CS 15-859: Algorithms for Big Data Fall 2017

Lecture 9b — November 2

Prof. David Woodruff *Scribe: Ruosong Wang*

1 Some Other Distance Measures

Definition (Kullback-Leibler Divergence)**.** For two discrete probability distributions P and Q, the Kullback-Leibler Divergence from Q to P is defined to be

$$\mathsf{KL}(P, Q) = \sum_i P_i \log \left(\frac{P_i}{Q_i} \right).$$

Notice that the KL-divergence is asymmetric and can be infinite.

Definition (Jensen-Shannon Distance)**.** For two discrete probability distributions P and Q, the Jensen-Shannon Distance between P and Q is defined to be

$$\mathsf{JS}(P, Q) = \frac{1}{2} \left(\mathsf{KL}(P, R) + \mathsf{KL}(Q, R) \right),$$

where $R = (P + Q)/2$ is the average distribution.

Notice that the JS distance is symmetric and always finite. Furthermore, we can use the JS distance to lower bound information. Formally, the following inequality holds.

Proposition 1. Suppose X and B are (possibly dependent) random variables and B is a uniform bit, then we have

$$I(X; B) \geq \mathsf{JS}(X \mid B = 0, X \mid B = 1).$$

Meanwhile, we have the following relations between distance measures.

Proposition 2. For two discrete probability distributions P and Q,

1. $\mathsf{JS}(\mathsf{P}, \mathsf{Q}) \geq h^2(P, Q)$;
2. $h^2(P, Q) \geq D_{TV}^2(P, Q)$.
3. If one can distinguish P from Q with a sample with probability $\frac{1}{2} + \delta/2$, then

$$D_{TV}(P, Q) \geq \delta.$$

2 Lower Bound of Communication Complexity of INDEX

2.1 Problem Setting of INDEX

A randomized 1-way communication protocol of INDEX is consisted of the following stages.

1. Alice receives a binary string $x \in \{0,1\}^n$, Bob receives an integer $j \in \{1, 2, \ldots, n\}$.
2. Alice sends a single message M to Bob. Here M depends only on x and Alice's random coins.
3. Bob outputs b, which is his guess of x_j. Here b depends only on M and Bob's random coins.

We say a protocol is correct if for any $x \in \{0,1\}^n$ and $j \in [n]$,

$$\mathbb{P}[b = x_j] \geq \frac{2}{3}.$$

2.2 The Lower Bound

Theorem 1. *For a correct randomized 1-way communication protocol of* INDEX*, $|M| = \Omega(n)$.*

Proof. Consider a uniform distribution on all possible $x \in \{0,1\}^n$. For any j, denote x'_j to be Bob's guess of x_j. Since $\mathbb{P}[x_j = x'_j] \geq \frac{2}{3}$ and $x_j \to M \to x'_j$ is a Markov chain, by Fano's inequality, we have for all j,

$$H(x_j \mid M) \leq H\left(\frac{2}{3}\right) + \frac{2}{3}(\log_2 2 - 1) = H\left(\frac{1}{3}\right).$$

By the chain rule,

$$\begin{aligned} I(x; M) &= \sum_i I(x; M \mid x_{<i}) \\ &= \sum_i H(x_i \mid x_{<i}) - H(x_i \mid M, x_{<i}). \end{aligned}$$

Since different coordinates of x are independent, we have $H(x_i \mid x_{<i}) = 1$. We also have $H(x_i \mid M, x_{<i}) \leq H(x_i \mid M)$ as conditioning will not increase entropy. Thus,

$$|M| \geq H(M) \geq I(x; M) \geq n - \sum_i H(x_i \mid M) \geq n - H\left(\frac{1}{3}\right) n.$$

■

Remark 1. The lower bound holds for any constant success probability. More specifically, for a protocol which succeeds with probability at least $1 - \delta$, we have

$$|M| \geq (1 - H(\delta))n.$$

Moreover, this lower bound is tight.

Remark 2. The lower bound holds even if Alice and Bob have unlimited amount of public random coins.

