CS 591 B1: Communication Complexity, Fall 2019

Problem Set 2

Due: 5:00PM, Friday, October 25, 2019.

Homework Policies:

- Submit your completed assignment by email to `mbun[at]bu[dot]edu`. Please include the string "CS591PS2" somewhere in your subject line.
- Solutions must be typeset, e.g., using LaTeX or Microsoft Word.
- To help your instructor calibrate the length and difficulty of future assignments, please include with each problem an estimate of how long it took you to solve it.
- You are encouraged to collaborate on the homework problems with each other in small groups (2-3 people). Collaboration may include brainstorming or exploring possible solutions together on a whiteboard, but should not include one person telling the others how to solve a problem. You must also write up the solutions independently (in your own words) and acknowledge your collaborators at the beginning of the first page.
- You may freely use without proof any results proved in class, in Mark's lecture notes posted on the class webpage, or in the main body of the texts assigned as reading. Note that this excludes results that appear in the texts as problems and exercises. You may, of course, use such results but you have to prove them first.
- You may read papers and other outside sources to help you solve these problems. If you do so, you must acknowledge these sources and write the solutions in your own words.
- Start early! The problems are presented roughly in the order of the course content they correspond to, so you may get started on the first few problems as soon as the assignment is released. Late assignments will receive credit only with prior permission of the instructor.

Problem 1 (Zero- and One-Sided Error Communication)**.** Our discussion in class has focused on randomized communication with two-sided error, but we could make other choices. For example, we could consider protocols that only make one-sided error. A randomized protocol Π computes a function f with one-sided error if for every (x, y) with $f(x, y) = 0$,

$$\Pr[\Pi(x, y) = 0] = 1$$

and for every (x, y) with $f(x, y) = 1$,

$$\Pr[\Pi(x, y) = 1] \geq \frac{1}{2}.$$

The $\mathbf{RP^{cc}}$ communication cost of a protocol is the length of the shortest public coin protocol computing f. We can similarly define the $\mathbf{coRP^{cc}}$ complexity by exchanging the roles of 0s and 1s above.

(a) What is the $\mathbf{RP^{cc}}$ communication complexity of the Equality function EQ_n? How about the $\mathbf{coRP^{cc}}$ communication complexity?

(b) We may also define randomized protocols with *zero* error, but which may with some probability output $\perp$ indicating "I don't know." A randomized protocol $\Pi : X \times Y \to \{0, 1, \perp\}$ computes a function f with zero error if for every (x, y),

$$\Pr[\Pi(x, y) \in \{f(x, y), \perp\}] = 1$$

and

$$\Pr[\Pi(x, y) = \perp] \leq \frac{1}{4}.$$

The $\mathbf{ZPP^{cc}}$ communication complexity of a protocol is length of the shortest public coin protocol computing f with zero error.

Show that if f has a fooling set of size s, then $\mathbf{ZPP^{cc}}(f) \geq \log s - O(1)$.

(c) Letting $\mathbf{RP^{cc}}, \mathbf{coRP^{cc}}, \mathbf{ZPP^{cc}}$ consist of sequences of functions with polylogarithmic complexity in each of these models respectively, show that $\mathbf{ZPP^{cc}} = \mathbf{RP^{cc}} \cap \mathbf{coRP^{cc}}$.

Problem 2 (Shearer's Lemma)**.** Shearer's Lemma is an important tool in information theory with lots of applications to combinatorics. It states that if $X = (X_1, \ldots, X_n)$ is a collection of random variables and $A \subseteq [n]$ is independent from X such that $\Pr[j \in A] \geq \varepsilon$ for every $j \in [n]$, then

$$\varepsilon H(X) \leq H(X_A|A),$$

where $X_A = (X_j)_{j \in A}$ is the subcollection of X of indices in A.

(a) Derive the subadditivity of entropy, i.e., $H(X_1 X_2 \ldots X_n) \leq H(X_1) + \cdots + H(X_n)$, as a special case of Shearer's Lemma.

