Exact counting

Exact counting is rare:

- #2-COLORINGS.
- #PERFECT MATCHINGS in planar graphs.
- #SPANNING TREES in general graphs.

Approximate counting

Definition

A fully polynomial randomized approximation scheme (fpras) for a counting problem $f : \Sigma^* \to \mathbb{N}$ is a randomized algorithm that takes as input an instance $x \in \Sigma^*$, an error tolerance $0 < \varepsilon < 1$, and $0 < \delta < 1$, and outputs a number $\widehat{f(x)} \in \mathbb{N}$ such that

$$\Pr[(1-\varepsilon)f(x) \leq \widehat{f(x)} \leq (1+\varepsilon)f(x)] \geq 1-\delta.$$

The algorithm must run in time polynomial in $|x|$, $1/\varepsilon$ and $\log(1/\delta)$.

- For example, given $\varepsilon = 0.1$, we would have

$$0.9 \leq \frac{\widehat{f(x)}}{f(x)} \leq 1.1$$

  with high probability.

- Given $|x|$, ε can be an inverse polynomial of $|x|$, and δ can be inversely exponential in $|x|$.

Alternatively we could define the fpras so that on input (x, ε) outputs $\widehat{f(x)}$ satisfying

$$\Pr[(1-\varepsilon)f(x) \leq \widehat{f(x)} \leq (1+\varepsilon)f(x)] \geq \frac{3}{4} \quad \text{or}$$

$$\Pr[e^{-\varepsilon}f(x) \leq \widehat{f(x)} \leq e^{\varepsilon}f(x)] \geq \frac{3}{4}$$

The probability $\frac{3}{4}$ can be boosted to $1-\delta$ for any desired $\delta > 0$ using $\mathcal{O}(\log \delta^{-1})$ repeated trials.

Boost the success probability to $1 - \delta$

Chernoff bound

Let $X_1, ..., X_m$ be independent, identically distibuted $\{0, 1\}$ random variables, where $p = \mathbb{E}[X_i]$. For all $\varepsilon \leq 3/2$,

$$\Pr[|\sum_{i=1}^{n} X_i - pn| > \varepsilon pn] \leq 2\exp(-\varepsilon^2 pn/3).$$

- Let an fpras for $f(x)$ such that

$$\Pr[(1-\varepsilon)f(x) \leq \widehat{f(x)} \leq (1+\varepsilon)f(x)] = \frac{3}{4}.$$

Boost the success probability to $1 - \delta$

Chernoff bound

Let $X_1, \ldots, X_m$ be independent, identically distibuted $\{0, 1\}$ random variables, where $p = \mathbb{E}[X_i]$. For all $\varepsilon \leq 3/2$,

$$\Pr[|\sum_{i=1}^{n} X_i - pn| > \varepsilon pn] \leq 2\exp(-\varepsilon^2 pn/3).$$

- Let an fpras for $f(x)$ such that

$$\Pr[(1-\varepsilon)f(x) \leq \widehat{f(x)} \leq (1+\varepsilon)f(x)] = \frac{3}{4}.$$

- Then run this algorithm for $k = 36\log(2/\delta)$ times, obtaining outputs $y_1, \ldots, y_k$. Output the median of these outputs, let's say y_m.

Boost the success probability to $1 - \delta$

Chernoff bound

Let $X_1, ..., X_m$ be independent, identically distibuted $\{0, 1\}$ random variables, where $p = \mathbb{E}[X_i]$. For all $\varepsilon \leq 3/2$,

$$\Pr[|\sum_{i=1}^{n} X_i - pn| > \varepsilon pn] \leq 2\exp(-\varepsilon^2 pn/3).$$

- Let an fpras for $f(x)$ such that

$$\Pr[(1-\varepsilon)f(x) \leq \widehat{f(x)} \leq (1+\varepsilon)f(x)] = \frac{3}{4}.$$

- Then run this algorithm for $k = 36\log(2/\delta)$ times, obtaining outputs $y_1, ..., y_k$. Output the median of these outputs, let's say y_m.
- Let $X_i = \begin{cases} 1, & \text{if } y_i \in (1 \pm \varepsilon)f(x) \\ 0, & \text{otherwise} \end{cases}.$

Boost the success probability to $1 - \delta$

Chernoff bound

Let $X_1, \ldots, X_m$ be independent, identically distibuted $\{0,1\}$ random variables, where $p = \mathbb{E}[X_i]$. For all $\varepsilon \leq 3/2$,

$$\Pr[|\sum_{i=1}^{n} X_i - pn| > \varepsilon pn] \leq 2\exp(-\varepsilon^2 pn/3).$$

- Let an fpras for $f(x)$ such that

$$\Pr[(1-\varepsilon)f(x) \leq \widehat{f(x)} \leq (1+\varepsilon)f(x)] = \frac{3}{4}.$$

- Then run this algorithm for $k = 36\log(2/\delta)$ times, obtaining outputs $y_1, \ldots, y_k$. Output the median of these outputs, let's say y_m.
- Let $X_i = \begin{cases} 1, & \text{if } y_i \in (1 \pm \varepsilon)f(x) \\ 0, & \text{otherwise} \end{cases}.$
- $\mathbb{E}[X_i] = \frac{3}{4}$ and $\mathbb{E}[\sum X_i] = \frac{3}{4}k$.

Boost the success probability to $1 - \delta$

Chernoff

Let $X_1, \ldots, X_m$ be independent, identically distibuted $\{0, 1\}$ random variables, where $p = \mathbb{E}[X_i]$. For all $\varepsilon \leq 3/2$,

$$\Pr[|\sum_{i=1}^{n} X_i - pn| > \varepsilon pn] \leq 2\exp(-\varepsilon^2 pn/3).$$

$$\begin{aligned}
\Pr[y_m \notin (1 \pm \varepsilon) f(x)] &\leq \Pr[\sum_{i=1}^{k} X_i < \frac{k}{2}] \\
&\leq \Pr[|\sum X_i - \mathbb{E}[\sum X_i]| > \frac{k}{4}] \\
&\leq \Pr[|\sum X_i - \frac{3}{4}k| > \frac{1}{3} \cdot \frac{3}{4} \cdot k] \\
&\leq 2\exp\left(-\frac{\left(\frac{1}{3}\right)^2 \frac{3}{4}k}{3}\right) = 2\exp(-k/36) = \delta.
\end{aligned}$$

Uniform sampling

- A sampling problem is specified by a relation $R \subseteq \Sigma^* \times \Sigma^*$ between problem instances and solutions, i.e. $(x, w) \in R$ iff w is a solution for the problem instance x.
- We denote the solution set $\{w \mid (x, w) \in R\}$ by $R(x)$.
- A uniform sampler for a solution set $R \in \Sigma^* \times \Sigma^*$ is a randomized algorithm that takes as input an instance $x \in \Sigma^*$ and outputs a solution $W \in R(x)$ uniformly at random.

Total variation distance

To define *approximate sampling* we first need to define the following notion of distance between two probability distributions.

Definition

For two probability distributions μ and ν on a countable set Ω, define the total variation distance between π and π' to be

$$||\mu - \nu||_{TV} = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)|.$$

Claim

For two probability distributions μ and ν,

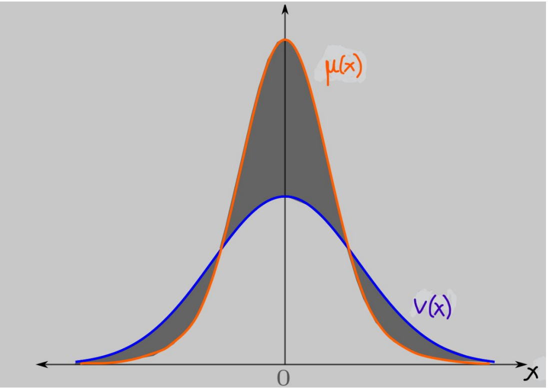
$$||\mu - \nu||_{TV} = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|.$$

For two probability distributions μ and ν,

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

Proof.

