CS 229r: ITCCS
LECTURE 20

4/11/2019

TODAY: STREAMING ALGORITHMS & LIMITS
- STREAMING MODEL
- EXAMPLE ALGORITHMS
- LOWER BOUND

———— x ————

MODEL: - Input = Stream $x_1, x_2, \ldots x_m$ $\quad x_i \in [n]$

- Target/Goal: Compute $f(x \ldots x_m)$ $\quad f: [n]^m \to \Gamma$

- Restriction: Small space s

  - ~~So~~ Algorithm $\begin{cases} A: \{0,1\}^s \times [n] \to \{0,1\}^s \\ g: \{0,1\}^s \to \Gamma \end{cases}$

  - ~~Output~~ Want

$f(x_1 \ldots x_m)$ derivable from σ_m (so $f(x \ldots x_m) = g(\sigma_m)$)

where $\sigma_0 = \bar{0}$

$\sigma_i = A(\sigma_{i-1}, x_i)$

- ~~Question~~: Allow randomization so σ_0 is random

& only need $\Pr_{\sigma_0}[f(x_1 \ldots x_m) = g(\sigma_m)] \geq \frac{2}{3}$

What can you compute this way? with $s = \mathrm{poly}(\log m, \log n)$

Examples:

- (Warning: - Algorithms below use lots of randomness (like $m_* n$ bits))

  - Can be implemented with $\log m \times \log n$ bits; but we'll skip this.)

I. Frequency Moments:

- frequency $f_i(x_1 \ldots x_m) = \left| \{ j \mid x_j = i \} \right| \quad \forall i \in [n]$

- k^{th} moment of Frequencies

$$F_k(x_1 \ldots x_m) \triangleq \sum_{i \in [n]} f_i^k (x_1 \ldots x_m)$$

- ~~Can we compute k^{th} moments?~~

$$F_0(x_1 \ldots x_m) \triangleq \lim_{k \to 0} F_k(x_1 \ldots x_m)$$

$$= \left| \{ i \mid f_i > 0 \} \right|$$

- Can we compute k^{th} moments with non-trivial space?

~~[Flajolet-Martin '85]~~:

- Trivial: $F_1(x_1 \dots x_m) = m$ $\boxtimes$

- [Flajolet-Martin '85]:

  - Pick random $h : [n] \to [0,1]$ (recall warning)

  - Compute $\min_{j \in [m]} \{h(x_j)\} =: h_{min}$

  - Output: $\frac{1}{h_{min}}$ $\boxtimes$

- Claim: $\mathbb{E}\left[\frac{1}{h_{min}}\right] \approx F_0$.

  [Slightly cleaner to pick $h(i)$ i.i.d $\sim \exp(1)$].

- Slightly better algorithm / guarantee by picking t^{th} smallest element & outputting $\frac{t}{h_{t\text{-}min}}$

- Leads to $O(\frac{1}{\epsilon^2}\log^c n)$ alg to output $(1 \pm \epsilon)$-approx to F_0.
- Exercise: Verify all sentences above.

[Alon-Matias-Szegedy]: F_2 !

- <u>Algorithm</u>: Pick $h: [n] \rightarrow \{+1, -1\}$ unif. at random.

  <u>Compute</u>: $V = \sum_{j=1}^{m} h(x_j)$

  <u>Output</u>: V^2.

- <u>Claim</u>: $\mathbb{E}_h \left[V^2 \right] = F_2(x_1 \dots x_m)$

- <u>Proof</u>: $\mathbb{E}_h \left[V^2 \right] = \mathbb{E}_h \left[\left(\sum_{j=1}^{m} h(x_j) \right) \left(\sum_{k=1}^{m} h(x_k) \right) \right]$

$$= \sum_{j=1}^{m} \sum_{k=1}^{m} \underbrace{\mathbb{E}_h \left[h(x_j) h(x_k) \right]}_{\substack{=1 \text{ if } x_j = x_k \\ =0 \text{ o.w.}}}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{m} \mathbb{1}_{x_j = x_k = i}$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} \mathbb{1}_{x_j = i} \right)^2 = \sum_{i=1}^{n} f_i^2 = F_2$$

[Variance can be est. $\Rightarrow 1 \pm \epsilon$- approx in space $(1/\epsilon^2 \dots)$

- What about F_3 F_4 etc. (can also consider $F_{1.5}$, $F_{.5}$ ... but we won't)

- AMS: F_k can be computed in space $\tilde{O}(n^{1-1/k})$.

  poly not polylog!

  Necessary? Will see shortly.