(b) Here is an application to counting satisfying assignments to boolean functions. Let $f_1, \ldots, f_m : \{0,1\}^n \to \{0,1\}$ be boolean functions, and let $F(x) = \wedge_{i=1}^m f_i(x)$. Suppose $\Pr_{x\sim U}[f_i(x) = 1] = p$ for every i where U is the uniform distribution over $\{0,1\}^n$. If the f_i's depend on *disjoint* sets of variables, i.e., each f_i depends on some subset $S_i \subseteq [n]$ of variables and $S_i \cap S_j = \emptyset$ for all $i \neq j$, then by independence we have

$$\Pr_{x\sim U}[F(x) = 1] = \prod_{i=1}^{m} \Pr_{x\sim U}[f_i(x) = 1] = p^m.$$

Prove a similar result when each variable x_j is allowed to appear in exactly k of the subsets S_i. Namely, suppose each f_i depends only on the variables in $S_i \subseteq [n]$ and $\#\{i \in [m] : j \in S_i\} = k$ for every $j \in [n]$. Show that

$$\Pr_{x\sim U}[F(x) = 1] \leq p^{m/k}.$$

Hint: Apply Shearer's Lemma with X being uniform over the satisfying assignments to F.

Problem 3 (Internal vs. External Information)**.** Show that if μ is a product distribution, i.e., A and B are independent when $(A, B) \sim \mu$, then $IC_\mu^{\text{ext}}(\Pi) = IC_\mu(\Pi)$ for every protocol Π.

Problem 4 (Correlated Sampling Revisited)**.** Consider the following correlated sampling problem. Let π be a distribution over X which decomposes as a product of two functions $p \cdot q$. Suppose Alice holds p and Bob holds q, and moreover, that they hold estimates q' and p' of each others' functions such that

1. $p \cdot q'$ and $p' \cdot q$ are valid probability mass functions, i.e., they take values in $[0, 1]$ and sum to 1 and
2. There is a known parameter $M > 0$ such that $p(x) \leq Mp'(x)$ and $q(x) \leq Mq'(x)$ for every $x \in X$.

Show that there is a public coin protocol that allows Alice and Bob to agree on a sample $x \sim \pi$ with probability at least $1 - \varepsilon$ using communication $\text{poly}(M, \log(1/\varepsilon))$.

Hint: Interpret the public randomness as a sequence of the form (x_i, a_i, b_i) where the goal is to discover the first index i for which $a_i \leq p(x_i)$ and $b_i \leq q(x_i)$.

Problem 5 (Augmented Indexing)**.** Consider the following Augmented Indexing Problem: Alice is given a string $x \in \{0,1\}^n$ and Bob is given an index i *and* the prefix $x_1, \ldots, x_{i-1}$. The goal is for Bob to output x_i given one-way communication from Alice to Bob with probability at least $1 - \delta$.

(a) Prove a lower bound of $(1-h(\delta))n$ on the one-way randomized communication complexity of Augmented Indexing, where $h(\delta) = \delta \log(1/\delta) + (1-\delta)\log(1/(1-\delta))$ is the binary entropy function.

(b) Augmented Indexing is a fairly contrived problem, but it turns out to be useful in applications. In the strict turnstile streaming model, a pair (i_t, v_t) arrives in every round $t = 1, \ldots, m$. The index i_t represents an element of the universe $[n]$ and $v_t \in \{-M, -M+1, \ldots, M\}$ represents an additive update to the count of how many times element i_t has been seen. The strict turnstile model imposes the condition that at every point in time t, the count of every element is non-negative.

In the Distinct Elements problem, we wish to estimate the number of elements which have non-zero count at the end of the stream. Show that any randomized strict turnstile streaming algorithm which, with probability ≥ 0.99, estimates this count to within a factor of 2 requires space $\Omega(\log n)$ even for some $m = O(n)$ and $M = O(1)$.

(c) Optional Challenge Problem: Show that any algorithm which estimates this count to within a factor of $(1+\varepsilon)$ requires space $\Omega(\log(\varepsilon^2 n)/\varepsilon^2)$ as long as $\log(\varepsilon^2 n)/\varepsilon^2 \leq n$.

Hint: Use a reduction from the Gap-Hamming problem to Indexing. In the Gap-Hamming problem, Alice and Bob receive $x, y \in \{0,1\}^n$ with the promise that $|x-y| \geq n/2 + \sqrt{n}$ or $|x-y| \leq n/2 - \sqrt{n}$ and their goal is to decide which is the case.