Lecture #2 1/14/2021 [AA]

Problem: count up to n, approx.

Algo: Morris's.

- init: $X = 0$.
- @ button: $X = \begin{cases} X+1, & \text{with } Pr = 2^{-X} \\ X, & \text{oth.} \end{cases}$
- estimator: $\hat{n} = 2^X - 1$.

$\hat{n}$ r.v.

Goal: $\hat{n} \approx n$ with good prob

Claim 1: $\mathbb{E}[\hat{n}] = n$.

Pf: X_n = r.v. val. X after n button

$\hat{n} = 2^{X_n} - 1$

$\mathbb{E}[\hat{n}] = \mathbb{E}_{X_1, X_2, \ldots X_n}[2^{X_n} - 1]$

IH: $\mathbb{E}[\hat{n}] = n$.

Base case: $n=0$, $\hat{n} = 2^{X_0} - 1 = 2^0 - 1 = 0$ ✓

Inductive step: assume $\mathbb{E}[2^{X_{n-1}} - 1] = n - 1$

$$\mathbb{E}[\hat{n}] = \mathbb{E}_{X_1 \ldots X_n}[2^{X_n} - 1]$$

$$= \mathbb{E}_{X_1 \ldots X_{n-1}}\left[\mathbb{E}_{X_n}[2^{X_n} - 1]\right]$$

$$= \mathbb{E}_{X_1 \ldots X_{n-1}}\Big[\underbrace{2^{-X_{n-1}} \cdot (2^{X_{n-1}+1} - 1)}_{X \text{ was increm.}} +$$

$$+ (1 - 2^{-X_{n-1}}) \cdot (2^{X_{n-1}} - 1)\Big]$$

$$= \mathbb{E}\left[\cancel{2} - \cancel{2^{-X_{n-1}}} + \cancel{2^{X_{n-1}}} - \cancel{1} - 1 + 2^{X_{n-1}}\right]$$

$$= \mathbb{E}\left[2^{X_{n-1}} - 1 + 1\right]$$

$$= n - 1 + 1 = n. \quad \boxtimes$$

Bound how close $\hat{n}$ is to $n = \mathbb{E}[\hat{n}]$

Pearl: space bound: $O(\lg X)$.

$O(\lg(\lg n))$ with $\geq 90\%$

Claim 2: space $O(\lg\lg n)$

pf: $\hat{n} = 2^{X_n} - 1$

Markov bnd on $\hat{n}$: $\Pr[\hat{n} > 10\,\mathbb{E}[\hat{n}]]$

$$\le \frac{\mathbb{E}[\hat{n}]}{10\,\mathbb{E}[\hat{n}]}$$

$$\le 0.1.$$

When $\hat{n} \le 10 \cdot n$, have that:

$$2^{X_n} - 1 \le 10n \Rightarrow X_n \le \lg_2(10n+1)$$

$$\Rightarrow \lg_2 X_n = O(\lg\lg n).$$

Bounding how close $\hat{n}$ is to $n = \mathbb{E}[\hat{n}]$

Will Chebyshev: $\Pr[|\hat{n} - \mathbb{E}[\hat{n}]| > \lambda]$

$$\le \frac{\mathrm{Var}[\hat{n}]}{\lambda^2}.$$

Claim 3: $\mathrm{Var}[\hat{n}] \le \frac{3n(n+1)}{2} + 1. \quad = O(n^2)$

pf: similar $\mathbb{E}[\hat{n}]$.

By induction.

~~Base case:~~ $\cancel{\mathrm{Var}[2^{X_0} - 1] = 0.}$

~~Assume:~~ $\cancel{\mathrm{Var}[2^{X_{n-1}} - 1] \le 3\frac{(n-1)\cdot n}{2} + 1.}$

$$\mathrm{Var}[2^{X_n}-1] = \mathbb{E}[(2^{X_n}-1)^2] - (\mathbb{E}[2^{X_n} \ldots$$

$$= \mathbb{E}[2^{2X_n}] \underbrace{- 2\mathbb{E}[2^{X_n}] + 1 - \ldots}_{n+1 \qquad \le 0}$$

$$\Rightarrow \; = \mathbb{E}_{X_1 \ldots X_{n-1}}\left[\mathbb{E}_{X_n}[2^{2X_n}]\right]$$

$$= \mathbb{E}\left[2^{-X_{n-1}} \cdot 2^{2(X_{n-1}+1)} + (1-2^{-X_{n-1}}) \cdot 2^{2X_{n-1}} \ldots\right.$$

$$= \mathbb{E}[4 \cdot 2^{X_{n-1}} + 2^{2X_{n-1}} - 2^{X_{n-1}}]$$

$$= 3 \cdot \mathbb{E}[2^{X_{n-1}} - 1 + 1] + \mathbb{E}[2^{2X_{n-1}}].$$

$$\overset{Cl.1}{=} 3(n \cancel{-1} \cancel{+1}) + \mathbb{E}[2^{2X_{n-1}}].$$

$$= 3n + \mathbb{E}[2^{2X_{n-1}}].$$

Inductive Hyp: $\mathbb{E}[2^{2X_n}] \le \frac{3n(n+1)}{2} + 1$

Ind. step: $\mathbb{E}[2^{2X_n}] \le 3n + \mathbb{E}[2^{2X_{n-1}}]$

$$\le 3n + 3\frac{(n-1)\cdot n}{2} + 1$$

$$= \frac{3n(n+1)}{2} + 1.$$

$\hat{n}$: $\mathbb{E}[\hat{n}] = n$

$$Var[\hat{n}] \le \frac{3n(n+1)}{2} + 1 \le 2n^2 .$$

<u>Chebyshev</u>: $Pr[|\hat{n} - n| > \lambda] \le \frac{2n^2}{\lambda^2} \le 0.1$

$$2n^2 \le 0.1 \cdot \lambda^2 \to \lambda \ge \sqrt{20}\, n .$$

Enough to have $\lambda = 5n$.