AMS Algorithm:

Pick $j \in [m]$ uniformly

Let $i = a_j$

Let $f^+ = |\{ j' \mid a_{j'} = i, \; j' \geq j \}|$

Output $X = (f^+)^k - (f^+ - 1)^k$

Claim: $\mathbb{E}[X] = F_k$ (exercise)

Claim $V[X] \approx n^{1-1/k}$ (not-exercise).

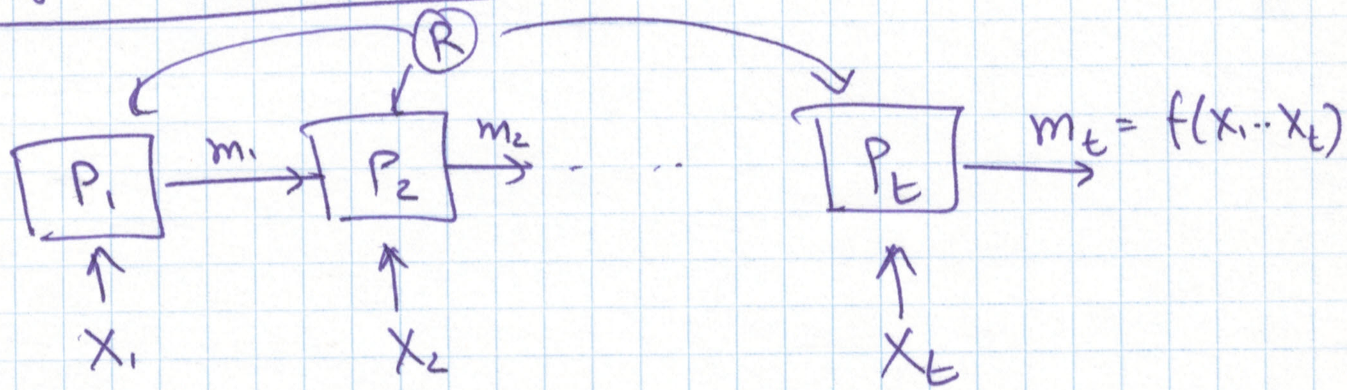
[BJKS] (the Disjointness via IC paper):

- Computing F_3 requires $n^{1/3}$ space in one-pass.
- computing F_4 $n^{1/4}$ space in $O(1)$-passes.

- Proof: from t-party Set-Disjointness

——x——

t-party Communication Model:



- One-way: $P_1 \to P_2 \to \cdots P_t$
- ~~t~~ r-pass-one-way: $\underbrace{P_1 \to \cdots P_t \to \cdots \to \cdots}_{r \text{ times}}$

- one way cc ≥ r-pass-cc ≥ arbit-cc.

——x——

~~Streaming ≤ t-party~~ Streaming ≥ t-pass-cc.

~~Player~~ Divide stream into t pieces. P_i gets i^{th} piece.
Together simulate s-space algorithm to get $O(s)$ comm. com for problem.

t-party-lower bounds of t-Set-Disj

t-Set-Disj : YES : $X_1 \dots X_t \in \{0,1\}^k$

& $\exists i$ s.t. $X_1^{(i)} \dots X_t^{(i)} = 1$

& $\forall i' \neq i$ $\sum_{j=1}^{t} X_j^{(i')} \leq 1$ } "otherwise disjoint"

NO : $X_1 \dots X_t \in \{0,1\}^k$

$\forall i$ $\sum_{j=1}^{t} X_j^{(i)} \leq 1$ [Strongly disjoint]

'Distinguish YES from NO

Thm: Oneway CC $\gtrsim \frac{k}{t}$

arbit - CC $\geq \frac{k}{t^2}$

Proof: As usual. k-bit DISJ $\geq \underbrace{\frac{k}{t} \text{ 1-bit DISJ}}_{\text{C.I.C.}}$

Review Exercise!

$\text{C.I.C}^{\text{one-way}}$ (1-bit Disj) $\geq \frac{1}{t}$

$\text{CIC}^{\text{arbit.}}$ (") $\geq 1/t^2$

} Omitted

Streaming Lower Bound For F_3

Use $t = (2n)^{1/k}$;

$X_1 \dots X_t \in \{0,1\}^k \longrightarrow$ ~~a...a~~ ~~$a_1 \dots a$~~

$$\underbrace{a_1 \dots}_{\text{elements in } X_1} \dots \underbrace{a_m}_{\text{el'ts in } X_t}$$

YES $\Rightarrow$ $F_k \geq t^k$ (some el't in all t sets)

$\geq 2n$

NO $\Rightarrow$ $f_i \leq 1 \quad \forall i$

$\Rightarrow F_k \leq n$. ☒