Homework 7 – Due Thursday, October 20, 2022

Page limit You can submit at most 2 pages per problem, even if the problem has multiple parts. If you submit a longer solution for some problem, only the first sheet of paper will be graded.

Reminder Collaboration is permitted, but you must write the solutions by yourself without assistance, and be ready to explain them orally to the instructor if asked. You must also identify your collaborators and whether you gave help, received help, or worked something out together. Getting solutions from outside sources such as the Web or students not enrolled in the class is strictly forbidden.

Exercises Please practice on exercises in Chapter 4 of Mitzenmacher-Upfal.

Problems

1. (Random vectors) Consider $d$-dimensional vectors, where $d$ is a sufficiently large integer. Recall that two vectors $a$ and $b$ are orthogonal if their dot product $a \cdot b = \sum_{i \in [d]} a_i b_i$ is zero. For a real number $\epsilon > 0$, two vectors are $\epsilon$-close to being orthogonal if their dot product is in the interval $[-\epsilon, \epsilon]$.

   (a) You pick two $d$-dimensional unit vectors independently at random by setting each coordinate uniformly and independently to $1/\sqrt{d}$ or $-1/\sqrt{d}$. Give the best upper bound you can on the probability that the two vectors are not $1/10$-close to orthogonal.

   (b) In $d$ dimensions, at most $d$ vectors can be pairwise orthogonal. However, exponentially many (in $d$) vectors can be close to pairwise orthogonal. Moreover, a random collection of exponentially many vectors is likely to be close to pairwise orthogonal. Find as large $k$ as you can (as a function of $d$) such that $k$ unit vectors chosen independently as specified in part (a) are pairwise $1/10$-close to being orthogonal with (at least) constant probability.

2. (Randomized Rounding and Randomized Response)

   (a) (Randomized Rounding) This problem is similar to the Set Balancing problem discussed in class. We are given an $n \times n$ matrix $A$ with entries in $\{0, 1\}$ and a real-valued vector $p$ with $n$ entries in $[0, 1]$. We would like to “round” the entries of $p$ to obtain an $n$-bit vector $q$ such that $Aq$ is close to $Ap$ in every component. In other words, we would like to ensure that $\|A(p - q)\|_\infty$ is small.

   To obtain $q$, we can use the following approach, called randomized rounding: Each entry $q_i$ of $q$, for $i \in [n]$, is (independently) set to 1 with probability $p_i$ and to 0 with probability $1 - p_i$. Derive a bound on $\|A(p - q)\|_\infty$ that is exceeded with probability at most $2/n$.

   (b) (Randomized Response) You are collecting sensitive data from $n$ people. Each person $i$ has a secret bit $x_i \in \{0, 1\}$. For example, this bit might represent whether each person already had COVID-19 or not. You would like to approximate the sum of the secret bits. (In our example, it would represent the number of people who had COVID-19.) To protect privacy of your respondents, you ask each of them to report $X_i$, which is set to $x_i$ with probability $1/2 + \epsilon$ and to $1 - x_i$ with probability $1/2 - \epsilon$, where $\epsilon \in (0, 1/2)$ is a privacy parameter.
3. (Randomized Routing on the Hypercube) Recall the Randomized Routing Algorithm and its analysis from class. As part of the analysis, we stated a lemma that we did not have time to prove. In this problem, you will prove this lemma.

(a) Consider each route in Phase 1 of the algorithm as a directed path from the source $x$ to the designation $z$. Prove that once two routes separate, they do not rejoin.

(b) Does part (a) imply that, for any two packets $i$ and $j$, there is at most one node such that $i$ and $j$ are waiting in queue at that node at the same time step?

(c) Consider any packet $i$. Let $p_i = (v_1, \ldots, v_k)$ be its path in phase 1. Let $S$ be the set of packets (other than $i$) whose routes pass through at least one edge of $p_i$. Recall that the delay of a packet is the number of time steps it waits in queues (in Phase 1). Show that the delay of packet $i$ is at most $|S|$. 

**Hint:** Use part (a).

**Guidelines:** For each unit of delay that packet $i$ encounters, we would like to “charge” one of the packets in $S$. We define the lag of a packet $i'$ on the edge $(v_j, v_{j+1})$ as $t - j$, where $t$ is the time step when $i'$ traverses the edge $(v_j, v_{j+1})$. We say that a packet $i'$ leaves $p_i$ with lag $\ell$ if the lag of packet $i'$ on the last edge of $p_i$ it traverses is $\ell$.

i. Argue that if the delay of the packet $i$ increases from $\ell$ to $\ell + 1$ (for any integer $\ell$), then there exists a packet $i'$ from $S$ that leaves $p_i$ with lag $\ell$. (You can charge $i'$ for this unit of delay.)

ii. Argue that you are charging each packet in $S$ for at most one unit of delay and conclude that the delay of packet $i$ is at most $|S|$.

As part of our analysis of the Randomized Routing Algorithm that we will continue on Tuesday, we will define a random variable $Y_e$ to be the number of packets whose routes in Phase 1 use edge $e$ of the hypercube. Calculate the expectation of $Y_e$ using the following guidelines.
(d) Let $L_i$ be the number of edges in the route of packet $i$. Argue that $\sum_e Y_e = \sum_i L_i$ for appropriately specified $e$ and $i$. (Make sure you specify what $e$ and $i$ we are summing over.)

(e) Find the expectation of the summation in part (d).

(f) Use part (e) to derive $\mathbb{E}[Y_e]$.

4* (Optional, no collaboration) Compute the variance of the estimate $Z$ for the number of triangles in Problem 2(c). Express your answer using asymptotic notation (specifically, $\Theta(\cdot)$), as a function of $n, \epsilon$, and $C_4$, where $C_4$ is the number of 4-cycles in the input graph.