CS 15-851: Algorithms for Big Data Spring 2024

Lecture 7 — 02/29/2024

Prof. David Woodruff *Scribe: Harry H. Jiang*

1 p-norm Estimation

Recall the sketching matrices $P \cdot D$, where P consists of a CountSketch matrix, and D consists of a diagonal matrix with diagonal elements $1/E_i^{1/p}$, with E_i being independent standard exponential random variables.

For arbitrary y, $||Dy||_\infty$ looks like

$$||Dy||_\infty^2 = \max_i \frac{||y_i||^p}{E_i} = \frac{1}{\min_i \frac{E_i}{||y_i||^p}} \equiv \frac{1}{E/||y_i||_p^p} = \frac{||y_i||_p^p}{E} \tag{1}$$

and the probability of a reasonable value of E is $\mathbf{Pr}\Big[E \in [1/10, 10]\Big] = (1-e^{-10}) - (1-e^{-1/10}) > 4/5$ (this actually evaluates to just over 9/10).

As such, $||Dy||_\infty^p$ is a good estimate for $||y||_p^p$, but $Dy \in \mathbb{R}^n$ is a large vector, so sketching using matrix $P \in \mathbb{R}^{s\times n}$ is needed to reduce computation cost.

Intuitively, P is hashing coordinates of Dy into buckets and taking a signed sum; most items cancel out and then $||PDy||_\infty \simeq ||Dy||_\infty$. It is known previously that P is composed of hash functions $h : [n] \to [s]$ and $\sigma : [n] \to \{-1, 1\}$ (assuming they are truly random). Given that $||Dy||_\infty/||y||_p \in [1/10^{1/p}, 10^{1/p}]$ with probability $> 4/5$, to achieve $||PDy||_\infty \simeq ||Dy||_\infty$ with good probability, it is necessary to have:

1. in each bucket i not containing the maximum value, $|(PDy)_i| \leq ||y||_p/100$
2. in each bucket i containing the maximum value, $\Big||(PDy)_i| - ||Dy||_\infty\Big| \leq ||y||_p/100$

Let $\delta(\text{event}) = 1$ if a given event holds and $\delta(\text{event}) = 0$ otherwise. It is then possible to define a given element of PDy as $(PDy)_i = \sum_j \delta(h(j) = i) \cdot \sigma_j \cdot (Dy)_j$. Due to σ, its expectation is $\mathbb{E}[(PDy)_i] = 0$. The evaluation of its variance as follows:

$$\mathbb{E}_P[(PDy)_i^2] = \sum_{j,j'} \mathbb{E}[\delta(h(j)=i)\cdot\delta(h(j')=i)\cdot\sigma_j\cdot\sigma_{j'}](Dy)_j(Dy)_{j'} = \frac{1}{s}||Dy||_2^2 \tag{2}$$

$$\mathbb{E}_D[||Dy||_2^2] = \sum_i y_i^2\mathbb{E}[D_{i,i}^2] \tag{3}$$

$$\mathbb{E}[D_{i,i}^2] = \int_0^\infty t^{\frac{2}{p}}e^{-t}dt = \int_0^1 t^{\frac{2}{p}}e^{-t}dt + \int_1^\infty t^{\frac{2}{p}}e^{-t}dt \tag{4}$$

$$\leq \int_0^1 t^{\frac{2}{p}}dt + \int_1^\infty e^{-t}dt = \left(\frac{1}{1-\frac{2}{p}}\right)t^{-\frac{2}{p}}\Big|_0^1 - e^{-t}\Big|_1^\infty = O(1) \tag{5}$$

$$\mathbb{E}_P[(PDy)_i^2] = O\left(\frac{1}{s}\right)||y||_2^2 = O\left(\frac{1}{s}\right)(n^{1-\frac{2}{p}}||y||_p^2). \tag{6}$$

The last line holds due to Hölder's Inequality ($||y||_2^2 = \sum_i^n y_i^2 \leq (\sum_i^n (y_i^2)^{p/2})^{2/p}(\sum_i^n 1^q)^{1/q} = ||y||_p^2 \cdot n^{1-2/p}$).

Definition (Bernstein's Bound)**.** Suppose independent random variables $R_1,\ldots,R_n$, and for all j, $|R_j| \leq K$, and $\text{Var}[\sum_j R_j] = \sigma^2$. Then, there exists constants c, C such that for all $t > 0$,

$$\mathbf{Pr}\left[\left|\sum_j R_j - \mathbb{E}[\sum_j R_j]\right| > t\right] \leq C\left(e^{-\frac{ct^2}{\sigma^2}} + e^{-\frac{ct}{K}}\right) \tag{7}$$

In order to get $1/\texttt{poly}(n)$ error probability, set $R_j = \delta(h(j)=i)\dot{\sigma_j}(Dy)_j$, $t = ||y||p/100$, and $s = \Theta(n^{1-2/p}\log n)$ to handle all parameters required for Bernstein's bound other than K.

