University of Wrocław: Algorithms for Big Data (Fall'19) 21/10/2019

Lecture 3: Sketches for L_p norms

Lecturer: *Przemysław Uznański* Scribes: *Wiktor Garbarek*

1 p-stable distributions

Definition 1. *A distribution $\mathcal{D}$, with mean 0, is called stable, if for $X_1, .., X_n \sim \mathcal{D}$ which are independent and $a_1, .., a_n \in \mathbb{R}$ there is $\sum_i a_i X_i = b \cdot Z$ for some $b \in \mathbb{R}$ and $Z \sim \mathcal{D}$.*

Definition 2. *A distribution D is p-stable if is stable and coefficient b from previous definition satisfies*

$$b = \Big(\sum_i |x_i|^p\Big)^{1/p}$$

Remark 3. *(Zolotarev, 1986) p-stable distribution exists if and only if $0 < p \leq 2$.*
For $p \in \{\frac{1}{2}, 1, 2\}$ we know closed form formulas, e.g.:

- *Normal distribution, that is $f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$, is 2-stable.*
- *Cauchy distribution, that is $f(x) = \frac{1}{(1+x^2)\pi}$, is 1-stable.*
- *Lévy distribution is $\frac{1}{2}$-stable.*

Remark 4. *Except for $p = 2$, those distributions are heavy tailed, that is $\mathbb{E}[|\mathcal{D}|] = \infty$ and $\mathbb{E}[\mathcal{D}^2] = \infty$ (this has to be, as by Central Limit Theorem distributions with finite moments cannot be stable, unless its normal distribution).*

2 Sketches for L_p norm [Ind00]

Pick random coefficients $r_i \sim \mathcal{D}_p$ for $i = 1..n$, where $\mathcal{D}_p$ is a p-stable distribution. Then

$$Z = \sum_i x_i r_i$$

is a sketch of vector $\mathbf{x}$, since $Z \sim |\mathbf{x}|_p \mathcal{D}_p$. We will of course run many parallel instances of sketching process (as usual).

Remark 5. *Our sketches are linear functions, so linear combination of sketches is also a sketch.*

Whenever update (x_i, c_i) comes, to maintain the sketch we compute $Z := Z + c_i \cdot r_{x_i}$
Challenges:

- How to draw random values from p-stable distribution?
- How to extract the result?
- How much independence is required?

2.1 Drawing from p-stable

- $p = 1$: If $U \sim \mathcal{U}(0,1)$ then $\tan \pi(U - \frac{1}{2})$ is distributed with the Cauchy distribution
- $p = 2$: We can use Box-Muller transformation. If $U, V \sim \mathcal{U}(0,1)$ and are iid, then $\sqrt{-2\ln U} \cdot \cos 2\pi V$ is distributed as a normal distribution.
- $p \in (0,2)$ and $p \neq 1$: In this case p-stable distribution can be simulated by method derived by Chambers, Mallows and Stuck (1976). If $U, V \sim \mathcal{U}(0,1)$ and are iid and let's set $\Theta(U) = \pi(U - \frac{1}{2})$. Then
$$\frac{\sin p\Theta(U)}{\cos^{\frac{1}{p}} \Theta(U)} \left(\frac{\cos\left(\Theta(U) \cdot (1-p)\right)}{-\ln V} \right)^{\frac{1-p}{p}}$$
is distributed as a p-stable distribution.

2.2 Extracting the result via median

Recall $Z \sim |\mathbf{x}|_p \mathcal{D}_p$ Expected value is useless, since it is infinite, so let's consider median:
$$\text{median}(|Z|) \sim |\mathbf{x}|_p \cdot \text{median}(|\mathcal{D}_p|)$$

How to extract median of a distribution X (on $\mathbb{R}_+$)?

Let F be CDF of distribution X. Let's also sample k values $x_1, \ldots, x_k \sim X$, and output median$(x_1, \ldots, x_k)$. By Chernoff bound, if we have $k = O(\log(\frac{1}{\delta})/\varepsilon^2)$, then
$$F^{-1}(\frac{1}{2} - \varepsilon) \leq \text{median}(x_1, \ldots, x_k) \leq F^{-1}(\frac{1}{2} + \varepsilon)$$

If F' is not too flat around $F^{-1}(\frac{1}{2})$ (i.e. median), then $F^{-1}(\frac{1}{2} \pm \varepsilon)$ are actually $1 \pm C \cdot \varepsilon$ approximations of median (required $F'(x) \geq \frac{1}{C}$ in a given range). This becomes an issue when $p \to 0$, but we don't have to care for constant p.