$\hat{n} \in [n - 5n, n + 5n]$ with Pr ≥ 9

C

<u>Goal</u>: $\hat{n} \in \boxed{[\underbrace{n - \varepsilon n}_{=(1-\varepsilon)n}, \underbrace{n + \varepsilon n}_{=(1+\varepsilon)n}]}$

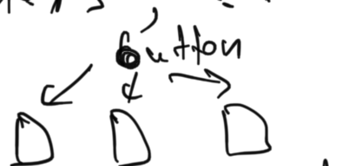
$$= \boxed{(1 \pm \varepsilon)\, n.}$$

where $\varepsilon > 0$ small const.

$\varepsilon = 0.01$.

<u>Morris+ algo</u>

take k counters, i.i.d.



$X^1 \quad X^2 \quad \dots X$.

each of them is $\boxed{\text{Morris algo}}$

counter X^i, $i=1..k$.

<u>Est:</u> $\hat{n}^k$ = avg. of estimators $X^1 \dots X$

$$= \frac{1}{k} \sum_{i=1}^{k} \left[2^{X^i} - 1\right].$$

<u>Claim 1':</u> $\mathbb{E}[\hat{n}^k] = n$.

pf: $\mathbb{E}[\hat{n}^k] = \mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k} \underbrace{2^{X^i}-1}_{\mathbb{E}=n}\right]$

$$= \frac{1}{k}\sum_{i=1}^{k} n = n. \quad \boxtimes$$

<u>Claim 2':</u> space is $O(k \cdot \lg \lg n)$.

with prob. $\geq 90\%$.

<u>Claim 3':</u> $\text{var}[\hat{n}^k] = \frac{1}{k} \text{var}[\underbrace{\hat{n}}_{\text{one coun.}}]$

pf: $\text{Var}[\hat{n}^k] = \text{Var}\left[\frac{1}{k}\sum_{i=1}^{k}(2^{X^i}-1)\right]$

$$= \frac{1}{k^2} \cdot \text{Var}\left[\sum_{i=1}^{k}(2^{X^i}-1)\right]$$

$$= \frac{1}{k^2} \cdot \sum^{k} \text{Var}[2^{X^i}-1]$$

$$= \frac{1}{k} Var[\hat{n}].$$

Goal: want $\hat{n}^k = (1 \pm \varepsilon) \cdot n$

Chebyshev: $Pr[|\hat{n}^k - n| > \varepsilon n] \le \underbrace{\frac{Var[\hat{n}^k]}{\varepsilon^2 n^2}}_{\le 0.1}$

$$Var[\hat{n}^k] \le 0.1 \cdot \varepsilon^2 n^2$$

$$\frac{Var[\hat{n}]}{k} \le 0.1 \cdot \varepsilon^2 n^2$$

$$k \ge \frac{Var[\hat{n}]}{0.1 \varepsilon^2 n^2}$$

enough to have $k = \frac{2\cancel{n^2}}{0.1 \cdot \varepsilon^2 \cdot \cancel{n^2}} = \frac{20}{\varepsilon^2}$

$$= \Theta(1/\varepsilon^2).$$

Thm [Morris + algo]: can achieve $1 \pm \varepsilon$ approx with 90% prob, using space $O\left(\frac{\lg \lg n}{\varepsilon^2}\right)$

Hashing:

Problem: [Dictionary] fix a universe U, given a set $S \subset U$, $|S| = n$, preprocess S into a data str., s.t.: "is $x \in S$ or not?"

Sol:

1) iterate $S \to O(n)$ runtime.
2) binary search $\to O(\lg n)$ r.t.
3) full-index: store a table T of size $|U|$, with $T_i = 1$ iff $i \in$

q.t. $\lg n$
sp: $O(n)$

q.t. $O(1)$
sp.: $O(|U|)$.

Q: can we get $O(1)$ q.t. & $O(n)$ space

$|U| = 2^{32}$ = all possible IP's. (all bin strings of len. 32).

$U \xrightarrow{h} \{1, \dots, m\} = [m]$

$\boxed{h : U \to [m]}$.

$\circledast$ s.t.: $\forall i \in S$ and given x we have: $h(i) = h(x)$ iff $i = x$.

Solution to Dict.:

reduce univ. U to $[m]$

- compute $h(i)$, $i \in S$.
- store table $T_j = 1$ iff $j = h(i)$, $i \in S$.
- @ query x, check if $T_{h(x)} = 1$.

apply sol #3.

Perf: - space $O(m)$
- q.t. $O(1)$ + $\boxed{\text{time to comp. } h(x}$

Solution #4: pick h randomly, hope property $\circledast$ holds with good prob.

Def: collision $h(i) = h(x)$.

for m as small as possible.