Let $S = \{x \mid \mu(x) \geq \nu(x)\}$.



Proof cont.

Since μ, ν are probability distributions $\sum_{x \in \Omega} \mu(x) = \sum_{x \in \Omega} \nu(x) = 1$. So,

$$\sum_{x \in S} \mu(x) - \nu(x) = \sum_{x \notin S} \nu(x) - \mu(x) =$$

Also, $\sum_{x \in S} \mu(x) - \nu(x) + \sum_{x \notin S} \nu(x) - \mu(x) = \sum_{x \in \Omega} |\mu(x) - \nu(x)|$. So,

$$= \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| = ||\mu - \nu||_{TV}.$$

For any set $S' \neq S$, $\sum_{x \in S'} \mu(x) - \nu(x) \leq \sum_{x \in S} \mu(x) - \nu(x)$. So,

$$\sum_{x \in S} \mu(x) - \nu(x) = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|.$$

$\square$

Almost uniform sampler

Let π denote the uniform distribution on a solution set $R(x)$, that is for any $w \in R(x)$, $\pi(x) = \frac{1}{|S(x)|}$.

Definition

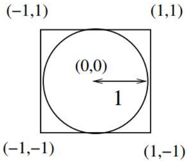
A fully polynomial almost uniform sampler (fpaus) for a solution set $R \in \Sigma^* \times \Sigma^*$ is a randomized algorithm that takes as input an instance $x \in \Sigma^*$ and a sampling tolerance $\delta > 0$ and outputs a solution $W \in R(x)$ sampled from a distribution π', such that

$$||\pi - \pi'||_{TV} \leq \delta.$$

The algorithm must run in time polynomial in $|x|$ and $\log(1/\delta)$.

The Monte Carlo method (examples)

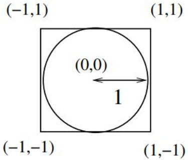
Example 1: An estimation of π



- Choose u.a.r. a point (x, y) in the unit square centered at (0,0) (choose x, y u.a.r. from the continuous distribution on $[-1, 1]$).
- Let $Z = \begin{cases} 1, & \text{if } (x, y) \in \text{ unit circle} \\ 0, & \text{otherwise} \end{cases}$

The Monte Carlo method (examples)

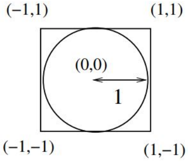
Example 1: An estimation of π



- Choose u.a.r. a point (x, y) in the unit square centered at (0,0) (choose x, y u.a.r. from the continuous distribution on $[-1, 1]$).
- Let $Z = \begin{cases} 1, & \text{if } (x, y) \in \text{ unit circle} \\ 0, & \text{otherwise} \end{cases}$
- $\Pr[Z = 1] = \frac{\text{area of the circle}}{\text{area of the square}} = \frac{\pi}{4}$, and so $\mathbb{E}[Z] = \Pr[Z = 1] = \frac{\pi}{4}$.

The Monte Carlo method (examples)

Example 1: An estimation of π

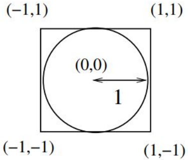


- Choose u.a.r. a point (x, y) in the unit square centered at (0,0) (choose x, y u.a.r. from the continuous distribution on $[-1, 1]$).
- Let $Z = \begin{cases} 1, & \text{if } (x, y) \in \text{ unit circle} \\ 0, & \text{otherwise} \end{cases}$

- $\Pr[Z = 1] = \frac{\text{area of the circle}}{\text{area of the square}} = \frac{\pi}{4}$, and so $\mathbb{E}[Z] = \Pr[Z = 1] = \frac{\pi}{4}$.
- We run N times and let $W = \sum_{i=1}^{N} Z_i$.

The Monte Carlo method (examples)

Example 1: An estimation of π

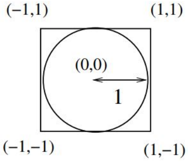


- Choose u.a.r. a point (x, y) in the unit square centered at (0,0) (choose x, y u.a.r. from the continuous distribution on $[-1, 1]$).
- Let $Z = \begin{cases} 1, & \text{if } (x,y) \in \text{ unit circle} \\ 0, & \text{otherwise} \end{cases}$

- $\Pr[Z = 1] = \frac{\text{area of the circle}}{\text{area of the square}} = \frac{\pi}{4}$, and so $\mathbb{E}[Z] = \Pr[Z = 1] = \frac{\pi}{4}$.
- We run N times and let $W = \sum_{i=1}^{N} Z_i$.
- $\mathbb{E}[W] = \sum_{i=1}^{N} \mathbb{E}[Z_i] = \frac{N\pi}{4}$ and $W' = \frac{4}{N} W$ is our estimate of π.

The Monte Carlo method (examples)

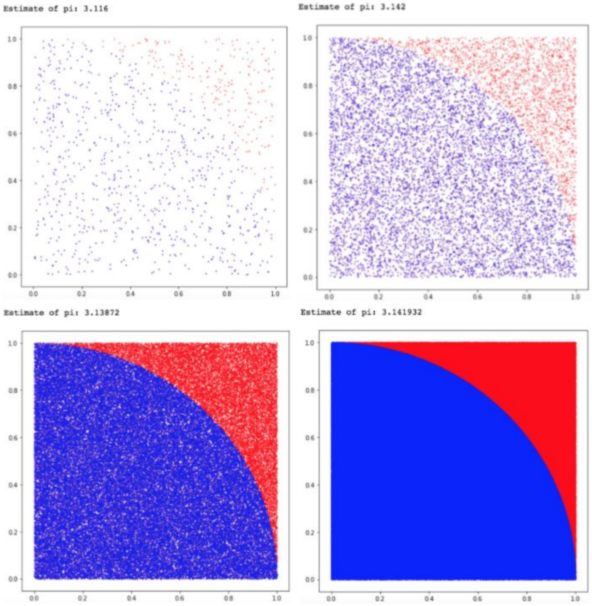
Example 1: An estimation of π



- Choose u.a.r. a point (x, y) in the unit square centered at (0,0) (choose x, y u.a.r. from the continuous distribution on $[-1, 1]$).
- Let $Z = \begin{cases} 1, & \text{if } (x, y) \in \text{ unit circle} \\ 0, & \text{otherwise} \end{cases}$
- $\Pr[Z = 1] = \frac{\text{area of the circle}}{\text{area of the square}} = \frac{\pi}{4}$, and so $\mathbb{E}[Z] = \Pr[Z = 1] = \frac{\pi}{4}$.
- We run N times and let $W = \sum_{i=1}^{N} Z_i$.
- $\mathbb{E}[W] = \sum_{i=1}^{N} \mathbb{E}[Z_i] = \frac{N\pi}{4}$ and $W' = \frac{4}{N} W$ is our estimate of π.
- By Chernoff bound $\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$, we have

$$\Pr[|W - \frac{N\pi}{4}| \geq \varepsilon \frac{N\pi}{4}] \leq 2e^{-N\pi\varepsilon^2/12}.$$

Estimating π



Example 2: An estimation of #DNF

- Let S denote the set of satisfying assignments.
- Choose N truth assignments $A^1, \ldots, A^N$ u.a.r.

Example 2: An estimation of #DNF

- Let S denote the set of satisfying assignments.
- Choose N truth assignments $A^1, ..., A^N$ u.a.r.
- Let $Y_i = \begin{cases} 1, & \text{if } A^i \text{ is satisfying} \\ 0, & \text{otherwise} \end{cases}$. Then, $\mathbb{E}[Y_i] = \frac{|S|}{2^n}$.
- Let $Y = \sum_{i=1}^{N} Y_i$.

