CS 15-859: Algorithms for Big Data Fall 2019

Lecture 10 — November 7, 2019

Prof. David Woodruff *Scribe: Chirag Gupts*

In last lecture we looked at the indexing problem and various lower bounds that it implies. We deduced lower bounds on the indexing problem using information theory concepts. We will continue discussing information theory in this lecture. The notable difference is that in the last lecture, we thought of the Alice's string X and Bob's index i being arbitrary, and the randomness was over the message being sent. Today we will also consider the randomness over the draw of (X, Y) assuming they come from a (joint) distribution.

1 General aspects of 1-Way communication

Recall the 1-way communication model. Alice has $X \in \{0,1\}^n$. Bob has $i \in [n]$. Alice sends a randomized message M to Bob and Bob must infer X_i. Suppose there is public knowledge of some shared random bits denoted as R. We are now interested in the entity $I(M; X|R)$. It can be bounded as follows:

$$\begin{aligned}
I(M; X|R) &= \sum_i I(M; X_i|X_{<i}, R) \\
&= \sum_i (H(X_i|X_{<i}, R) - H(X_i|X_{<i}, R, M)) \\
&\geq \sum_i (H(X_i|X_{<i}, R) - H(X_i|R, M)) \\
&= \sum_i I(M; X_i|R) \\
&= \sum_i (H(X_i|R) - H(X_i|M, R)) \\
&= n - \sum_i H(X_i|M, R).
\end{aligned}$$

In last lecture we saw Fano's inequality: if Bob can guess X_i with probability at least $1 - \delta$, then $H(X_i|M, R) \leq H(\delta)$. Then

$$CC_\delta(\text{Index}) \geq I(M; X|R) \geq n(1 - H(\delta)),$$

where $CC_\delta(\text{Index})$ is the communication complexity defined as the length of Alice's message to ensure that Bob can recover the bit with probability $1 - \delta$. We have seen a proof of the above fact in the previous lecture:

$$|M| \geq H(M) \geq H(M|R) \geq H(M|R) - H(M|X, R) = I(M; X|R).$$

These lower bounds we have discussed also apply if the protocol is only correct on average over X and i drawn independently from a uniform distribution. Under such a distribution, suppose that for half the fraction of $i \in [n]$ the probability that Bob outputs X_i when his input is i is $1 - 2\delta$ and for the other half Bob gets it right with probability 1. Our overall error guarantee for the indexing problem is still $1 - \delta$. To study this framework in detail we introduce distributional communication complexity.

2 Distributional communication complexity

In this section we introduce the 1-way communication problem for general strings x and y that Alice and Bob hold respectively (y need not be a single bit). We also allow for (x, y) to be drawn from a distribution μ. For some Boolean function $f : X \times Y \to \{0, 1\}$ Alice wants to send a short message to Bob so that he can output $f(x, y)$ with error probability at most δ. Allowing a distribution over (X, Y) leads to a framework that is more powerful than the indexing problem (for reasons discussed shortly). In particular, as we discuss in Section 2.2, to obtain a lower bound for the Gap_∞ (which can be used to obtain lower bounds for p-norm estimation) we need to consider this more general framework.

The μ-distributional complexity $D_\mu(f)$ is defined as the minimum communication cost of a protocol which outputs $f(x, y)$ with probability $2/3$ for $(x, y) \sim \mu$. Using Yao's minimax principle and strong duality, it is possible to deduce that the maximum communication complexity for the non-distributional problem (but with randomized message M) is the same as the maximum communication complexity over the distributional problem:

$$R(f) = \max_\mu D_\mu(f).$$

Note that the randomized guarantee considers randomness over bits drawn by Alice and Bob, whereas the μ-distributional guarantee is both over $(x, y) \sim \mu$ as well as the random bits of Alice and Bob.

Suppose now that μ is a product distribution $\mu_X \times \mu_Y$ (in other words X and Y are independent). Consider solving the problem

$$\max_{\mu=\mu_X\times\mu_Y} D_\mu(f).$$

Kremer, Nisan and Ron [2] showed that solving the above is equivalent to solving the indexing problem. Their argument is summarized in the following section.

2.1 Indexing is universal for product distributions

Going back to the indexing problem, let the support of X be $\{0, 1\}^n$ and that of Y be $[n]$. The 1-way communication problem can be written down as a Boolean function $f : X \times Y \to \{0, 1\}$. That is for every possibility of Alice's bit string $x \in \{0, 1\}^n$ and Bob's single bit $y \in [n]$, f evaluates to the y'th bit of x. We denote this Boolean function as a communication matrix A_f where the rows are indexed by Alice's string and the columns by Bob's index, and entry (x, y) is given by $f(x, y)$. Consider the following matrix where $n = 3$ and Alice's string distribution is supported on 6 out of the total 8:

$$A_f = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Consider the first two columns of the matrix and rows 2, 3, 4, and 6. These create the submatrix

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

which represents all possible outputs depending on X and Y. Clearly to represent all these states, the message must be at least 2 bits long. In general the VC dimension of A_f is the largest number of columns for which we can find a set of rows that contain all possibly values over these two columns. It can be shown that

$$\max_{\mu=\mu_X\times\mu_Y} D_\mu(f) = \Theta(\text{VC dimension of } A_f).$$

The indexing problem is universal for product distributions but there are limitations with using such product distribution based lower bounds. For example, we get the same $\Omega(n)$ communication lower bound for 1/3 as well as zero probability of error, which seems unsatisfying. For norm estimation, the indexing problem can be used to show a $\Omega(\log n + \epsilon^{-2} + \log(1/\delta))$ lower bound (see note below). The is comparable to the best known bound $\Omega(\epsilon^{-2}\log n \log(1/\delta))$. We know that this lower bound is tight for $(0, 2]$-norm estimation since we can use one of our $p-$stable $\log n/\epsilon^2$ sketches and repeat it $\log(1/\delta)$ times to get high probability correctness. However it is not tight for $p > 2$. To provide tight lower bounds in the $p > 2$ case, we introduce non-product distributions that are more general than indexing.