Estimator: $\frac{\text{median}(|Z|)}{\text{median}(|D_p|)} = \frac{\text{median}(|Z_1|, \ldots, |Z_k|)}{\text{median}(|D_p|)}$

2.3 Geometric mean estimator [Li08]

Use geometric mean as an estimator.

Output: $(\Pi_{i=1}^{k} |Z_i|)^{1/k}/\alpha$, where $\alpha = e^{\mathbb{E}(\ln |D_p|)}$, and $k = O(\log(\frac{1}{\delta})/\varepsilon^2)$.

2.4 Independence

To use p-stability we assumed full n-wise independence - this means in theory huge space in streaming applications, and there is no easy workaround. We present two workarounds that are theoretic in nature and require very sophisticated machinery.

2.4.1 First workaround [Ind00]

Apply Nisan's pRNG which is able to:

- Store its seed/local state in a polylog(n) space
- Extract its next random number in a small working space and small time
- "Fool" small space (arbitrary) computations as if they were given perfect randomness

2.4.2 Second workaround [KNPW11]

Prove that k-wise independence is enough, for

$$k = \mathcal{O}(\log \frac{1}{\varepsilon} \log \log \log \frac{1}{\varepsilon})$$

3 L_p sketching for $p > 2$

3.1 2-pass Algorithm

Algorithm (only for sequence of values, not for turnstile):

1. Pick uniformly and at random $i \in [n]$
2. In the first pass: Select $x = x_i$
3. In the second pass: Compute $f_x = |\{j : x_j = x\}|$
4. return output $Y = n(f_x)^{p-1}$

$$E[Y] = \frac{1}{n} \sum_i n(f_{x_i})^{p-1} = \sum_x (f_x)^{p-1} f_x = \sum_x (f_x)^p = F_p$$

$$E[Y^2] = \frac{1}{n} \sum_i n^2 (f_{x_i})^{2p-2} = n \sum_x (f_x)^{2p-2} f_x = nF_{2p-1}$$

Claim 6. *Following inequality holds*

$$nF_{2p-1} \leq m^{1-1/p}(F_p)^2$$

Proof.

$$nF_{2p-1} = n\|f\|_{2p-1}^{2p-1} \leq n\|f\|_p^{2p-1} = \|f\|_1 \|f\|_p^{2p-1} \leq m^{1-1/p}\|f\|_p \|f\|_p^{2p-1} = m^{1-1/p}\|f\|_p^{2p} = m^{1-1/p} F_p^2$$

For $\|f\|_1 \leq m^{1-1/p}\|f\|_p$ and $\|f\|_{2p-1} \leq \|f\|_p$ see Wikipedia [1]. □

3.2 1-pass Algorithm

How to pick random i in a stream of unknown length?

1. Initialize $r \leftarrow 0$.
2. At i-th step of algorithm: with probability $\frac{1}{i}$ we set $x \leftarrow x_i$ and $r \leftarrow 0$
3. If $x_i = x$ then $r \leftarrow r + 1$
4. Return output $Y' = n(r^p - r^{p-1})$

$$E[Y'] = \sum_x \sum_{r \leq f_x} (r^p - r^{p-1}) = \sum_x (f_x)^p = F_p Y' \leq npr^{p-1} \leq np(f_x)^{p-1} = pY$$

$$E[(Y')^2] \leq p^2 E[Y^2] = p^2 m^{1-1/p}(F_p)^2$$

And thus the number of parallel runs should be $\mathcal{O}\left(\frac{p^2 m^{1-\frac{1}{p}}}{\varepsilon^2}\right)$

[1] https://en.wikipedia.org/wiki/Lp_space

Fact 7. *There is a lower bound for size of sketches for L_p, $p > 2$: $\Omega(m^{1-\frac{2}{p}}/\varepsilon^2)$ [BJKS04]*

Fact 8. *Best complexity of sketching is $\mathcal{O}(m^{1-\frac{2}{p}}/\varepsilon^2)$ – achieved in turnstile model. [IW05]*

References

[BJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[Ind00] Piotr Indyk. Stable distributions, pseudorandom generators, embeddings and data stream computation. In *41st Annual Symposium on Foundations of Computer Science, FOCS 2000, 12-14 November 2000, Redondo Beach, California, USA*, pages 189–197, 2000.

[IW05] Piotr Indyk and David P. Woodruff. Optimal approximations of the frequency moments of data streams. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 202–208, 2005.

[KNPW11] Daniel M. Kane, Jelani Nelson, Ely Porat, and David P. Woodruff. Fast moment estimation in data streams in optimal space. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 745–754, 2011.

[Li08] Ping Li. Estimators and tail bounds for dimension reduction in $l_\alpha (0 < \alpha \leq 2)$ using stable random projections. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2008, San Francisco, California, USA, January 20-22, 2008*, pages 10–19, 2008.