Example 2: An estimation of #DNF

- Let S denote the set of satisfying assignments.
- Choose N truth assignments $A^1, ..., A^N$ u.a.r.
- Let $Y_i = \begin{cases} 1, & \text{if } A^i \text{ is satisfying} \\ 0, & \text{otherwise} \end{cases}$. Then, $\mathbb{E}[Y_i] = \frac{|S|}{2^n}$.
- Let $Y = \sum_{i=1}^{N} Y_i$.
- Then, $\mathbb{E}[Y] = N \cdot \frac{|S|}{2^n}$, and $Y' = \frac{2^n}{N} \cdot Y$ is our estimate of $|S|$.

Example 2: An estimation of #DNF

- Let S denote the set of satisfying assignments.
- Choose N truth assignments $A^1, ..., A^N$ u.a.r.
- Let $Y_i = \begin{cases} 1, & \text{if } A^i \text{ is satisfying} \\ 0, & \text{otherwise} \end{cases}$. Then, $\mathbb{E}[Y_i] = \frac{|S|}{2^n}$.
- Let $Y = \sum_{i=1}^N Y_i$.
- Then, $\mathbb{E}[Y] = N \cdot \frac{|S|}{2^n}$, and $Y' = \frac{2^n}{N} \cdot Y$ is our estimate of $|S|$.
- By Chernoff bound $\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$, we have

$$\Pr[|Y - N\frac{|S|}{2^n}| \geq \varepsilon N\frac{|S|}{2^n}] \leq 2e^{-N\varepsilon^2|S|/(3\cdot 2^n)}.$$

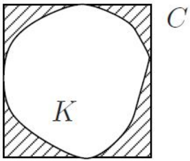
Example 2: An estimation of #DNF

- Let S denote the set of satisfying assignments.
- Choose N truth assignments $A^1, ..., A^N$ u.a.r.
- Let $Y_i = \begin{cases} 1, & \text{if } A^i \text{ is satisfying} \\ 0, & \text{otherwise} \end{cases}$. Then, $\mathbb{E}[Y_i] = \frac{|S|}{2^n}$.
- Let $Y = \sum_{i=1}^{N} Y_i$.
- Then, $\mathbb{E}[Y] = N \cdot \frac{|S|}{2^n}$, and $Y' = \frac{2^n}{N} \cdot Y$ is our estimate of $|S|$.
- By Chernoff bound $\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$, we have

$$\Pr[|Y - N\frac{|S|}{2^n}| \geq \varepsilon N\frac{|S|}{2^n}] \leq 2e^{-N\varepsilon^2|S|/(3 \cdot 2^n)}.$$

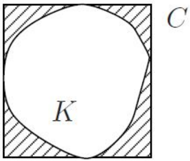
- If we want probability $\leq 2e^{-1/3}$, then $N \approx \frac{2^n}{\varepsilon^2 \cdot |S|}$.

Example 3: An estimation of the volume of a convex body



- Given K, shrink a box C around K as tightly as possible.
- Sample points $x_1, ..., x_N$ u.a.r. from C.
- In a similar way, estimate the volume of K based on the number of points that belong to K (an oracle for the membership in K is needed here).

Example 3: An estimation of the volume of a convex body



- Given K, shrink a box C around K as tightly as possible.
- Sample points $x_1, ..., x_N$ u.a.r. from C.
- In a similar way, estimate the volume of K based on the number of points that belong to K (an oracle for the membership in K is needed here).

- Let $K = B_n(0,1)$ be the unit ball and $C = [-1,1]^n$ be the smallest enclosing cube.
- $\dfrac{vol_n K}{vol_n C} = \dfrac{2\pi^{n/2}}{2^n n\Gamma(n/2)}$, which decays rapidly in n.
- In high dimensions, exponentially many points are required.

Monte Carlo method

Theorem

Let $X_1, \ldots, X_m$ be independent and identically distributed indicator variables and $\mu = \mathbb{E}[X_i]$. Then if $m \geq \frac{3\log(2/\delta)}{\varepsilon^2\mu}$, we have

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

So for this m, sampling gives an (ε, δ)-approximation of μ.

Note that if $\frac{1}{\mu}$ is polynomial in the input size, then this theorem gives an fpras for μ.

An fpras for $\#\mathrm{DNF}$

The following technique is due to Karp, Luby and Madras (1989).

- Suppose we have m sets $S_1, ..., S_m$ and we want to estimate $|\bigcup_{i=1}^{m} S_i|$.

An fpras for $\#\mathrm{DNF}$

The following technique is due to Karp, Luby and Madras (1989).

- Suppose we have m sets $S_1, \ldots, S_m$ and we want to estimate $|\bigcup_{i=1}^{m} S_i|$.

1. $|\bigcup_{i=1}^{m} S_i| \leq \sum_{i=1}^{m} |S_i|$.
2. $\sum_{i=1}^{m} |S_i| \leq m \cdot \max_{1 \leq i \leq m} |S_i| \leq m \cdot |\bigcup_{i=1}^{m} S_i|$.

An fpras for $\#\mathrm{DNF}$

The following technique is due to Karp, Luby and Madras (1989).

- Suppose we have m sets $S_1, \ldots, S_m$ and we want to estimate $|\bigcup_{i=1}^{m} S_i|$.

1. $|\bigcup_{i=1}^{m} S_i| \leq \sum_{i=1}^{m} |S_i|$.
2. $\sum_{i=1}^{m} |S_i| \leq m \cdot \max_{1 \leq i \leq m} |S_i| \leq m \cdot |\bigcup_{i=1}^{m} S_i|$.

- By 1 and 2, we have that

$$\frac{1}{m} \leq \frac{|\bigcup_{i=1}^{m} S_i|}{\sum_{i=1}^{m} |S_i|} \leq 1.$$

- Create a new universe U with $|U| = \sum_{i=1}^{m} |S_i|$.
  - For each S_i and each $a \in S_i$, add (a, i) to U.

- For every $a \in \bigcup_{i=1}^{m} S_i$, mark (a, j) as special, where j is the minimum index among all i's such that $(a, i) \in U$ (or $a \in S_i$).

$$
\begin{array}{cccccc}
 & S_1 & S_2 & S_3 & \dots & S_m \\
a_1 & \overline{*} & & * & & \\
a_2 & \overline{*} & * & & & \\
a_3 & & \overline{*} & & & \\
\vdots & & & & &
\end{array}
$$

Requirements for having an fpras for $|\bigcup_{i=1}^{m} S_i|$.

1. Sample an element efficiently from U.
2. Determine efficiently if any $(a, i) \in U$ is marked.
3. Calculate efficiently $|U|$.

Requirements for having an fpras for $|\bigcup_{i=1}^{m} S_i|$.

1. Sample an element efficiently from U.
2. Determine efficiently if any $(a, i) \in U$ is marked.
3. Calculate efficiently $|U|$.

Then, the following steps describe the fpras for $|\bigcup_{i=1}^{m} S_i|$.

1. Sample elements $e_1, \ldots, e_N$ from U.

Requirements for having an fpras for $|\bigcup_{i=1}^{m} S_i|$.

1. Sample an element efficiently from U.
2. Determine efficiently if any $(a, i) \in U$ is marked.
3. Calculate efficiently $|U|$.

Then, the following steps describe the fpras for $|\bigcup_{i=1}^{m} S_i|$.

1. Sample elements $e_1, ..., e_N$ from U.
2. Let $X_i = \begin{cases} 1, & \text{if } e_i \text{ is marked} \\ 0, & \text{otherwise} \end{cases}$.
   - Let $X = \sum_{i=1}^{N} X_i$. Then, $\mathbb{E}[X] = N \cdot \frac{|\bigcup_{i=1}^{m} S_i|}{|U|}$.

