can be computed easily.

Lemma: $p = \frac{|G|}{|U|} \geq \frac{1}{m}$.

Not done is to tell you how to sample from U.

Proof: $|U| = \sum_{v \in H} |cov(v)|$

$$\leq \sum_{v \in H} m$$

$$= m|H|.$$

$$= m \cdot |G|. \Rightarrow \frac{|G|}{|U|} \geq \frac{1}{m}.$$

[Estimator theorem]

Theorem: Let $\rho \geq 1/m$. Then the Monte Carlo method yields an ε-approximation to $|G|$ w. prob $\geq 1-\delta$ provided

$$N \geq \frac{4m}{\varepsilon^2} \cdot \ln \frac{2}{\delta}.$$

T_1 T_2
v
$(v,1)$ $(v,2)$

Recall that $u_1, \dots, u_N$ were sampled uniformly and ind.

$(v_1, i_1), (v_2, i_2), \dots, (v_N, i_N)$.

for (v,i) we define a way to sample it uniformly.

→ First sample $i \in [m]$ w.p $\frac{|H_i|}{|U|} = \frac{|H_i|}{\sum_{i=1}^{m} |H_i|}$ ← no. of assignments satisfying term i.

$U = H_1 \uplus H_2 \uplus \dots \uplus H_m$.

$p_i = \Pr[\text{sampling } i] = \frac{|H_i|}{|U|}$ and $\sum_{i=1}^{m} p_i = 1$

→ Sample v from H_i uniformly ↗ $(\underbrace{\quad}_{r_i} \downarrow\downarrow \downarrow)$ 2^{n-r_i}

$\Pr[(v,i) \text{ is sampled}] = \Pr[v \text{ is sampled from } H_i \mid i \text{ is sampled from } [m]]$

$\cdot \Pr[i \text{ is sampled}].$

$= \frac{1}{|H_i|} \times \frac{|H_i|}{|U|} = \frac{1}{|U|}.$

$N = \text{poly}(m, \frac{1}{\varepsilon}, \frac{1}{\delta})$.

$$Z' = \frac{|U'|}{N} \cdot \sum Y_i$$

$$\mathbb{E}[Z'] = |G|.$$

$T_1 \vee T_2 \vee \cdots \vee T_m$. $|H_i| = 2^{n-r_i}$

→ Define a distribution

i, $\Pr[i] = \frac{2^{n-r_i}}{\sum_{i=1}^{m} 2^{n-r_i}}$. ✓

→ Sample $v \in H_i$

$N \geq \frac{4m}{\varepsilon^2} \ln \frac{2}{\delta}$

$1-\delta$.

ε.