Last lecture we introduced the AMS algorithm for estimating F_k. Today we continue the proof and analysis of that algorithm.

Lemma 3 (3.2.1 Continuation) $F_1 F_{2k-1} \leq n^{1-\frac{1}{k}}(F_k)^2$

Proof.

We recall $F_k = \sum_{i=1}^{m} f_i^k = \ell_k^k = ||f||_k^k$. Also, if X is the output of the algorithm we showed

$$E[X] = F_k$$

$$Var[X] \leq k(F_1 F_{2k-1})$$

Recall that we define $\max_i f_i = F_\infty$ and that $F_\infty^{k-1} = (F_\infty^k)^{\frac{k-1}{k}}$. Also due to convexity, for all $k \geq 1$: $\frac{\sum x_i}{n} \leq (\frac{\sum x_i^k}{n})^{\frac{1}{k}}$.

$$\begin{aligned} F_1 F_{2k-1} =& (\sum_{i=1}^{n} f_i)(\sum_i {f_i}^{2k-1}) \\ \leq& (\sum_i f_i) F_\infty^{k-1} (\sum_i {f_i}^k) \\ \leq& (\sum_i f_i)(\sum_i f_i^k)^{\frac{k-1}{k}} (\sum_i f_i^k) \\ \leq& n^{1-\frac{1}{k}} (\sum_i f_i)^{\frac{1}{k}} (\sum_i f_i^k)^{\frac{k-1}{k}} (\sum_i f_i^k) \\ =& n^{1-\frac{1}{k}} (\sum f_i^k)^2 \end{aligned}$$

■

So $Var[X] \leq k F_1 F_{2k-1} \leq k n^{1-\frac{1}{k}} F_k^2$.

Note: A good explanation about frequency moments can be found in [MM13].

Now we use the median of means trick. Suppose that we run $h = \frac{c}{\epsilon^2} k n^{1-\frac{1}{k}}$ copies of the basic algorithm. Let X' be the average of the estimator:

$$E[X'] = F_k$$

$$Var[X'] \leq \frac{Var[X]}{h} \leq \frac{\epsilon^2}{c} F_k^2$$

Using Chebyshev we have:

$$Pr[|X' - E[X']| \geq \epsilon E[X']] \leq \frac{Var[X']}{\epsilon^2 E[X']^2} \leq \frac{1}{c}$$

We can get a $(\epsilon, \frac{1}{3})$-estimator by choosing $c = 3$, call this an intermediate estimator. If we use $t = c' \log \frac{1}{\delta}$ for parallel copies of this and return the median $\rightarrow$ hence we get a (ϵ, δ)-estimator. The space for the overall use is $O(\log \frac{1}{\epsilon} \cdot \frac{k}{\epsilon^2} \cdot n^{1-\frac{1}{k}})$. And the space for each of the copies of the basic estimator is $O(\log m + \log n)$.

4.1 Linear - Sketching

The algorithm of AMS we saw last time for estimating F_k works for all $k \geq 2$ with space $\tilde{O}(n^{1-\frac{1}{k}})$, but we prefer to have polylog space. AMS also gave an amazingly simple algorithm for estimating F_2. This is a sketching algorithm in the following sense. Suppose we have two streams: σ_1 and σ_2, and an algorithm that computes a structure $z(\sigma_1)$ and $z(\sigma_2)$. We call these structures sketch if $\exists$ an efficient (space) combining algorithm A such that for any two streams σ_1 and σ_2 if $\sigma_1\sigma_2$ is the stream obtained by concatinating the two then then $A(z(\sigma_1), z(\sigma_2)) = z(\sigma_1\sigma_2)$.

Suppose the values of a stream $\sigma_1 ... \sigma_m$ where from $[n]$, and we start with a n-dimensional vector $x = (0, ..., 0)$ and each time a new token comes, we update x. So, each token i corresponds to an updated (i, a) where $x_i \leftarrow x_i + a$. Typically $a = 1$ but it could be different.

- If a is allowed to be negative, we have a turnstile stream model.
- If we require x_i's be always non-negative, we have a strict turnstile model.
- If a is required to be positive, we have a cash register model.

Linear Sketch: Corresponds to a $k \times n$ matrix M and the sketch for a vector x becomes Mx. So composing two linear sketches $Mx + Mx' = M(x + x')$.

4.2 Estimating F_2 by Sketching

Estimating $||x||_2$ (ℓ_2 norm) of a data vector x has lots of applications. So estimating $||x||_2^2$ is probably the most important of all other frequency moments. AMS algorithm for $k = 2$ is an amazingly simple algorithm that produces a sketch.

Through the use of the generic algorithm in the Lecture 2.1 to estimate the F$_k$, we can develope an algorithm for F_2, which is useful in case, for example, we require to gather analytical meaning of the data that is being streamed.

> <u>AMS F_2 Estimator</u>[AMS99]
> Let h be a random hash function from a 4 universal family $\mathcal{H}$, $h_i[n] \rightarrow \{-1, +1\}$
>
> $x \leftarrow 0$
> While the stream is non empty do
>   let a_j be next element
> $x \leftarrow x + h(a_j)$
> return x^2

4.2.1 Analysis

The previous algorithm can be described in the following way as well: one can think of $Y_1 ... Y_n$ as 4-wise independent random variables $\{-1, +1\}$ and in each round $x \leftarrow x + Y_{aj}$. Therefore, we can get $Y_i = h(i)$. Let $X = \sum f_i Y_i$ be value of x at the end of the stream. For all $E[Y_i] = 0$ and $E[{Y_i}^2] = 1$. Since the Y_i's are also 2-wise independent $\underset{i \neq i'}{E}[Y_i {Y_i}'] = 0$.