Requirements for having an fpras for $|\bigcup_{i=1}^{m} S_i|$.

1. Sample an element efficiently from U.
2. Determine efficiently if any $(a, i) \in U$ is marked.
3. Calculate efficiently $|U|$.

Then, the following steps describe the fpras for $|\bigcup_{i=1}^{m} S_i|$.

1. Sample elements $e_1, \ldots, e_N$ from U.
2. Let $X_i = \begin{cases} 1, & \text{if } e_i \text{ is marked} \\ 0, & \text{otherwise} \end{cases}$.
   - Let $X = \sum_{i=1}^{N} X_i$. Then, $\mathbb{E}[X] = N \cdot \frac{|\bigcup_{i=1}^{m} S_i|}{|U|}$.
3. $X' = \frac{|U|}{N} \cdot X$ is our estimate of $|\bigcup_{i=1}^{m} S_i|$.

Requirements for having an fpras for $|\bigcup_{i=1}^{m} S_i|$.

1. Sample an element efficiently from U.
2. Determine efficiently if any $(a, i) \in U$ is marked.
3. Calculate efficiently $|U|$.

Then, the following steps describe the fpras for $|\bigcup_{i=1}^{m} S_i|$.

1. Sample elements $e_1, ..., e_N$ from U.
2. Let $X_i = \begin{cases} 1, & \text{if } e_i \text{ is marked} \\ 0, & \text{otherwise} \end{cases}$.
   - Let $X = \sum_{i=1}^{N} X_i$. Then, $\mathbb{E}[X] = N \cdot \frac{|\bigcup_{i=1}^{m} S_i|}{|U|}$.
3. $X' = \frac{|U|}{N} \cdot X$ is our estimate of $|\bigcup_{i=1}^{m} S_i|$.
4. For an (ε, δ) approximation of μ, $N \geq \frac{3m \log(2/\delta)}{\varepsilon^2}$, since $\frac{1}{\mu} \leq m$.

Application to #DNF

- S_i is the set of satisfying assignments of the i-th clause. So, m is polynomial in the input size (step 4).
- $|S_i| = \#(\text{satisfying assignments of the } i^{th} \text{ clause})$, and $|U| = \sum_{i=1}^{m} |S_i|$ (requirement 3).
- Given an element $(a, i) \in U$, we can determine in polynomial time whether the i-th clause is the first one satisfied by truth assignment a (requirement 2).

Application to $\#\mathrm{DNF}$

We can sample an (a, i) u.a.r. from U as follows (requirement 1).

1. Calculate $|S_i|$, for every $1 \leq i \leq m$.
2. Choose i with probability $\dfrac{|S_i|}{\sum_{i=1}^{m} |S_i|}$.
3. Choose a satisfying assignment a of the i-th clause u.a.r.
   In other words, with probability $\frac{1}{|S_i|}$.

Then, (a, i) has been chosen with probability $\dfrac{1}{\sum_{i=1}^{m} |S_i|} = \dfrac{1}{|U|}$. $\square$

Counting versus Sampling

For any self-reducible problem,

- counting and sampling are closely related as shown below.

$$\begin{array}{ccc} \text{Exact Counter} & \Rightarrow & \text{Exact Sampler} \\ \Downarrow & & \Downarrow \\ \text{Approximate Counter} & \Leftrightarrow & \text{Approximate Sampler} \end{array}$$

Counting versus Sampling

For any self-reducible problem,

- counting and sampling are closely related as shown below.

$$\begin{array}{ccc} \text{Exact Counter} & \Rightarrow & \text{Exact Sampler} \\ \Downarrow & & \Downarrow \\ \text{Approximate Counter} & \Leftrightarrow & \text{Approximate Sampler} \end{array}$$

- if there exists a polynomial-time randomized algorithm for counting within a polynomial factor, then there exists an fpras.

Self-reducible problems

Example: Consider $\textsc{Sat}$ and let ϕ be a CNF fomrula. Then,

$$S(\phi) = S(\phi \restriction_{x_1=0}) \cup S(\phi \restriction_{x_1=1})$$

where $\phi \restriction_{x_1=0}$ (resp. $\phi \restriction_{x_1=1}$) is ϕ after setting the variable x_1 to false (resp. true).

Self-reducible problems

Example: Consider $\textsc{Sat}$ and let ϕ be a CNF fomrula. Then,

$$S(\phi) = S(\phi \restriction_{x_1=0}) \cup S(\phi \restriction_{x_1=1})$$

where $\phi \restriction_{x_1=0}$ (resp. $\phi \restriction_{x_1=1}$) is ϕ after setting the variable x_1 to false (resp. true).

Definition

An NP problem is self-reducible if the set of solutions can be partitioned into polynomially many sets each of which is the set of solutions of a smaller instance of the problem.
Also, these smaller instances are efficiently computable.

Approximate sampler $\Rightarrow$ Approximate counting

- We prove that an fpaus implies an fpras in the context of a specific combinatorial structure, namely matchings in a graph.

Approximate sampler $\Rightarrow$ Approximate counting

- We prove that an fpaus implies an fpras in the context of a specific combinatorial structure, namely matchings in a graph.
- Let $\mathcal{M}(G)$ denote the set of matchings of all sizes in a graph G.

Approximate sampler $\Rightarrow$ Approximate counting

- We prove that an fpaus implies an fpras in the context of a specific combinatorial structure, namely matchings in a graph.
- Let $\mathcal{M}(G)$ denote the set of matchings of all sizes in a graph G.
- Let G be a graph with n vertices and m edges, where $m \geq 1$ (to avoid trivialities).

Approximate sampler $\Rightarrow$ Approximate counting

- We prove that an fpaus implies an fpras in the context of a specific combinatorial structure, namely matchings in a graph.
- Let $\mathcal{M}(G)$ denote the set of matchings of all sizes in a graph G.
- Let G be a graph with n vertices and m edges, where $m \geq 1$ (to avoid trivialities).

Proposition

If there is an almost uniform sampler for $\mathcal{M}(G)$ with run-time bounded by $T(n, m, \varepsilon)$, then there is a randomized approximation scheme for $|\mathcal{M}(G)|$ with run-time bounded by $cm^2\varepsilon^{-2}T(n, m, \varepsilon/6m)$ for some constant c.
In particular,

$$\text{fpaus for } \mathcal{M}(G) \Rightarrow \text{ fpras for } |\mathcal{M}(G)|.$$

Proof.

- Let $\mathcal{S}$ denote the almost uniform sampler.

Proof.

- Let $\mathcal{S}$ denote the almost uniform sampler.
- Given G with $E(G) = \{e_1, ..., e_m\}$, we consider the graphs

$$G_i := (V(G), \{e_1, ..., e_i\}), 0 \leq i \leq m.$$

In particular, G_0 has no edge and $G_m = G$.

Proof.

- Let $\mathcal{S}$ denote the almost uniform sampler.

- Given G with $E(G) = \{e_1, ..., e_m\}$, we consider the graphs

$$G_i := (V(G), \{e_1, ..., e_i\}), 0 \leq i \leq m.$$

  In particular, G_0 has no edge and $G_m = G$.

- Then,

$$|\mathcal{M}(G)| = \Big(\frac{|\mathcal{M}(G_0)|}{|\mathcal{M}(G_1)|} \cdot \frac{|\mathcal{M}(G_1)|}{|\mathcal{M}(G_2)|} \cdots \frac{|\mathcal{M}(G_{m-1})|}{|\mathcal{M}(G_m)|}\Big)^{-1}.$$

  where we consider $|\mathcal{M}(G_0)| = 1$.

Proof.