Note: ($\Omega(\log n + \epsilon^{-2} + \log(1/\delta))$ lower bound for norm-estimation). In the previous lecture, we have already seen $\Omega(\log n + \epsilon^{-2})$ lower bound for norm estimation. For the δ dependence we consider the framework of Indexing on Large Alphabets. Alice has $x \in \{0, 1\}^{n/\delta}$ with $\text{wt}(x) = n$ and Bob has $i \in [n/\delta]$. Bob wants to decide whether $x_i = 1$ with error probability at most δ. This problem has a lower bound of $\Omega(n \log(1/\delta))$ [1]. Let us see how this gives us a lower bound for norm-estimation. Suppose norm estimation was possible with failure probability $o(\log(1/\delta)$. We will embed an indexing problem in the norm estimation problem as follows. Since $\text{wt}(x) = 1$, Alice's vector is a unit vector, say e_j. Alice sends this vector to Bob with a message that has size $o(\log(1/\delta)$. Bob now updates the 'stream' with a unit vector generated from his index i: $-e_i$. Bob then uses the streamed message to estimate the norm of $e_j - e_i$. For any $p-$norm, if $j = i$ then $\|e_j - e_i\|_p = 0$, and if $j \neq i$ then $\|e_j - e_i\|_p = \sqrt[p]{2} > 1$. Since our $o(\log(1/\delta)$ allows for norm-estimation with error probability δ, it can be used to solve the original indexing problem, which would lead to a $o(\log(1/\delta)$ upper bound for indexing on large alphabets. This is a contradiction and hence the length of Alice's message must be at least $\Omega(\log(1/\delta)$.

2.2 Non-product distributions and the Gap_∞ problem

We have seen that

$$R(f) = \max_\mu D_\mu(f).$$

But it is possible that product distributions do not capture $R(f)$, that is

$$\max_\mu D_\mu(f) \gg \max_{\mu=\mu_X\times\mu_Y} D_\mu(f).$$

In particular for lower bounds on $p-$norm estimation, the product distribution is not enough. Suppose that $p = \infty$ and we want to estimate $\|x\|_\infty$ up to a multiplication factor of B in a stream. A sub-case of this problem is the $\text{Gap}_\infty(x, y)$ problem for which we will show lower bounds. Alice holds x and Bob holds y. We are guaranteed that either $\|x - y\|_\infty \leq 1$ or $\|x - y\|_\infty \geq B$. Observe that any $p-$norm estimation problem can solve the Gap_∞ problem by setting $B = \Omega(n^{1/p})$. Using non-product distributions, the 1-way communication framework can be used to prove a lower bound of $\Omega(n/B^2)$ for the Gap_∞ problem [3]. This would then lead to the tight $\Omega(n^{1-2/p})$ lower bound for p-norm-estimation problem.

To discuss this lower bound, we use the technique of direct sums (see next part of the lecture). Before that we discuss what happens to the information content of a message Π given a particular distribution μ (which need not be a product distribution μ of X and Y. Suppose Alice and Bob both hold single bits A and B and are trying to compute $A \wedge B$. This is a simple problem to solve for every A and B: Alice simply sends the message $Pi = A$ and Bob computes $A \wedge B$. It can be seen easily that if A and B are both independently and uniformly 0 or 1 then

$$I(\Pi; A, B) = 1.$$

Consider now a new distribution: $A, B \overset{\text{unif}}{\sim} \{(0,0), (0,1), (1,0)\}$. Now the mutual information is computed as follows:

$$\begin{aligned} I(\Pi; A, B) &= \underbrace{H(A,B)}_{\text{uniform distribution supported on 3 elements}} - H(A, B|\Pi) \\ &= \log_2(3) - \underbrace{H(A,B|\Pi=1)}_{=0}\mathbb{P}[\Pi = 1] - \underbrace{H(A,B|\Pi=0)}_{1}\mathbb{P}[\Pi = 0] \\ &= \log_2(3) - (2/3) \\ &\approx 0.77 < 1. \end{aligned}$$

Thus the mutual information can *decrease* when A, B follow a particular distribution, even though the message Π works for every possible choice of A and B.

References

[1] Rajesh Jayaram and David P Woodruff. Perfect lp sampling in a data stream. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 544–555. IEEE, 2018.

[2] Ilan Kremer, Noam Nisan, and Dana Ron. On randomized one-round communication complexity. Springer.

[3] Michael Saks and Xiaodong Sun. Space lower bounds for distance approximation in the data stream model. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 360–369. ACM, 2002.