CS 395T: Sublinear Algorithms Fall 2016

Lecture 11 — 29th September, 2016

Prof. Eric Price *Scribe: Vatsal Shah, Yanyao Shen*

1 Overview

In previous lectures, we have shown how to estimate l_2 norm using AMS-sketch and how to estimate number of distinct elements. As a result, we get $O(\frac{1}{\epsilon^2})$ for l_2 and $O(\log^c n)$ for l_0 respectively.

In this lecture, we will estimate pth moment, where $p \in (0, 2)$ and show that it is doable in $O(\log^c n)$ space. On the other hand, when $p > 2$, linear sketches require $\Omega(n^{1-2/p})$ words, and is not doable in $poly \log n$ space.

2 A special case when $p = 1$

First, recall the algorithm for $p = 2$.

- Select $v \in \mathbb{R}^n$, where $v_i \sim N(0, 1)$, then $\langle v, x\rangle \sim N(0, \|x\|^2)$
- Take $\langle v_1, x\rangle, \langle v_2, x\rangle, \langle v_3, x\rangle, \cdots$: samples at $N(0, \|x\|^2)$ and then estimate. This requires $O(\frac{1}{\epsilon^2} \log \frac{1}{\delta})$ samples.

For $p = 1$, instead of choosing samples from Gaussian distribution, we use Cauchy distribution.

2.1 Cauchy Distribution Basics

In this section, we will look at the definitions and properties of Cauchy distribution:

- Standard form of Cauchy Distribution:

$$p(x) = \frac{1}{\pi} \frac{1}{x^2 + 1} \tag{1}$$

- General Form: Cauchy Distribution with scale factor γ

$$p(x) = \frac{1}{\pi\gamma} \frac{1}{\left(\frac{x}{\gamma}\right)^2 + 1} \tag{2}$$

- **Claim:** If $X_1 \sim Cauchy(\gamma_1)$, $X_2 \sim Cauchy(\gamma_2)$, then $X_1 + X_2 \sim Cauchy(\gamma_1 + \gamma_2)$

Proof: The Fourier Transform of standard Cauchy Distribution is given as follows:

$$\begin{aligned}\mathcal{F}_x(t) &= \mathbb{E}[cos(2\pi tx)] \\ &= \int_{-\infty}^{\infty} \exp(2\pi itx)\frac{1}{\pi(1+x^2)}dx \\ &= \exp(-2\pi|t|)\end{aligned}$$

Inverse Fourier Transform of standard Cauchy Distribution is:

$$\begin{aligned}\mathcal{F}_x^{-1}(t) &= \int_{-\infty}^{\infty} \exp(2\pi itx)\exp(2\pi|t|)dx \\ &= \int_{-\infty}^{0} \exp(2\pi itx - 2\pi t)dt + \int_{0}^{\infty} \exp(2\pi itx + 2\pi t)dt \\ &= \frac{1}{2\pi(1-ix)} + \frac{1}{2\pi(1+ix)} \\ &= \frac{1}{\pi(1+x^2)} = p(x)\end{aligned}$$

We can similarly show that the Fourier transform of $\mathcal{F}_{Cauchy(\gamma)}(t)$ is $\exp(-2\pi|t|\gamma)$.

Thus, if $X_1 \sim Cauchy(\gamma_1)$, $X_2 \sim Cauchy(\gamma_2)$, then:

$$\begin{aligned}\mathcal{F}_{X_1+X_2}(t) = \mathcal{F}_{X_1}(t)\mathcal{F}_{X_2}(t) &= \exp\left(-2\pi|t|(\gamma_1+\gamma_2)\right) \\ &= \mathcal{F}_{Cauchy(\gamma_1+\gamma_2)}(t)\end{aligned}$$

Now, if we have $X_1, X_2, \ldots, X_n \sim Cauchy(1)$, then $\dfrac{X_1 + X_2 + \ldots X_n}{n} \sim Cauchy(1)$.

This seemingly violates the law of large numbers which states that the distribution of the average of n random variables approaches that of a Gaussian as n increases. The law of large numbers is only applicable for random variables that have a finite expectation which does not hold for Cauchy random variables.