- Let $\mathcal{S}$ denote the almost uniform sampler.
- Given G with $E(G) = \{e_1, ..., e_m\}$, we consider the graphs

$$G_i := (V(G), \{e_1, ..., e_i\}), 0 \leq i \leq m.$$

  In particular, G_0 has no edge and $G_m = G$.

- Then,

$$|\mathcal{M}(G)| = \Big(\frac{|\mathcal{M}(G_0)|}{|\mathcal{M}(G_1)|} \cdot \frac{|\mathcal{M}(G_1)|}{|\mathcal{M}(G_2)|} \cdots \frac{|\mathcal{M}(G_{m-1})|}{|\mathcal{M}(G_m)|}\Big)^{-1}.$$

  where we consider $|\mathcal{M}(G_0)| = 1$.

- Let ρ_i denote the i-th ratio $\dfrac{|\mathcal{M}(G_{i-1})|}{|\mathcal{M}(G_i)|}$.

Proof cont.

1. $\mathcal{M}(G_i)$ contains all matchings in $\mathcal{M}(G_{i-1})$.

Proof cont.

1. $\mathcal{M}(G_i)$ contains all matchings in $\mathcal{M}(G_{i-1})$.

2. Also, the size of $\mathcal{M}(G_i)$ is at most twice the size of $\mathcal{M}(G_{i-1})$.
   - It holds $|\mathcal{M}(G_i)| = 2 \cdot |\mathcal{M}(G_{i-1})|$ if every $M \in \mathcal{M}(G_i)$ can be extended to an $M' = M \cup \{e_i\}$ in $\mathcal{M}(G_i)$.

Proof cont.

1. $\mathcal{M}(G_i)$ contains all matchings in $\mathcal{M}(G_{i-1})$.

2. Also, the size of $\mathcal{M}(G_i)$ is at most twice the size of $\mathcal{M}(G_{i-1})$.
   - It holds $|\mathcal{M}(G_i)| = 2 \cdot |\mathcal{M}(G_{i-1})|$ if every $M \in \mathcal{M}(G_i)$ can be extended to an $M' = M \cup \{e_i\}$ in $\mathcal{M}(G_i)$.

By 1 and 2,

$$\frac{1}{2} \leq \frac{|\mathcal{M}(G_{i-1})|}{|\mathcal{M}(G_i)|} \leq 1.$$

Proof cont.

- We want to have an ε-approximation of $|\mathcal{M}(G)|$ with prob. $\geq \frac{3}{4}$.

Proof cont.

- We want to have an ε-approximation of $|\mathcal{M}(G)|$ with prob. $\geq \frac{3}{4}$.
- The idea is to use the sampler $\mathcal{S}$ to approximate every ratio ρ_i.

Proof cont.

- We want to have an ε-approximation of $|\mathcal{M}(G)|$ with prob. $\geq \frac{3}{4}$.
- The idea is to use the sampler $\mathcal{S}$ to approximate every ratio ρ_i.
- We run our sampler $\mathcal{S}$ on G_i with $\delta = \frac{\varepsilon}{6m}$ and obtain a matching $M_i \in \mathcal{M}(G_i)$ sampled from μ.

Proof cont.

- We want to have an ε-approximation of $|\mathcal{M}(G)|$ with prob. $\geq \frac{3}{4}$.
- The idea is to use the sampler $\mathcal{S}$ to approximate every ratio ρ_i.
- We run our sampler $\mathcal{S}$ on G_i with $\delta = \frac{\varepsilon}{6m}$ and obtain a matching $M_i \in \mathcal{M}(G_i)$ sampled from μ.
- Let π denote the uniform distribution on $\mathcal{M}(G_i)$.

Proof cont.

- We want to have an ε-approximation of $|\mathcal{M}(G)|$ with prob. $\geq \frac{3}{4}$.
- The idea is to use the sampler $\mathcal{S}$ to approximate every ratio ρ_i.
- We run our sampler $\mathcal{S}$ on G_i with $\delta = \frac{\varepsilon}{6m}$ and obtain a matching $M_i \in \mathcal{M}(G_i)$ sampled from μ.
- Let π denote the uniform distribution on $\mathcal{M}(G_i)$.
- Let $Z_i = \begin{cases} 1, & \text{if } M_i \in \mathcal{M}(G_{i-1}) \\ 0, & \text{otherwise} \end{cases}$, and set $\mu_i = \mathbb{E}(Z_i) = \Pr[Z_i = 1]$.

Proof cont.

- We want to have an ε-approximation of $|\mathcal{M}(G)|$ with prob. $\geq \frac{3}{4}$.
- The idea is to use the sampler $\mathcal{S}$ to approximate every ratio ρ_i.
- We run our sampler $\mathcal{S}$ on G_i with $\delta = \frac{\varepsilon}{6m}$ and obtain a matching $M_i \in \mathcal{M}(G_i)$ sampled from μ.
- Let π denote the uniform distribution on $\mathcal{M}(G_i)$.
- Let $Z_i = \begin{cases} 1, & \text{if } M_i \in \mathcal{M}(G_{i-1}) \\ 0, & \text{otherwise} \end{cases}$, and set $\mu_i = \mathbb{E}(Z_i) = \Pr[Z_i = 1]$.
- How close is μ_i to $\dfrac{|\mathcal{M}(G_{i-1})|}{|\mathcal{M}(G_i)|}$ (or how close is μ_i to ρ_i)?

Proof cont.

Let $A = \{M \mid M \in \mathcal{M}(G_{i-1})\}$.
By definition of the TV distance $||\mu - \nu||_{TV} = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|$:

$$|\mu(A) - \pi(A)| \leq \frac{\varepsilon}{6m} \Leftrightarrow |\sum_{M \in A} \mu(M) - \sum_{M \in A} \pi(M)| \leq \frac{\varepsilon}{6m} \Leftrightarrow$$

$$|\Pr_{M \sim \mu}[M \in A] - \Pr_{M \sim \pi}[M \in A]| \leq \frac{\varepsilon}{6m} \Leftrightarrow |\mu_i - \rho_i| \leq \frac{\varepsilon}{6m} \Leftrightarrow$$

$$\rho_i - \frac{\varepsilon}{6m} \leq \mu_i \leq \rho_i + \frac{\varepsilon}{6m} \Leftrightarrow \qquad \frac{1}{2} \leq \rho_i \leq 1$$

$$\rho_i - \frac{\varepsilon \cdot \frac{1}{2}}{3m} \leq \mu_i \leq \rho_i + \frac{\varepsilon \cdot \frac{1}{2}}{3m} \Leftrightarrow$$

$$\left(1 - \frac{\varepsilon}{3m}\right)\rho_i \leq \mu_i \leq \left(1 - \frac{\varepsilon}{3m}\right)\rho_i$$

Proof cont.

Let $A = \{M \mid M \in \mathcal{M}(G_{i-1})\}$.
By definition of the TV distance $||\mu - \nu||_{TV} = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|$:

$$|\mu(A) - \pi(A)| \leq \frac{\varepsilon}{6m} \Leftrightarrow |\sum_{M \in A} \mu(M) - \sum_{M \in A} \pi(M)| \leq \frac{\varepsilon}{6m} \Leftrightarrow$$

$$|\Pr_{M \sim \mu}[M \in A] - \Pr_{M \sim \pi}[M \in A]| \leq \frac{\varepsilon}{6m} \Leftrightarrow |\mu_i - \rho_i| \leq \frac{\varepsilon}{6m} \Leftrightarrow$$

$$\rho_i - \frac{\varepsilon}{6m} \leq \mu_i \leq \rho_i + \frac{\varepsilon}{6m} \Leftrightarrow \qquad \frac{1}{2} \leq \rho_i \leq 1$$

$$\rho_i - \frac{\varepsilon \cdot \frac{1}{2}}{3m} \leq \mu_i \leq \rho_i + \frac{\varepsilon \cdot \frac{1}{2}}{3m} \Leftrightarrow$$

$$\left(1 - \frac{\varepsilon}{3m}\right)\rho_i \leq \mu_i \leq \left(1 - \frac{\varepsilon}{3m}\right)\rho_i$$

So, μ_i is an $\frac{\varepsilon}{3m}$-approximation of ρ_i.