Remark 3. The lower bound holds even for the Augmented-INDEX problem, which is a (seemingly) easier version of INDEX. In the Augmented-INDEX problem, Bob receives not only j but also $x_1, x_2, \ldots, x_{j-1}$. The lower bound still applies since by Fano's inequality, if the protocol succeeds with probability at least $1 - \delta$, then we have

$$H(x_j \mid M, x_1, x_2, \ldots, x_{j-1}) \leq H(\delta).$$

Thus,

$$|M| \geq H(M) \geq I(x; M) \geq n - \sum_i H(x_i \mid M, x_1, x_2 \ldots, x_{i-1}) \geq (1 - H(\delta))n.$$

2.3 Applications of the Lower Bound

2.3.1 Lower Bound of the Distinct-Elements Problem

The goal of the Distinct-Elements problem is that, given $a_1, a_2, \ldots, a_m \in [n]$ in the streaming model, count the (exact) number of distinct elements in the stream.

Theorem 2. *Any algorithm for* Distinct-Elements *has space complexity* $\Omega(n)$.

Proof. Suppose there exists an algorithm $\mathcal{A}$ for Distinct-Elements with $o(n)$ space complexity. We show that based on $\mathcal{A}$, we can construct a protocol for the INDEX problem with $o(n)$ communication complexity, which contradicts Theorem 1.

1. After Alice receives x, she runs $\mathcal{A}$ with input $i_1, i_2, \ldots, i_r$ for all i_j where $x_{i_j} = 1$. Then Alice sends a message M to bob, where M is the current state of $\mathcal{A}$. Since the space complexity of $\mathcal{A}$ is $o(n)$, the size of the message Alice sent is also $o(n)$.
2. After Bob receives j and M, he runs $\mathcal{A}$ with M as the initial state and j as the input.
3. Bob outputs 1 if the number of distinct elements remains unchanged after inputing j, and 0 otherwise.

To see the correctness of this protocol, if $x_j = 1$, then j is already in the stream, and the number of distinct elements will remain unchanged after Bob adds j into the stream. If $x_j = 0$, then the number of distinct elements will increase by 1 after Bob adds j into the stream. Thus, Bob always outputs the correct answer. ■

2.3.2 Lower Bound for Estimating Norms

We have shown in Lecture 7 that estimating 1-norm and 2-norm of the input stream can be solved with space complexity $O\left(\frac{\log n}{\varepsilon^2}\right)$. Here we show an $\Omega(\log n)$ lower bound for the space complexity of estimating p-norms (for $p > 0$) with constant approximation. For simplicity we assume the approximation ratio is at most 2.

Proof. Suppose there exists an algorithm $\mathcal{A}$ for estimating p-norm with $o(\log n)$ space complexity. We show that based on $\mathcal{A}$, we can construct a protocol for the Augmented-INDEX problem with $o(\log n)$ communication complexity when Alice receives $x \in \{0,1\}^{\log n}$ and Bob receives $j \in [\log n]$, which contradicts Theorem 1.

1. After Alice receives x, she runs $\mathcal{A}$ with input w where w is vector with a single coordinate equal to $\sum_{i=1}^{n} 10^{n-i} x_i$. Then Alice sends a message M to bob, where M is the current state of $\mathcal{A}$. Since the space complexity of $\mathcal{A}$ is $o(\log n)$, the size of the message Alice sent is also $o(\log n)$.
2. After Bob receives j, $x_1, x_2, \ldots, x_{j-1}$ and M, he runs $\mathcal{A}$ with M as the initial state and w' as the input, where w' is a vector with a single coordinate equal to $-\sum_{i=1}^{j-1} 10^{n-i} x_i$.

3. Bob outputs 1 if the final estimated norm is at least $\frac{1}{2} \cdot 10^{n-j}$ and 0 otherwise.

To see the correctness of this protocol, if $x_j = 1$, then the estimated norm is at least $\frac{1}{2} \cdot 10^{n-j}$. If $x_j = 0$, then the estimated norm is at most $2 \cdot 10^{n-j-1}$. Thus, Bob will always output the correct answer. ■

We can also prove an $\Omega\left(\frac{1}{\varepsilon^2}\right)$ lower bound for the space complexity of estimating norms based on the communication complexity of Gap-Hamming problem. In the Gap-Hamming problem, Alice receives a string $x \in \{0,1\}^n$ and Bob receives $y \in \{0,1\}^n$. It is guaranteed that $\Delta(x,y) > n/2 + \sqrt{n}$ or $\Delta(x,y) < n/2$, where $\Delta(\cdot,\cdot)$ denotes the Hamming distance. The goal is to distinguish these two cases.