- $p - stable$ **distribution**:
  Let $p > 0$ be a real number. A probability distribution $\mathcal{D}$ is a $p - stable$ distribution if $\forall a_1, a_2, \ldots a_n$ and $X_1, \ldots X_n$ $\mathcal{D}$ are independently chosen, $\sum_i a_i X_i$, $\bar{a}X$ have the same distribution and $X \sim D$ and $\bar{a} = \|a\|_p$

  $p - stable$ distribution are typically defined $0 < p < 2$ and do not exist for $p > 2$

  Using the above definition, we can easily show that Cauchy is a $1-$stable distribution while Gaussian is a $2-$stable distribution.

2.2 Algorithm for $p = 1$

Here is the algorithm.

- For $i = 1, \cdots, m$, sample $A_{i,:}$'s elements from $Cauchy(\gamma = 1)$ distribution.

- Then, $y_i = A_{i,:}x$ follows distribution $Cauchy(\gamma = \|x\|_1)$, according to properties in section 2.1.
- Store $y_1, \cdots, y_m$.
- Let the median of $|y_i|$ be our estimator for $\|x\|_1$.

Question: Given m samples of Cauchy with unknown γ, how well can we estimate γ?

Use median. Then, how does median of $|y_i|, i \in [n]$ behave?

Claim: the median of $|y_i|, i \in [n] \to c\gamma$.

- How do we know c?
- How fast?

To answer the first question, we can compute c directly. c should satisfy:

$$\begin{aligned} \mathbb{P}(|y_i| < c\gamma) &= \frac{1}{2} \\ \Rightarrow 2\int_0^{c\gamma} \frac{1}{\pi\gamma((\frac{x}{\gamma})^2 + 1)} dx &= \frac{1}{2} \\ \Rightarrow 2\int_0^{c} \frac{1}{\pi(x^2 + 1)} dx &= \frac{1}{2} \\ \Rightarrow \frac{2}{\pi} \arctan x|_0^c &= \frac{1}{2} \\ \Rightarrow \arctan c &= \frac{\pi}{4} \Rightarrow c = 1. \end{aligned}$$

To answer the second question, we use the same methodology as in Lecture 7. Notice that the following is correct

$$\mathbb{P}(|y_i| > (1+\epsilon)c\gamma) = \frac{1}{2} - \Omega(\epsilon) \tag{3}$$

Then, using Chernoff bound (refer to Lecture 7), we know that a (ϵ, δ) estimate requires $m = O(\frac{1}{\epsilon^2} \log \frac{1}{\delta})$.

Here is an intuitive explanation of the result for m. For the probability density distribution f of $|y_i|$, the density at $x = 0$ is constant, and so does the density at the median point x_{med} where $\int_0^{x_{med}} f(x)dx = 1/2$. Now, if you have m balls and draw them randomly according to distribution f, then, the total number of balls that lie less than x_{med} are $\frac{m}{2} + O(\sqrt{m})$. Now, for example, suppose we have $\frac{m}{2} - \sqrt{m}$ that fall below x_{med}, then, in order to find the median ball, we need to move slightly right from x_{med} to cover another $\sqrt{m}$ balls. Notice that in total we have m balls, that is saying, the increased mass due to moving x_{med} to the right should be approximately $1/\sqrt{m}$. Furthermore, since the density at x_{med} is constant, the distance moving right is $O(1/\sqrt{m})$. Therefore, in order to achieve an ϵ approximation, $m = O(\frac{1}{\epsilon^2})$ is required.

2.3 Other p in the interval $(1, 2)$

For other p, we can use the similar methodology as we did when $p = 1$, i.e., we build a p-stable distribution $\mathcal{D}$. If all the elements of $A_{i,:}$ are sampled from $\mathcal{D}(1)$, then, one can show that y_i follows distribution $\mathcal{D}(\|x\|_p)$, and all the rest parts of the algorithm remains the same.

3 Lower Bound for $p > 2$

Proving the lower bound for $p > 2$ will be shown in next class. We will use ideas from information theory (Shannon-Hartley theorem) to prove it.