Proof cont.

- So we need a good estimate of μ_i.

Proof cont.

- So we need a good estimate of μ_i.
- $\mu_i \geq \left(1 - \frac{\varepsilon}{3m}\right)\rho_i \geq \left(1 - \frac{\varepsilon}{3m}\right)\frac{1}{2} \geq \frac{1}{3}$ (using $\varepsilon \leq m$) (which also implies that $\frac{1}{\mu_i} \leq 3$).

Proof cont.

- So we need a good estimate of μ_i.
- $\mu_i \geq \left(1 - \frac{\varepsilon}{3m}\right)\rho_i \geq \left(1 - \frac{\varepsilon}{3m}\right)\frac{1}{2} \geq \frac{1}{3}$ (using $\varepsilon \leq m$) (which also implies that $\frac{1}{\mu_i} \leq 3$).
- $Var(Z_i) = \mathbb{E}[(Z_i - \mu_i)^2] = \Pr[Z_i = 1](1 - \mu_i)^2 + \Pr[Z_i = 0]\mu_i^2 = \mu_i(1 - \mu_i)$.
- $\frac{Var(Z_i)}{\mu_i^2} = \frac{\mu_i(1-\mu_i)}{\mu_i^2} = \frac{\mu_i - \mu_i^2}{\mu_i^2} = \frac{1}{\mu_i} - 1 \leq 2$.

Proof cont.

- So we need a good estimate of μ_i.
- $\mu_i \geq \left(1 - \frac{\varepsilon}{3m}\right)\rho_i \geq \left(1 - \frac{\varepsilon}{3m}\right)\frac{1}{2} \geq \frac{1}{3}$ (using $\varepsilon \leq m$) (which also implies that $\frac{1}{\mu_i} \leq 3$).
- $Var(Z_i) = \mathbb{E}[(Z_i - \mu_i)^2] = \Pr[Z_i = 1](1 - \mu_i)^2 + \Pr[Z_i = 0]\mu_i^2 = \mu_i(1 - \mu_i)$.
- $\frac{Var(Z_i)}{\mu_i^2} = \frac{\mu_i(1-\mu_i)}{\mu_i^2} = \frac{\mu_i - \mu_i^2}{\mu_i^2} = \frac{1}{\mu_i} - 1 \leq 2$.
- If we take the outputs $Z_i^{(1)}, \ldots, Z_i^{(s)}$ of s independent runs of $\mathcal{S}$ on G_i, and set $\overline{Z}_i := \frac{\sum_{j=1}^{s} Z_i^{(j)}}{s}$, then $\mathbb{E}[\overline{Z}_i] = \mu_i$ and

$$\frac{Var(\overline{Z}_i)}{\mu_i^2} = \frac{\frac{1}{s^2}\sum_{j=1}^{s} Var(Z_i^{(j)})}{\mu_i^2} \leq \frac{2}{s}.$$

Proof cont.

- $\mathbb{E}[\overline{Z}_i] = \mu_i$ and $\frac{Var(\overline{Z}_i)}{\mu_i^2} \leq \frac{2}{s}$.

Proof cont.

- $\mathbb{E}[\overline{Z}_i] = \mu_i$ and $\frac{Var(\overline{Z}_i)}{\mu_i^2} \leq \frac{2}{s}$.
- Let $s := \lceil 74\varepsilon^{-2}m \rceil$.
- Then, $\frac{Var(\overline{Z}_i)}{\mu_i^2} \leq \frac{2}{s} \leq \frac{\varepsilon^2}{37m}$.

Proof cont.

- $\mathbb{E}[\overline{Z}_i] = \mu_i$ and $\frac{Var(\overline{Z}_i)}{\mu_i^2} \leq \frac{2}{s}$.
- Let $s := \lceil 74\varepsilon^{-2} m \rceil$.
- Then, $\frac{Var(\overline{Z}_i)}{\mu_i^2} \leq \frac{2}{s} \leq \frac{\varepsilon^2}{37m}$.
- Our estimator for $|\mathcal{M}(G)|$ is the random variable

$$N := \Big(\prod_{i=1}^{m} \overline{Z}_i\Big)^{-1}.$$

Proof cont.

- $\mathbb{E}[\overline{Z}_i] = \mu_i$ and $\frac{Var(\overline{Z}_i)}{\mu_i^2} \leq \frac{2}{s}$.

- Let $s := \lceil 74\varepsilon^{-2} m \rceil$.

- Then, $\frac{Var(\overline{Z}_i)}{\mu_i^2} \leq \frac{2}{s} \leq \frac{\varepsilon^2}{37m}$.

- Our estimator for $|\mathcal{M}(G)|$ is the random variable

$$N := \left(\prod_{i=1}^{m} \overline{Z}_i \right)^{-1}.$$

- $\mathbb{E}[\overline{Z}_1 \cdots \overline{Z}_m] = \mu_1 \cdots \mu_m$.

Proof cont.

$$
\begin{aligned}
\frac{Var(\overline{Z}_1 \cdots \overline{Z}_m)}{(\mu_1 \cdots \mu_m)^2} &= \frac{\mathbb{E}[\overline{Z}_1^2 \cdots \overline{Z}_m^2]}{\mu_1^2 \cdots \mu_m^2} - 1 && \text{since } Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \prod_{i=1}^{m} \frac{\mathbb{E}[\overline{Z}_i^2]}{\mu_i^2} - 1 && \text{since } \overline{Z}_i \text{ are independent} \\
&= \prod_{i=1}^{m} \left(1 + \frac{var(\overline{Z}_i)}{\mu_i^2}\right) - 1 && \text{since } \mathbb{E}[X^2] = Var(\overline{Z}_i) + \mathbb{E}[X]^2 \\
&\leq \left(1 + \frac{\varepsilon^2}{37m}\right)^m - 1 && \text{since } \frac{Var(\overline{Z}_i)}{\mu_i^2} \leq \frac{\varepsilon^2}{37m} \\
&\leq \exp\left(\frac{\varepsilon^2}{37}\right) - 1 && \text{since } (1 + \frac{x}{k})^k \leq e^x \\
&\leq \frac{\varepsilon^2}{36} && \text{since } e^{x/(k+1)} \leq 1 + x/k \text{ for } 0 \leq x \leq 1
\end{aligned}
$$

Proof cont.

By Chebychev's inequality $\Pr[|X - \mathbb{E}(X)| \leq a) \geq 1 - \frac{Var(X)}{a^2}$, we have that

$$\Pr[|\overline{Z}_1 \cdots \overline{Z}_m - \mu_1 \cdots \mu_m| \leq \frac{\varepsilon}{3}\mu_1 \cdots \mu_m] \geq 1 - \frac{\frac{\varepsilon^2}{36}(\mu_1 \cdots \mu_m)^2}{\frac{\varepsilon^2}{9}(\mu_1 \cdots \mu_m)^2} \Leftrightarrow$$

Proof cont.

