Nearly Optimal Distinct Elements and Heavy Hitters on Sliding Windows

Vladimir Braverman

Elena Grigorescu

Harry Lang

David P. Woodruff

Samson Zhou

Johns Hopkins University

Carnegie Mellon University

Purdue University

Streaming Model

- Input: Elements of an underlying data set S, which arrives sequentially
- Output: Evaluation (or approximation) of a given function
- Goal: Use space *sublinear* in the size of the input S

- Sliding Window: "Only the W most recent updates form the underlying data set S"
  - Recent interactions, time sensitive

Distinct Elements (L_0 Norm)

- Given a set S of m elements from $[n]$, let F be the number of distinct elements in S. (How many elements of $[n]$ appear *at least once* in S)
- Goal: Give $(1+\epsilon)$-approximation of F.
- Best-known algorithm: $O\left(\frac{1}{\epsilon^3}\log^2 n + \frac{1}{\epsilon}\log^3 n\right)$ bits of space [KaneNelsonWoodruff10, BravermanOstrovsky07]

Our result: $O\left(\frac{1}{\epsilon^2}\log n\left(\log\log n \log\frac{1}{\epsilon}\right) + \frac{1}{\epsilon}\log^2 n\right)$ bits of space

Distinct Elements

Upper Bound	Lower Bound
$O\left(\frac{1}{\epsilon^3}\log^2 n + \frac{1}{\epsilon}\log^3 n\right)$ [KNW10, BO07]	$\Omega\left(\frac{1}{\epsilon^2} + \log n\right)$ [AMS99, IW03]
$O\left(\frac{1}{\epsilon^2}\log n\left(\log\log n \log\frac{1}{\epsilon}\right) + \frac{1}{\epsilon}\log^2 n\right)$ [Here]	$\Omega\left(\frac{1}{\epsilon^2}\log n + \frac{1}{\epsilon}\log^2 n\right)$ [Here]

Optimal up to $\log\log \mathrm{n}, \log\frac{1}{\epsilon}$ factors

314	293	812	758	314	211
112	067	183	447		
033	314	905	717	623	576
128	443	007	889	572	511
223	981	961	011	314	414
314	000	668	295	223	366
552	877	256	505	566	314
191	993	314	054	007	314

Heavy Hitters

Heavy-Hitters

- Given a set S of m elements from $[n]$, let f_i be the frequency of element i. (How often it appears)
- Let L_2 be the norm of the frequency vector:

$$L_2 = \sqrt{f_1^2 + f_2^2 + \cdots + f_n^2}$$

- Goal: Given a set S of m elements from $[n]$ and a parameter ϵ, output the elements i such that $f_i > \epsilon\, L_2$...and no elements j such that $f_j < \frac{\epsilon}{16}\, L_2$.

Heavy-Hitters in the Sliding Window Model

Upper Bound	Lower Bound
$O\left(\frac{1}{\epsilon^4}\log^3 n\right)$ [BGO14]	$\Omega\left(\frac{1}{\epsilon^2}\log n\right)$ [JST11]
$O\left(\frac{1}{\epsilon^2}\log^2 n\left(\log\log n + \log\frac{1}{\epsilon}\right)\right)$ [Here]	$\Omega\left(\frac{1}{\epsilon^2}\log^2 n\right)$ [Here]

Optimal up to $\log\log \mathrm{n}, \log\frac{1}{\epsilon}$ factors