Sublinear Edge Sampling and Bernoulli Factories

University of Toronto Theory Student Seminar
(February 2023)

Morgan Shirley

Edge Sampling

Given $G = (V, E)$, sample an edge uniformly at random using a **sublinear** number of queries

$|V| = n$, $|E| = m$

$$[v_1, \dots, v_n]$$
$$\downarrow$$
$$[v_2, v_7, \dots]$$

Allowed queries:

- degree of ith vertex
- jth neighbor of ith vertex
- know approx #edges m

Edge Sampling: Results

What is the expected number of queries to sample an ϵ-approximately-uniform edge?

- Eden and Rosenbaum (2018): $O\left(\frac{n}{\sqrt{\epsilon m}}\right)$
- Tětek and Thorup (2022): $O\left(\frac{n \log\left(\frac{1}{\epsilon}\right)}{\sqrt{m}}\right)$
- Lower bound: $\Omega\left(\frac{n}{\sqrt{m}}\right)$

$\sqrt{m}$ vertices

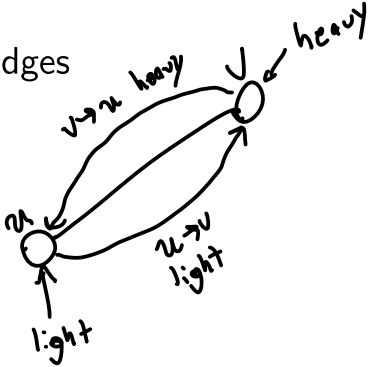
m edges

- Eden, Narayanan, and Tětek (2023): $O\left(\frac{n}{\sqrt{m}}\right)$

<u>Exactly</u> uniform!

Eden and Rosenbaum: Setup

Split edges into two directed edges



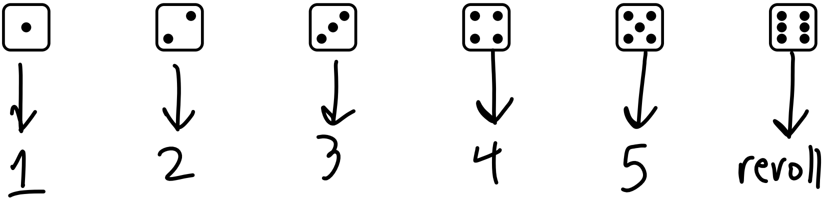
Degree threshold: θ

$\sqrt{\frac{m}{\varepsilon}}$

$\deg(v) \leq \theta$: v is *light*
$\deg(v) > \theta$: v is *heavy*

Rejection Sampling

How do I simulate a 5-sided die if I only have a 6-sided die?



Eden and Rosenbaum: Light Edges

Sampling a light edge *exactly* uniformly

- Sample random v and reject if heavy
- Sample random edge incident to v
- Return edge with probability $\deg(v)/\theta$

For a given light edge, the probability we sampled it:

$$\frac{1}{n} \cdot \frac{1}{\deg(v)} \cdot \frac{\deg(v)}{\theta} = \frac{1}{n\theta}$$

Probability we return an edge:

$$\sum_{\text{light } v} \frac{1}{n} \frac{\deg(v)}{\theta} = \frac{\#\text{ light edges}}{n\theta}$$

Eden and Rosenbaum: Heavy Edges

Sampling a heavy edge *approximately* uniformly

- Sample a uniform random light edge $u \to v$
- If v is light, reject
- Otherwise return a random edge incident to v

For a given heavy edge, the probability we sampled it:

$$\frac{1}{n\Theta} \cdot \left[\overbrace{\deg_\ell(v)}^{\#\text{ light vertices adj. to } v} \cdot \frac{1}{\deg(v)} \right]$$

$\in [(1-\varepsilon), 1]$ so $\deg_\ell(v) > (1-\varepsilon)\deg(v)$

$\#$ heavy vertices $< \frac{m}{\Theta}$

$= \frac{m}{\sqrt{\frac{m}{\varepsilon}}} = \sqrt{\varepsilon m}$

$\deg_h(v) \leq \#$ heavy vertices

$< \sqrt{\varepsilon m} < \varepsilon \cdot \deg(v)$

Probability we return an edge:

$$\sum_{\text{heavy } e} \Pr[e \text{ returned}] > \frac{(1-\varepsilon)}{n\Theta} \#\text{ heavy edges}$$

Eden and Rosenbaum: Final algorithm

Sampling an edge *approximately* uniformly

- Flip a fair coin
  - If heads, sample a light edge
  - If tails, sample a heavy edge
- Repeat until the above step returns an edge

Probability we return an edge on a given iteration:

$$\geq \frac{1}{2} \frac{\#\text{light edges}}{n\theta} + \frac{1}{2} \frac{(1-\varepsilon)\#\text{heavy edges}}{n\theta} \geq \frac{(1-\varepsilon)m}{n\theta} \approx \frac{\sqrt{\varepsilon m}}{n}$$

Expected number of attempts:

$$O\left(\frac{n}{\sqrt{\varepsilon m}}\right)$$

How do we get rid of ϵ?

Sampling a light edge *exactly* uniformly

- Sample random v and reject if heavy
- Sample random edge incident to v
- Return edge with probability $\deg(v)/\theta$

Edge probability: $\frac{1}{n\theta}$

Sampling a heavy edge *approximately* uniformly

- Sample a uniform random light edge $u \to v$
- If v is light, reject
- Otherwise return a random edge incident to v

Edge probability: $\boxed{\frac{\deg_\ell(v)}{\deg(v)}}\frac{1}{n\theta}$

$$p\frac{1}{n\theta}\frac{1}{p}=\frac{1}{n\theta}$$

Set $\boxed{\theta = \sqrt{cm}}$ for some constant c.