By Chebychev's inequality $\Pr[|X - \mathbb{E}(X)| \leq a) \geq 1 - \frac{Var(X)}{a^2}$, we have that

$$\Pr[|\overline{Z}_1 \cdots \overline{Z}_m - \mu_1 \cdots \mu_m| \leq \frac{\varepsilon}{3}\mu_1 \cdots \mu_m] \geq 1 - \frac{\frac{\varepsilon^2}{36}(\mu_1 \cdots \mu_m)^2}{\frac{\varepsilon^2}{9}(\mu_1 \cdots \mu_m)^2} \Leftrightarrow$$

$$\Pr[|\overline{Z}_1 \cdots \overline{Z}_m - \mu_1 \cdots \mu_m| \leq \frac{\varepsilon}{3}\mu_1 \cdots \mu_m] \geq \frac{3}{4} \Leftrightarrow$$

$$\left(1 - \frac{\varepsilon}{3}\right)\mu_1 \cdots \mu_m \leq \overline{Z}_1 \cdots \overline{Z}_m \leq \left(1 + \frac{\varepsilon}{3}\right)\mu_1 \cdots \mu_m \quad \text{with prob. } \geq \frac{3}{4} \Leftrightarrow$$

Proof cont.

By Chebychev's inequality $\Pr[|X-\mathbb{E}(X)| \leq a) \geq 1-\frac{Var(X)}{a^2}$, we have that

$$\Pr[|\overline{Z}_1\cdots\overline{Z}_m-\mu_1\cdots\mu_m| \leq \frac{\varepsilon}{3}\mu_1\cdots\mu_m] \geq 1-\frac{\frac{\varepsilon^2}{36}(\mu_1\cdots\mu_m)^2}{\frac{\varepsilon^2}{9}(\mu_1\cdots\mu_m)^2} \Leftrightarrow$$

$$\Pr[|\overline{Z}_1\cdots\overline{Z}_m-\mu_1\cdots\mu_m| \leq \frac{\varepsilon}{3}\mu_1\cdots\mu_m] \geq \frac{3}{4} \Leftrightarrow$$

$$\left(1-\frac{\varepsilon}{3}\right)\mu_1\cdots\mu_m \leq \overline{Z}_1\cdots\overline{Z}_m \leq \left(1+\frac{\varepsilon}{3}\right)\mu_1\cdots\mu_m \quad \text{with prob. } \geq \frac{3}{4} \Leftrightarrow$$

$$e^{-\varepsilon/2}\mu_1\cdots\mu_m \leq \overline{Z}_1\cdots\overline{Z}_m \leq e^{\varepsilon/2}\mu_1\cdots\mu_m \quad \text{with prob. } \geq \frac{3}{4} \ (1)$$

using $1+x \leq e^x$ and $e^{-x/k} \leq 1-x/(k+1)$ for $0 \leq x \leq 1$.

Proof cont.

By $\left(1-\frac{\varepsilon}{3m}\right)\rho_i \leq \mu_i \leq \left(1-\frac{\varepsilon}{3m}\right)\rho_i$ and similar calculations, we obtain that

$$e^{-\varepsilon/2}\rho_1 \cdots \rho_m \leq \mu_1 \cdots \mu_m \leq e^{\varepsilon/2}\rho_1 \cdots \rho_m \quad (2)$$

By (1) and (2), we have that

$$e^{-\varepsilon}\rho_1 \cdots \rho_m \leq \overline{Z}_1 \cdots \overline{Z}_m \leq e^{\varepsilon}\rho_1 \cdots \rho_m \quad \text{with prob. } \geq \frac{3}{4} \Leftrightarrow$$

$$e^{-\varepsilon}(\rho_1 \cdots \rho_m)^{-1} \leq (\overline{Z}_1 \cdots \overline{Z}_m)^{-1} \leq e^{\varepsilon}(\rho_1 \cdots \rho_m)^{-1} \quad \text{with prob. } \geq \frac{3}{4}$$

$$e^{-\varepsilon}|\mathcal{M}(G)| \leq \text{ output } \leq e^{\varepsilon}|\mathcal{M}(G)| \quad \text{with prob. } \geq \frac{3}{4}$$

Proof cont.

The run-time of the algorithm is bounded by

$$\begin{aligned}
\text{(number of samples)} \cdot \text{(time per sample)} &= \\
sm \cdot T(n, m, \frac{\varepsilon}{6m}) &\leq \\
75\varepsilon^{-2}m^2 \cdot T(n, m, \frac{\varepsilon}{6m})
\end{aligned}$$

$\square$

Overview

- We deal with discrete-time Markov chains on a finite state space Ω.

- We deal with discrete-time Markov chains on a finite state space Ω.
- A sequence $\{X_t \in \Omega\}_{t=0}^{\infty}$ of random variables is a Markov chain (MC), with state space Ω, if

$$\Pr[X_{t+1} = y \mid X_t = x_t, \ldots, X_0 = x_0] = \Pr[X_{t+1} = y \mid X_t = x_t]$$

for all $t \in \mathbb{N}$ and all $x_0, \ldots, x_t \in \Omega$.

- We deal with discrete-time Markov chains on a finite state space Ω.
- A sequence $\{X_t \in \Omega\}_{t=0}^{\infty}$ of random variables is a Markov chain (MC), with state space Ω, if

$$\Pr[X_{t+1} = y \mid X_t = x_t, ..., X_0 = x_0] = \Pr[X_{t+1} = y \mid X_t = x_t]$$

  for all $t \in \mathbb{N}$ and all $x_0, ..., x_t \in \Omega$.
- This is called the Markovian property.

- We deal with discrete-time Markov chains on a finite state space Ω.
- A sequence $\{X_t \in \Omega\}_{t=0}^{\infty}$ of random variables is a Markov chain (MC), with state space Ω, if

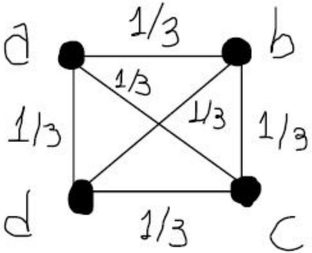
$$\Pr[X_{t+1} = y \mid X_t = x_t, ..., X_0 = x_0] = \Pr[X_{t+1} = y \mid X_t = x_t]$$

  for all $t \in \mathbb{N}$ and all $x_0, ..., x_t \in \Omega$.
- This is called the Markovian property.
- Time-homogeneous MCs are the ones for which the probability $\Pr[X_{t+1} = y \mid X_t = x]$ does not depend on t. In this case we write

$$P(x, y) = \Pr[X_{t+1} = y \mid X_t = x]$$

  where P is the transition matrix of the MC.

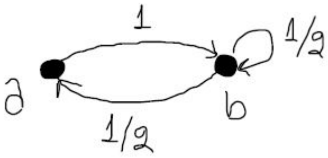
Example 1



$$P = \begin{array}{c c} & \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \end{array} & \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix} \end{array}$$

$$X_0 = a, X_1 = b, X_2 = d, X_3 = b, \ldots$$

Example 2



$$P = \begin{array}{c c} & \begin{array}{cc} a & b \end{array} \\ \begin{array}{c} a \\ b \end{array} & \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix} \end{array}$$

$$X_0 = a, X_1 = b, X_2 = b, X_3 = b, X_4 = a, X_5 = b, \ldots$$

Transition matrix

- Each row of the transition matrix P is a distribution.

Transition matrix

- Each row of the transition matrix P is a distribution.
- P describes single-step transition probabilities.