Thus:

$$\begin{aligned} E[X^2] =& E[\sum_i \sum_{i'} f_i f_i' E[Y_i Y_i']] \\ =& \sum_i {f_i}^2 E[{Y_i}^2] + \sum_{i \neq i'} f_i f_{i'} E[Y_i Y_i'] \\ =& \sum_i {f_i}^2 = F_2 \end{aligned}$$

To compute the variance:

$$Var[X^2] = E[X^4] - E[X^2]^2 = E[X^4] - F_2^2.$$

Also $E[X^4] = \sum_i \sum_j \sum_k \sum_\ell f_i f_j f_k f_\ell E[Y_i Y_j Y_k Y_\ell]$. Suppose one of i, j, k, ℓ appears exactly once in the 4-tuple , say $i \notin \{j, k, \ell\}$. Then by by 4-wise independent $E[Y_i Y_j Y_k Y_\ell] = E[Y_i] E[Y_j Y_k Y_\ell] = 0$, so, the only non zero terms in $E[X^4]$ is when all 4 indices are the same or when we have two pairs:

$$\begin{aligned}
E[X^4] =& \sum_i \sum_j \sum_k \sum_\ell f_i f_j f_k f_\ell E[Y_i Y_j Y_k Y_\ell] \\
=& \sum_i f_i^4 E[Y_i^4] + 6 \sum_{i=1} \sum_{j=i+1} f_i^2 f_j^2 E[Y_i^2 Y_j^2] \\
=& F_4 + 6 \sum_i \sum_{j=i+1} f_i^2 f_j^2
\end{aligned}$$

Therefore:

$$\begin{aligned}
Var[X^2] =& E[X^4] - E[X^2]^2 \\
=& E[X^4] - F_2^2 \\
=& F_4 + 6 \sum_i \sum_{j=i+1} f_i^2 f_j^2 - F_2^2 \\
=& F_4 + 6 \sum_i \sum_{j=i+1} f_i^2 f_j^2 - (\underbrace{\sum f_i^4}_{F_4} + 2 \sum_i \sum_{j=i+1} f_i^2 f_j^2) \\
=& 4 \sum_i \sum_{j=i+1} f_i^2 f_j^2 \\
\le& 2F_2^2
\end{aligned}$$

Using the (now standard) method of median of the means, we first use $O(1/\epsilon^2)$ estimators and apply Chebyshev's inequality to obtain an $(\epsilon, \frac{1}{3})$-estimator. Then use the median trick and $O(\log \frac{1}{\delta})$ independent average estimators we obtain an (ϵ, δ)-estimator using $O(\frac{1}{\epsilon^2} \log \frac{1}{\delta})$ parallel copies.

4.2.2 Space Usage

We have $E[X^2] = F_2$, where each average estimator uses O($\underbrace{\frac{1}{\epsilon^2}}_{basic}$), later we apply Chebyshev to obtain an $(\epsilon, \frac{1}{2})$ estimator (intermediate estimator to reduce variance). Then, we use the O($\log \frac{1}{3}$) of intermediate estimator to take the median and obtain an (ϵ, δ)-estimator using O($\frac{1}{\epsilon^2} \log \frac{1}{\delta}$) for the parallel copies. The overall space is O($\frac{1}{\epsilon^2} \log \frac{1}{\delta} (\log m + \log n)$).

4.2.3 Geometric intuition

Suppose we have t independent copies of the basic sketch. Let $M \in \mathbb{R}^{t \times n}$ matrix for the final sketch, which transforms the frequency vector into t dimensional vector x. M is a matrix with ± 1 entries, $M_{ij} = h_i(j)$ where h_i is for the $i'th$ copy using $t = \frac{6}{\epsilon^2}$ copies (and applying Chebyshev):

$$Pr[|\frac{1}{t} \sum_{i=1} X_i{}^2 - F_2| \ge \epsilon F_2] \le \frac{1}{3}$$

Recall that

$$\mathrm{F}_k = \sum_i {f_i}^k = ||F||_k^k$$

So, with probability $\geq \frac{2}{3}$: $||\frac{1}{\sqrt{t}} Mx||_2 = \frac{1}{\sqrt{t}}||x||_2 \in [\sqrt{1-\epsilon}||x||_2, \sqrt{1+\epsilon}||x||_2]$. We can think of $M/\sqrt{t}$ as a random matrix that "reduces dimension" of an n-dimensional vector x to a t-dimenstion sketch while preserving the ℓ_2-norm approximately. We can use the AMS sketch (a linear sketch) that gives us an estimate of the ℓ_2-norm and use it to estimate ℓ_2-difference between two streams σ and σ': $||f(\sigma) - f(\sigma')||_2$.

References

AMS99 N. Alon, Y. Matias, and M. Szegedy, The Space Complexity of Approximating the Frequency Moments. *J. Comput. Syst. Sci*, 31(2):137-147, 1999.

MM13 A. McGregor and S. Muthukrishnan, Data Stream Algorithms for Vectors: Draft Chapter*. October 27, 2013.