Last time:

given a large vector $v \in \mathbb{R}^n$
via updates, maintain
$O(1/\varepsilon^2)$ numbers so to estimate
$F_2 = ||v||_2$ within $1 \pm \varepsilon$.

Tug of War sketch.

"Detecting a DDOS attack"
Would not it be nice to estimate
$||v||_\infty$?

[AMS]: requires $\Omega(n)$ space!

Uses communication complexity lower bound for DISJ [BJKS].

Disjointness:

Alice $x \in [n]$ decide if $x \cap y = \emptyset$
Bob $y \in [n]$

It (hard) requires $\Omega(n)$ communication
even for randomized protocols with public randomness.

Claim: If $\exists$ sketch of size $O(s)$ for
estimating $||v||_\infty$ within $1 \pm \varepsilon \Rightarrow$ DISJ has
for $\varepsilon < 0.1$ protocol of
\( \| u \|_\infty \) is hard to estimate only if it does not matter.

**Heavy Hitters.**

**Def:** \( i \in [n] \) is \( \varphi \)-HH if \( u_i > \varphi \cdot \| u \|_1 \). \((\varphi > 0 \text{ think } \varphi = 0.01)\)

**Cl:** There are \( \leq \frac{1}{\varphi} \) \( \varphi \)-HH's

**Th:** \( \forall \varepsilon \varphi > 0 \) can maintain sketch of size \( O(\frac{\log n}{\varepsilon^2 \varphi^2}) \) and recover a set \( A \subseteq [n] \) s.t. whp \( \forall i \)

- if \( i \) is \( \varphi \)-HH \( \Rightarrow i \in A \)
- if \( i \in A \) \( \Rightarrow i \) is \((1-\varepsilon)\varphi\)-HH.

**Technique:** Count-Sketch, a generalization of Tug-of-War. Uses hashing and the median estimator!

**Morale:** the right modelling is the key!

**Count-Sketch:**

- enough to estimate fixed \( u_i \) up to \( \pm \varepsilon \varphi \| u \|_1 \).
- \( h : [n] \rightarrow [w] \) - 2-independent (\( w \) to be chosen later).
- \( \sigma : [n] \rightarrow \{-1, 1\} \)
Maintain $V_j \in \mathbb{R}^n$

$$v_j = \sum_{i:h(i) = j} \sigma_i u_i$$

Sparse dimension reduction (Tug-of-war is "dense").

Will see many times later.

Estimator: $u_i \approx \sigma_i \cdot v_{h(i)}$.

$V_{h(i)} = u_i \sigma_i + \text{leftovers}$.

**Cl** $E[\text{leftovers}] = 0$.

**Pf** by linearity of expectation.

$\text{Cl}$ $\text{Var}[\text{leftovers}] \leq \frac{\|u\|^2}{w} \leq \frac{\|u\|^2}{w}$

**Pf** $\text{Var}[\text{leftovers}] \leq E[\text{leftovers}^2] =$

$$= \mathbb{E} \left[ \left( \sum_{i \neq i'} \sigma_i u_i \right)^2 \right] = \mathbb{E} \left[ \sum_{i \neq i'} u_i^2 \right] = \frac{\|u\|^2}{w} \leq \frac{\|u\|^2}{w}$$

Cor $\Pr[|\text{leftovers}| \gg \frac{\|u\|_1}{\sqrt{w}}] \ll 1$.

If $w \sim \frac{1}{\varepsilon^2 n^2}$, whp get estimate for

$$u_i \pm \varepsilon \Phi \|u\|_1.$$
Problem to recover $HH$ need a good estimate

Median trick

$X, \hat{X}_1, \ldots, \hat{X}_k$ - estimators for $X$

$\Pr[\hat{X}_i \approx X] \geq \frac{2}{3}$

$\Pr[\text{med}(\hat{X}_1, \ldots, \hat{X}_k) \approx X] \geq 1 - 2^{-\Omega(k)}$

$k$ coins "heads" w.p. $\geq \frac{2}{3}$

$\Pr[> \frac{k}{2} \text{ heads}] \geq 1 - 2^{-\Omega(k)}$

Chebyshev merely gives $1 - \frac{4}{k}$

(but requires merely pairwise independence)

Th (Chernoff bound)

$X_1, \ldots, X_n$ - independent $\{0, 1\}$

$\mu = \mathbb{E}[X], \quad X = \sum X_i$

$\Pr[X > (1+\delta)\mu] < \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$

$\Pr[X < (1-\delta)\mu] < \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu$
\[ \Pr \left[ \geq \frac{k}{2} \text{ heads} \right] \geq 1 - 2^{-\frac{k^2}{2}} \]

Full independence yields sharper concentration.

Back to Count-Sketch.

Maintain \( O\left(\frac{1}{\varepsilon^2 \log^2} \right) \) numbers.

\[ \forall i \ \Pr \left[ \hat{u}_i \in u_i \pm \varepsilon \|u\|_1 \right] \geq 0.99. \]

Can boost 0.99 to \( 1 - \frac{1}{n^{10}} \) by storing \( O\left( \frac{\log n}{\varepsilon^2 \log^2} \right) \) numbers (estimator is the median of \( O(\log n) \) of them).

Union bound: w.p. \( 1 - \frac{1}{n^3} \)

\[ \forall i \ \hat{u}_i \in u_i \pm \varepsilon \|u\|_1. \] (quantifier is inside)

\( HH \) is solved!

Comments:

1) Supports "turnstile updates"

2) Solves a stronger problem \( \pm \varepsilon \|u\|_2 \).

For our problem Count-Min gives space \( O\left( \frac{\log n}{\varepsilon^2} \right) \).
l₀-sampling

Maintain a vector \( u \in \{-1, 0, 1\}^n \) under turnstile updates.

Goal: sample one non-zero element of \( u \) uniformly (or just find any non-zero). Strange thing to want but useful for graph sketching.

Case 1: Suppose \( u \) has at most 10 non-zeroes. Then can use HH to recover all of them using space \( O(\log n) \).

Case 2: Say \( u \) has \( \sqrt{n} \) non-zeroes. HH is expensive...

Idea:

\[
\hat{u}_i = \begin{cases} 
  u_i, & \text{w.p. } \frac{10}{\sqrt{n}} \\
  0, & \text{w.p. } 1 - \frac{10}{\sqrt{n}} 
\end{cases}
\]

(use PRG to sample).

\[E[\# \text{ non-zeroes in } \hat{u}] = 10.\]

By Markov, \( \Pr[\# \text{nnz} \geq 100] \leq 0.1. \)

\[\Pr[\# \text{nnz} = 0] = \left(1 - \frac{10}{\sqrt{n}}\right)^{\sqrt{n}} \approx e^{-10} \leq 0.01. \]

\[\Pr[\text{success}] \geq 0.89. \]
Overall solution:

- $F^*_2$ estimator (super-charged with median estimator).
- $O(\log n)$ scales
  
  $\geq 1 < 2$ non-$z$.  
  For each scale  
  downsample and  
  run HH.
  
  $\geq 2 < 4$ non-$z$.  
  use $F^*_2$ to check  
  if sampling worked
  
  $\vdots$

  $\geq \frac{n}{2} < n$

Repeat each scale $O(\log n)$ times.