CS 395T: Sublinear Algorithms, Fall 2020 September 10, 2020

Lecture 5: Lower bounds; concentration inequalities

Prof. Eric Price *Scribes: Aaron Lamoreaux, Stanley Wei*

NOTE: THESE NOTES HAVE NOT BEEN EDITED OR CHECKED FOR CORRECTNESS

1 Overview

In the last lecture we discussed lower bounds for the indexing problem and how to generate lower bounds for streaming algorithms. Today we will finish lower bounds for indexing. Then we will discuss some concentration inequalities.

2 Lower bound for indexing

Recall the indexing problem. Alice is given $x \in \{0,1\}^n$, Bob is given $i \in [n]$, and when Alice gives Bob a message m, and Bob needs to output x_i correctly with some constant probability, say 99%.

Claim 1. *The indexing problem requires $\Omega(n)$ communication complexity.*

Here's the intuition behind this. Bob runs this on all i, not just his input, in which Bob would obtain some $\hat{x} \in \{0,1\}^n$. In particular, if Alice's input is uniform on some set C (which we'll define later), then $E_{x \in C}[||x - \hat{x}||] \leqslant \frac{n}{100}$, since we want $Pr_{x \in C, i \in [n]}[\hat{x_i} \neq x] \leqslant \frac{1}{100}$.

Let's choose C to be of large size $2^{\Omega(n)}$, such that $\min_{x-y \in C, x \neq y} ||x-y|| \geqslant \frac{n}{10}$. We now proceed with the following algorithm to choose the points x_i for C greedily: first pick x_1, then pick any x_2 "far" from x_1, then pick x_3 "far" from x_1, x_2, etc. (*Far* in this case refers to the distance being $\geqslant \frac{n}{10}$.)

Claim 2. *This algorithm will find $2^{\Omega(n)}$ points for C.*

Proof. To see this, note that each x_i invalidates $\binom{n}{\frac{n}{10}} + \binom{n}{\frac{n}{10}-1} + \binom{n}{\frac{n}{10}-2} + ... \approx \binom{n}{\frac{n}{10}} \leqslant (10e)^{\frac{n}{10}}$ points (where the inequality comes from Stirling's). Furthermore, since there are 2^n total possible x_i, we'll find at least $\dfrac{2^n}{(10e)^{\frac{n}{10}}} \geqslant \dfrac{2^n}{(2^5)^{\frac{n}{10}}} = 2^{\frac{n}{2}}$ points. □

Now that we have a set C, Alice an $x \in C$, and Bob chooses $\hat{x} \in \{0,1\}^n$. Now, Bob rounds $\hat{x}$ to the nearest $\tilde{x} \in C$, and this implies $\tilde{x} = x$ if $||x - \hat{x}|| \leqslant \frac{n}{20}$. Therefore, $Pr[\tilde{x} \neq x] \leqslant Pr[||x - \hat{x}|| \geqslant \frac{n}{20}] \leqslant \dfrac{E[||x-\hat{x}||]}{\frac{n}{20}} \leqslant \dfrac{n/100}{n/20} = \frac{1}{5}$. From this, we see that 80% of the time, we correctly identify this vector.

Here, it's worth mentioning two famous "codes": the Hamming codes and the Shannon codes. The *former requires the minimum distance between two elements to be small, while the latter requires correcting random errors. Set C above is an example of a "code."*

We now consider the entropy functions. In particular, for the message m, we have

$$H(m)(x;m) \geqslant I(x;\tilde{x}) = H(x) - H(x|\tilde{x}) \geqslant Pr[x = \tilde{x}] * \log|C| - 1 \geqslant 0.8 \cdot \frac{n}{2} - 1 = \frac{2}{5}n - 1 = \Omega(n)$$

where the second inequality comes from Fano's inequality.

Lemma 3. *(Fano's inequality)*

$$H(x|\tilde{x}) \leqslant 1 + Pr[x \neq \tilde{x}] \cdot log|C|$$

Proof. We have $H(x|\tilde{x}) \leqslant H(x, \mathbf{1}_{x=\tilde{x}}|\tilde{x}) = H(\mathbf{1}_{x=\tilde{x}}|\tilde{x}) + H(x|\tilde{x}, \mathbf{1}_{x=\tilde{x}}) \leqslant 1 + Pr[x \neq \tilde{x}] \cdot log|C|$, where the last inequality follows because $H(x|\tilde{x}) \leqslant 1$ and $H(x|\tilde{x}, \mathbf{1}_{x=\tilde{x}}) \leqslant Pr[x \neq \tilde{x}] \cdot log|C|$. □

Combining the results, we see that we need $\geqslant \frac{2}{5}n - 1$ bits of communication to solve indexing 99% of the time. Note however that from our analysis, our constant is very loose.