Transition matrix

- Each row of the transition matrix P is a distribution.
- P describes single-step transition probabilities.
- The t-step transition probabilities are given inductively by

$$P^t(x,y) := \begin{cases} I(x,y), & \text{if } t = 0 \\ \sum_{y' \in \Omega} P^{t-1}(x,y')P(y',y), & \text{if } t > 0 \end{cases}$$

Transition matrix

- Each row of the transition matrix P is a distribution.
- P describes single-step transition probabilities.
- The t-step transition probabilities are given inductively by

$$P^t(x,y) := \begin{cases} I(x,y), & \text{if } t = 0 \\ \sum_{y' \in \Omega} P^{t-1}(x,y')P(y',y), & \text{if } t > 0 \end{cases}$$

$$\begin{array}{c} \\ y_1 \\ y_2 \\ \vdots \\ x \\ \vdots \\ y_n \end{array} \begin{array}{c} \begin{array}{cccc} y_1 & y_2 & \cdots & y_n \end{array} \\ \left[\begin{array}{cccc} & & & \\ & & & \\ & & & \\ P(x,y_1) & P(x,y_2) & \cdots & P(x,y_n) \\ & & & \\ & & & \end{array}\right] \end{array} \cdot \begin{array}{c} \\ y_1 \\ y_2 \\ \\ \\ y_n \end{array} \begin{array}{c} \begin{array}{ccccc} y_1 & y_2 & \cdots\ y & \cdots & y_n \end{array} \\ \left[\begin{array}{ccc} & P(y_1,y) & \\ & P(y_2,y) & \\ & \vdots & \\ & P(y_n,y) & \end{array}\right] \end{array} =$$

$$\begin{array}{c} \\ \\ x \\ \\ \end{array} \begin{array}{c} y \\ \left[\begin{array}{ccc} \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ & P^2(x,y) & \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{array}\right] \end{array} \qquad \boxed{P^t = P^{t-1} \cdot P}$$

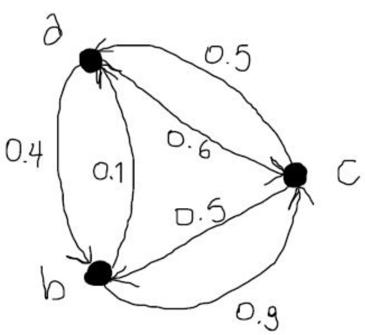
Transition matrix

- Each row of the transition matrix P is a distribution.
- P describes single-step transition probabilities.
- The t-step transition probabilities are given inductively by

$$P^t(x, y) := \begin{cases} I(x, y), & \text{if } t = 0 \\ \sum_{y' \in \Omega} P^{t-1}(x, y')P(y', y), & \text{if } t > 0 \end{cases}$$

- So P^t describes t-step transition probabilities.

Example 3



$$P = \begin{array}{c c} & \begin{array}{ccc} a & b & c \end{array} \\ \begin{array}{c} a \\ b \\ c \end{array} & \begin{bmatrix} 0 & 0.4 & 0.6 \\ 0.1 & 0 & 0.9 \\ 0.5 & 0.5 & 0 \end{bmatrix} \end{array}$$

Start distribution: $\sigma_0 = (1, 0, 0)$ (start in a)
After one step: $\sigma_1 = \sigma_0 P = (0, 0.4, 0.6)$
After two steps: $\sigma_2 = \sigma_1 P = \sigma_0 P^2 = (0.34, 0.3, 0.36)$
After t steps: $\sigma_t = \sigma_{t-1} P = \sigma_0 P^t$

Stationary distribution

$$\begin{bmatrix} \pi(a) & \pi(b) & \pi(c) & \pi(d) \end{bmatrix} \begin{bmatrix} P(a,a) & P(a,b) & P(a,c) & P(a,d) \\ P(b,a) & P(b,b) & P(b,c) & P(b,d) \\ P(c,a) & P(c,b) & P(c,c) & P(c,d) \\ P(d,a) & P(d,b) & P(d,c) & P(d,d) \end{bmatrix} =$$

$$\begin{bmatrix} \pi(a) & \pi(b) & \pi(c) & \pi(d) \end{bmatrix}$$

$$\pi(a) = \pi(a)P(a,a) + \pi(b)P(b,a) + \pi(c)P(c,a) + \pi(d)P(d,a)$$

Stationary distribution

- A stationary distribution of an MC with transition matrix P is a distribution $\pi : \Omega \to [0, 1]$ such that

$$\pi(y) = \sum_{x \in \Omega} \pi(x) P(x, y)$$

- In other words, $\pi \cdot P = \pi$.

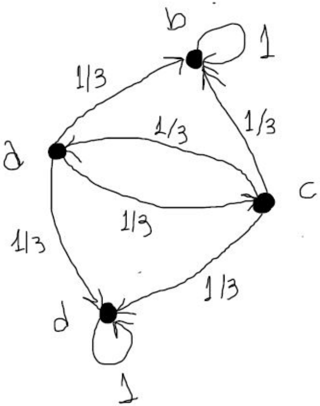
Definition of irreducibility

Definition

An MC is irreducible if for all $x, y \in \Omega$, there exists a $t > 0$, such that $P^t(x, y) > 0$ (there exists a path in the transition graph from every state to every other state).

Example 4

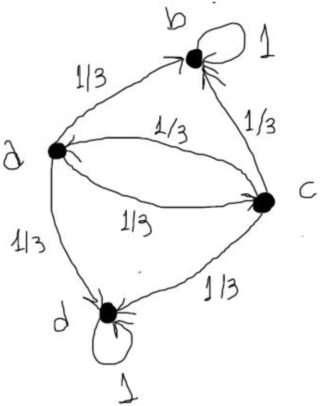
Not irreducible



$$P = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Example 4

Not irreducible



$$P = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Stationary distributions: $\pi_1 = (0, 1, 0, 0)$, $\pi_2 = (0, 0, 0, 1)$, $\pi_3 = (0, 0.5, 0, 0.5)$, ...

Definition of aperiodicity

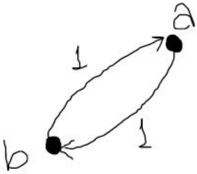
Definition

An MC is aperiodic if $\gcd\{t \mid P^t(x,x) > 0\} = 1$ for all $x \in \Omega$ (for each state x, the gcd of all walk lengths from x to x is 1).

In the case of an irreducible MC, it is sufficient to verify the condition $\gcd\{t \mid P^t(x,x) > 0\} = 1$ for just one state $x \in \Omega$.

Example 5

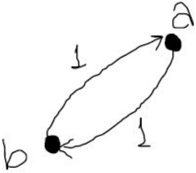
Not aperiodic



$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

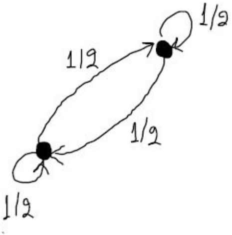
Example 5

Not aperiodic



$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Lazy MC (a self-loop at every state)



$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

Overview

Useful elements of probability theory

- $\mathbb{E}[X] = \sum_i x_i \cdot P(X = x_i)$.
- $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
- Chebychev's Inequality: $\Pr[|X - \mathbb{E}(X)| \geq a] \leq \frac{Var(X)}{a^2}$.
  - In particular, $\Pr[|X - \mathbb{E}(X)| \geq a\mathbb{E}(X)] \leq \frac{Var(X)}{a^2 \mathbb{E}(X)^2}$.
- Chernoff bound: $\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$ for all $0 < \delta < 1$, where $X = \sum_{i=1}^n X_i$, $X_i = \begin{cases} 1, & \text{with prob. } p_i \\ 0, & \text{with prob. } 1 - p_i \end{cases}$, all X_i are independent and $\mu = \mathbb{E}[X] = \sum_{i=1}^n p_i$.

Useful inequalities

1. $1 + x \leq e^x$.
2. $(1 + \frac{x}{k})^k \leq e^x$.
3. $e^{x/(k+1)} \leq 1 + x/k$ for $0 \leq x \leq 1$ and $k \in \mathbb{N}^+$.
4. $e^{-x/k} \leq 1 - x/(k+1)$ for $0 \leq x \leq 1$ and $k \in \mathbb{N}^+$.
5. $e^{-\frac{x}{k}} \leq \left(1 - \frac{x}{(k+1)n}\right)^n$ for $0 \leq x \leq 1$ and $k \in \mathbb{N}^+$.