It is possible to treat large R_j separately, where $R_j > \frac{\alpha||y||_p}{\log n}$ for a sufficiently small $\alpha > 0$. If $|R_j| > \frac{\alpha||y||_p}{\log n}$, then necessarily $(Dy)_j \geq \frac{\alpha||y||_p}{\log n}$ (define j as "large" if this is the case, "small" otherwise). Then, any j may be large with probability and expectation

$$\mathbf{Pr}[j \text{ is large}] = \mathbf{Pr}\left[\frac{|y_j|}{E_j^{1/p}} \leq \frac{\alpha||y||_p}{\log n}\right] = \mathbf{Pr}\left[\frac{|y_j|^p}{\alpha^p||y||_p^p}\log^p n \leq E_j\right] \tag{8}$$

$$= 1 - e^{-\frac{|y_j|^p\log^p n}{\alpha^p||y||_p^p}} \leq \frac{|y_j|^p\log^p n}{\alpha^p||y||_p^p} \tag{9}$$

$$\mathbb{E}[R_j \text{ for large } j] \leq \sum_j \frac{|y_j|^p\log^p n}{\alpha^p||y||_p^p} = \frac{\log^p n}{\alpha^p} \tag{10}$$

There are $s = O(n^{1-2/p}\log n)$ buckets and $\frac{\log^p n}{\alpha^p}$ items. By Markov bound, there are $O(\log^p n)$ large j with constant probability. D is conditioned on the above as well as $||Dy||_\infty \in [||y||_p/10^{1/p}, ||y||_p \cdot 10^{1/p}]$ (which happens with probability $> 4/5$). All the large j should then be perfectly hashed into separate buckets by P. (If there are b balls and Cb bins, $\mathbf{Pr}[\text{collision}] \leq \binom{b}{2}1/Cb \leq 1/2C$)

Bernstein's bound can then be applied separately for the small indices j for each hash bucket. $\mathbb{E}[(PDy)_i] = 0$ for each hash bucket i, and $\mathbb{E}[(PDy)_i^2] = O(1/s)(n^{1-2/p}||y||_p^2)$. Assuming $K = \max_j |R_j| \leq \alpha||y||_p/\log n$ for small j in a bucket (it can be shown that $\text{Var}[R_j]$ is $O(1/s)(n^{1-2/p}||y||_p^2)$

even if no j is large. Setting $t = ||y||_p/100$ and $s = \Theta(n^{1-2/p}\log n)$ in Bernstein's bound, for a bucket $(PDy)_i$

$$\mathbf{Pr}\left[\left|\sum_{\text{small } j} \delta(h(j)=i)\cdot\sigma_j\cdot(Dy)_i\right| > \frac{||y||_p}{100}\right] \le C\left(e^{-\Theta(\log n)} + e^{-c\frac{\log n}{100\alpha}}\right) \le \frac{1}{n^2} \tag{11}$$

By union bound over all s buckets, the signed sum of all small j in every bucket will be at most $||y||_p/100$. Therefore, for all i,

1. in each bucket i without large indices j, $|(PDy)_i| \le ||y||_p/100$
2. in each bucket i with one large index j, $|(PDy)_i| = |\sigma_j(Dy)_j| \pm ||y||_p/100$

and no bucket has more than one large j as shown in the perfect hashing assumption above. Conditioning on $||Dy||_\infty \in [||y||_p/10^{1/p}, ||y||_p\cdot 10^{1/p}]$,

$$\frac{||y||_p}{10^{\frac{1}{p}}} - \frac{||y||_p}{100} \le ||PDy||_\infty \le 10^{\frac{1}{p}}\cdot||y||_p + \frac{||y||_p}{100} \tag{12}$$

Therefore, it is reasonable to use $||PDy||_\infty$ as an estimate for $||y||_p$. The total space used is $s = O(n^{1-2/p}\log n)$, i.e. $O(n^{1-2/p}\log^2 n)$ bits. This space complexity still holds even when considering the pseudorandom generation of matrix P, see [1].

2 Heavy Hitters

l_1 **guarantee**: output a set containing all items j for which $|x_j| \ge \phi||x||_1$, and the set should not contain any j with $|x_j| \le (\phi-\varepsilon)||x||_1$.

l_2 **guarantee**: output a set containing all items j for which $x_j^2 \ge \phi||x||_2^2$, and the set should not contain any j with $x_j^2 \le (\phi-\varepsilon)||x||_2^2$. This guarantee is much stronger: suppose $x = [\sqrt{n}, 1, \ldots, 1]$, $\sqrt{n}$ is an l_2-heavy hitter for constant ϕ and ε, but not an l_1-heavy hitter. Also, if $|x_j| \ge \phi||x||_1$, it means that $x_j^2 \ge \phi^2||x||_1^2 \ge \phi^2||x||_2^2$ as well.

References

[1] N. Nisan. Pseudorandom generators for space-bounded computations. In *Proceedings of the Twenty-Second Annual ACM Symposium on Theory of Computing*, STOC '90, page 204â212, New York, NY, USA, 1990. Association for Computing Machinery. doi:10.1145/100216.100242.