Def: A polynomial randomized approx. sheme (PRAS) for a counting problem Π is a randomized alg. A that takes an input I, real number $\varepsilon > 0$, and in $poly(n)$ time produces output $A(I)$ s.t

$$P\left[(1-\varepsilon)\underbrace{\#(I)}_{\substack{\text{number}\\ \text{of sols}\\ \text{to } I.}} \le A(I) \le (1+\varepsilon)\#(I) \right] \ge \underbrace{3/4}_{\delta}$$

Def: FPRAS poly time in $\left(n, \frac{1}{\varepsilon}, \lg \frac{1}{\delta}\right)$.

counting DNF

$f: U \to \{0,1\}$ boolean function over U

$G = \{ u \in U : f(u) = 1 \}$.

Assume computing $f(u)$ takes unit time.

$F(X_1, \dots, X_n) :=$ Boolean formula in DNF form.

let $U = \{0,1\}^n$ set of all 2^n assignments

Classical Monte-Carlo method:

- N independent samples from U say $u_1, u_2, \ldots, u_N$.

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

estimator:

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

<u>Thm (Estimator theorem)</u>: (straightforward from Chernoff bound)

Let $p = \frac{|G|}{|U|}$. The Monte-Carlo method yields an (ε, δ)-approximation provided:

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

<u>problem</u>: $\frac{1}{p}$ could be exponentially large.

<u>idea</u>: importance sampling.

skewed sampling that concentrates the probability on the sub-problem of interest.

The coverage algorithm:

- Let V be a universe.
- $H_1, H_2, \dots, H_m \subseteq V$ s.t
  1) $\forall i$, $|H_i|$ can be computed in poly-time.
  2) It's possible to sample uar from any H_i
  3) $\forall v \in V$, one can determine whether $v \in H_i$ $\forall i$

<u>Goal:</u> Estimate $|H| = |H_1 \cup H_2 \cup \dots \cup H_m|$

- Consider a DNF formula $F(X_1, \dots, X_n)$
- let the i^{th} clause C_i be a conjunction of r_i literals.

- $V = 2^n$ assignments
- $H_i =$ assignments that satisfy C_i

- We have:

$$|H_i| = 2^{n - r_i}$$

- To sample from H_i, set all literals in H_i to true and the rest are set randomly.
- Testing $v \in H_i$ can be done in linear time.

- Def:

$$\mathcal{U} = H_1 \uplus H_2 \uplus \cdots \uplus H_m$$

multiset union.

- convention: the elms of $\mathcal{U}$ are ordered pairs (v, i) corresponding to $v \in H_i$

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

- Observe that:

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

- Def: the coverage set of $v \in V$

$$\mathrm{cov}(v) = \{(v,i) : (v,i) \in \mathcal{U}\}$$

the size of $\mathrm{cov}(v)$ is the multiplicity of v in $\mathcal{U}$.

- The following are immediate:

1) The number of cover sets is exactly $|H|$. These coverage sets are easy to compute.

2) The coverage sets partition $\mathcal{U}$

$$\mathcal{U} = \bigcup_v \mathrm{cov}(v).$$

3) $|U|$ is easily computed as

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

4) For all $v \in H$, $cov(v) \leq m$.

Def: The function $f: U \to \{0,1\}$ is defined as:

$$f((v,i)) = \begin{cases} 1 & \text{if } i = \min\{j : v \in H_j\} \\ 0 & \text{o.w.} \end{cases}$$

The set G can be rewritten as:

$$G = \{(v,j) \in U : f((v,i)) = 1\}.$$

(assign v to the lowest index C_j where v satisfies C_j).

Def. canonical elm. for $cov(v)$. as the elm that corresponds to the occurence of v in the lowest-index $H_j \ni v$.

Crucial observation: $|G| = |H|$.

Goal: Estimate the size $G \subseteq U$ s.t

$G = f^{-1}(1)$

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

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

$\leq m \cdot |H| = m \cdot |G|$.

$\Rightarrow \frac{|U|}{|G|} \leq m$.

Thm: The monte-carlo method yields an (ε, δ)-approximation w/

$N \geq \frac{4m}{\varepsilon^2} \lg \frac{2}{\delta}$ samples. in poly-time.

pf: To sample from U:

$\mathbb{P}[\text{pick } i] = \frac{|H_i|}{|U|} = \frac{2^{n-r_i}}{\sum_j |H_j|}$

Then sample $v \in H_i$ u.a.r

To check if $v \in G$: check if $\forall j < i$, v does not satisfy C_j

Intuition: sampling from V fails b/c V can be exponentially larger than H.

In the new approach each satisfying assignment is sampled proportional to the # clauses it satisfies.