3 Concentration Bounds

Recall the definition of Markov's:

Theorem 4 (Markov's inequality). *If $X \geq 0$, then*

$$\Pr[X \geq t] \leq \frac{E[x]}{t}.$$

Recall the definition of Chebyshev's:

Theorem 5 (Chebyshev's inequality). *If $\mu = E[x]$ and $\sigma^2 = E[(x-\mu)^2]$ then*

$$\Pr[|x - \mu| \geq t] \leq \frac{E[(x-\mu)^2]}{t^2} = \frac{\sigma^2}{t^2}.$$

Consider the following example. You are given n random variables $X_i \in [0,1]$ independently distributed with $E[X_i] = \mu$. Consider the sum

$$X = \sum X_i.$$

Analyzing the mean and variance gives that

$$E[X] = n\mu$$

and (since $\text{Var}(X_i) \leq 1$

$$\text{Var}(X) = \sum \text{Var}(X_i) \leq n.$$

Applying Chebyshev's gives us that

$$\Pr[|X - n\mu| \geq t] \leq \frac{n}{t^2}.$$

Question 6. *At what points is this bound from Chebyshev's really weak relative to the actual value?*

Consider that if the coins are balanced, then we can compute that

$$\Pr[X = n] \leq \frac{n}{(n/2)^2} = \frac{4}{n}.$$

In reality, the actual value of $\Pr[X = n]$ is

$$\Pr[X = n] = \frac{1}{2^n}.$$

In general, the Chebyshev bound is really lose at the tails of the distribution.

Question 7. *Consider how we derived Chebyshev's inequality, can you find inequalities for higher moments?*

Let's consider apply Markov's to a higher exponent to get

$$\Pr[|X - n\mu| \geq t] = \Pr[|X - n\mu|^4 \geq t^4] \leq \frac{E[(X - n\mu)^4]}{t^4}.$$

We can bound (using a similar method to a previous lecture when analyzing distinct elements)

$$E[(X-n\mu)^4] = E[(\sum(X_i-\mu))^4] = \sum E[(X_i-\mu)^4]+\sum_{i<j} 6E[(X_i-\mu)^2]\cdot E[(X_j-\mu)^2] \leq 3n^2-2n \leq 3n^2.$$

From Chebyshev's we get that

$$\Pr[|X - n\mu| \geq s\sqrt{n}] \leq \frac{1}{s^2}$$

and from the 4th moment that

$$\Pr[|X - n\mu| \geq s\sqrt{n}] \leq \frac{3}{s^4}.$$

Question 8. *What do we get for higher moments? I.e. what is*

$$\frac{E[(x - n\mu)^k]}{t^k}?$$

Doing similar analysis to before we get that

$$\Pr[|X - n\mu| \geq s\sqrt{n}] \leq \frac{k^{k/2}}{s^k}, \quad \forall k \text{ even integer.}$$

If we pick $k = \frac{s^2}{4}$, we get that

$$\Pr[|X - n\mu| \geq s\sqrt{n}] \leq 2^{-s^2/4}.$$

If we check the previous tail we now get that

$$\Pr[X = n] \leq 2^{-n/16}.$$

This is still a really loose constant in the bound, however this is much closer to the 2^{-n} truth that we expected. However our analysis is a bit sketch since we ignored other terms and made assumptions about k being even.

4 Moment generating functions

Define a random variable X, that

$$\Phi_X(\lambda) = E\left[e^{\lambda(X-E[X])}\right].$$

We can compute that

$$e^{\lambda X} = 1 + \lambda X + \frac{(\lambda X)^2}{2} + \cdots$$

which is a distribution over X. As λ increases though, the larger moments become more important.

Using moment generating functions, we an compute the following inequality

$$\Pr[X - n\mu \geq t] = \Pr[e^{\lambda(X-n\mu)} \geq e^{\lambda t}] \leq \frac{E[e^{\lambda(X-n\mu)}]}{e^{\lambda t}} = \frac{\Phi_X(\lambda)}{e^{\lambda t}}.$$

We can break this up into a product to get that

$$\Phi_X(\lambda) = \prod_i E[e^{\lambda(X_i-\nu)}] = \prod_i \Phi_{X_i}(\lambda).$$

Lemma 9 (Hoeffding's Lemma)**.** *If $Y \in [0,1]$, mean μ, then*

$$\Phi_Y(\lambda) \leq e^{-\lambda^2/8}.$$

This implies that

$$\Pr[X - n\mu \geq t] \leq e^{n\lambda^2/8} \cdot e^{-\lambda t}, \quad \forall \lambda > 0.$$

We can complete the square and pick $\lambda = 4t/n$ in order to get that

$$\Pr[X - n\mu \geq t] \leq e^{-2t^2/n}.$$

which is equivalent to

$$\Pr[X - n\mu \geq s\sqrt{n}] \leq e^{-2s^2}.$$

This is called the Chernoff Bound.

In particular, when we apply this to the unbiased coin example we get that

$$\Pr[X = n] \leq e^{-2(n/2)^2/n} = e^{-n/2}$$

This bound only gives us an upper bound, but for $\lambda < 0$, we can get a lower bound to conclude that

$$\Pr[|x - n\mu| \geq t] \leq 2e^{-2t^2/n}.$$

For a really biased coin, Bernstein bounds are very useful (which we will possibly cover later).