How do we get rid of ϵ?

Problem: $\frac{\deg_\ell(v)}{\deg(v)} = p$ varies.

Idea: Use rejection sampling with probability $\frac{1}{p}$.

Fix: reject w/p $\frac{1}{2} \to \frac{1}{2p}$

Bernoulli Factories

Given: Coin with <u>unknown</u> probability p

Goal: Simulate a coin with probability $f(p)$

Exactly and with low expected coin flips

Easy examples

$f(p) = \frac{1}{2}$

Flip twice

$00 \to$ reroll

$\left.\begin{array}{l} 01 \\ 10 \end{array}\right\}$ equal prob

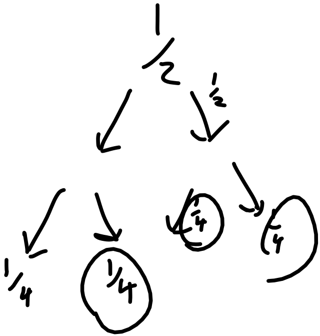
$11 \to$ reroll

$\Pr[\text{flip more than } n \text{ times}]$

$\leq \left(p^2 + (1-p)^2\right)^{\lfloor n/2 \rfloor}$

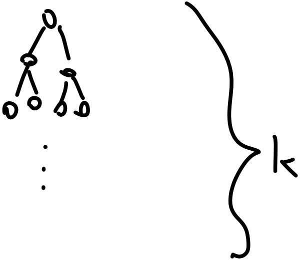
Easy examples

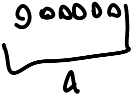
$f(p) = \frac{3}{4}$



Easy examples

$$f(p) = \frac{a}{2^k}$$





Easy examples

$f(p) = c$ for any $c \in [0, 1]$

$1/\pi$

0 → return 0

1 → continue

$0.0\,1\,01\,000\,10\,11111\ldots$

1 → return 1

0 → continue

Easy examples

$f(p) = p^2$

Flip p twice

$$\sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \quad a[i]$$

$\downarrow$

coeff $0 \leq a[i] \leq 1$

Our desired factory

$f(p) = \frac{1}{2p}$ (promised to have $p > \frac{1}{2} + c$)

Existential results: This is possible and can be done efficiently

Possibility results

Keane and O'Brien (1994): The following are *necessary* and *sufficient* to simulate $f(p)$ on domain $\mathcal{P} \subseteq [0,1]$:

1. f is continuous on $\mathcal{P}$
2. Either f is constant on $\mathcal{P}$ or $\exists t \in \mathbb{N}, \forall p \in \mathcal{P}$

$$\min\{f(p), 1 - f(p)\} \geq \min\{p^t, (1-p)^t\}.$$

Efficiency results

Nacu and Peres (2005): If f is *real analytic* on $\mathcal{P}$, f has an efficient simulation on $\mathcal{P}$.

A function f is real-analytic if f matches its Taylor series at all points

Step 1: Reduce to $2p$

2014 Hüber

$f(p) = \frac{1}{2p}$ (promised to have $p > \frac{1}{2} + c$)

$$\frac{1}{2p} = \frac{1}{\underbrace{1}_{p_1}+\underbrace{(2p-1)}_{p_2}}$$

Bernoulli Race

Flip a fair coin

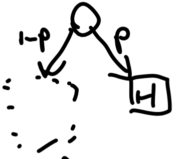
Heads: output heads w/p p_1

Tails: output tails w/p $(2p-1)$

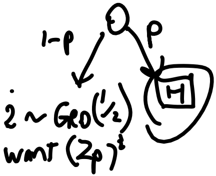
$$\Pr[\text{heads}] = \frac{1}{2}\sum_{k=0}^{\infty}\left(1-\left(\tfrac{1}{2}+\tfrac{1}{2}(2p-1)\right)\right)^k = \frac{1}{2}\sum_{k=0}^{\infty}(1-p)^k = \frac{1}{2p}$$

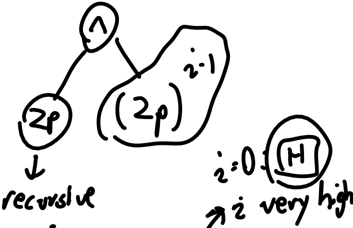
$$(2p-1) \leftarrow 1-(2p-1) = 2-2p = 2(1-p) \leftarrow 2p$$

Step 2: Factory for $2p$



want: $\frac{p}{1-p}$ $\quad p + (1-p)\frac{p}{1-p} = 2p$





$$\frac{p}{1-p} = p + p^2 + p^3 + \dots$$

$$= \frac{1}{2}(2p) + \frac{1}{4}(2p)^2 + \dots$$

$$= \sum_{i=1}^{\infty} \Pr[\mathrm{Geo}(1/2) = i](2p)^i$$

want $(2p)^{i+j}$

Step 2: Factory for $2p$

$$2p \le 0.99$$

$$(2p)^{100} < 1/2$$

$$\begin{tikzcd}
& \bigcirc \arrow[dl, "1/2"'] \arrow[dr, "1/2"] & \\
2(2p)^{100} & & \boxed{T}
\end{tikzcd}$$

$$= \left(2^{\frac{101}{100}} p\right)^{100}$$