In [2, 4, 3] it has been proved that the communication complexity of randomized 1-way communication protocol is $\Omega(n)$. In [1], Chakrabarti and Regev further proved that the same bound also holds for 2-way communication. Here we follow the approach in [3] and prove the communication complexity of Gap-Hamming based on the communication complexity of INDEX.

Suppose there exists a protocol $\mathcal{A}$ for Gap-Hamming with $o(n)$ space complexity. We show that based on $\mathcal{A}$, we can construct a protocol for the INDEX problem with $o(n)$ communication complexity, which contradicts Theorem 1.

1. Alice and Bob first agree on n independent random binary vectors $r^1, r^2 \ldots, r^n \in \{0,1\}^n$ by using their public random coins.
2. After Alice receives $x \in \{0,1\}^n$, she calculates $a \in \{0,1\}^n$ where for any $k \in [n]$,
$$a_k = \mathsf{Maj}_{i \text{ such that } x_i = 1} r_i^k,$$
where Maj denotes the majority function.
3. After Bob receives j, he calculates $b \in \{0,1\}^n$, where for any $k \in [n]$,
$$b_k = r_j^k.$$
4. Alice and Bob invokes the protocol $\mathcal{A}$ with a and b as input. If $\Delta(a,b) > n/2 + \sqrt{n}$ then they output 0. Otherwise they output 1.

To see the correctness of this reduction, we first state the following lemma.

Lemma 1. *For any $k \in [n]$,*
$$\mathbb{P}[a_k = b_k] = \begin{cases} \frac{1}{2} & \text{if } x_j = 0 \\ \frac{1}{2} + \Omega(1/\sqrt{n}) & \text{if } x_j = 1 \end{cases}.$$

With Lemma 1, the correctness of the reduction follows directly from an application of the Chernoff bound. The intuition behind Lemma 1, is that when $x_j = 0$, a_k and b_k are independent random bits, and thus the probability that a_k equals b_k is $\frac{1}{2}$. However, when $x_j = 1$, notice that b_k is the majority of at most n independent random bits, and one of these random bits is a_k in this case. Thus, the probability that a_k equals b_k will be larger than $\frac{1}{2}$ due to the bias.

To bound the probability that a_k equals b_k when $x_j = 1$, we assume b_k is the majority of $\hat{n}$[1] independent random bits (and one of them is a_k). W.l.o.g., we assume $\hat{n}$ is odd.

In this case, it is clear that

$$\mathbb{P}[a_k = b_k] - \mathbb{P}[a_k \neq b_k] = \binom{\hat{n}-1}{(\hat{n}-1)/2}\left(\frac{1}{2}\right)^{\hat{n}-1} = \Theta\left(\frac{1}{\sqrt{\hat{n}}}\right)$$

by Stirling's formula. Since $\hat{n} \leq n$,

$$\mathbb{P}[a_k = b_k] = \frac{1}{2} + \Theta(1/\sqrt{\hat{n}}) = \frac{1}{2} + \Omega(1/\sqrt{n}).$$

References

[1] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM Journal on Computing*, 41(5):1299–1317, 2012.

[2] Piotr Indyk and David Woodruff. Tight lower bounds for the distinct elements problem. In *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*, pages 283–288. IEEE, 2003.

[3] Thathachar S Jayram, Ravi Kumar, and D Sivakumar. The one-way communication complexity of hamming distance. *Theory of Computing*, 4(1):129–135, 2008.

[4] David Woodruff. Optimal space lower bounds for all frequency moments. In *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 167–175. Society for Industrial and Applied Mathematics, 2004.

[1] Also notice that $\hat{n} = |\{i \mid x_i = 1\}|$.