CS265/CME309, Fall 2019. Instructor: Gregory Valiant

Problem Set 1

Electronic submission via Gradescope (submission code 92EK55) due **11:59pm Tuesday 10/1**. You are strongly encouraged to submit a homework with a partner—that is, submit one homework with both of your names.

[You may discuss these problems with classmates. Feel free to look at wikipedia, course notes, etc. for reference material, but do not try to specifically search online for solutions to the problems. ***Your submission must be the original work of you and your partner, and you must understand everything that is written on your submission.*** *We strongly suggest that you write solutions using LaTex—see the course website for a latex solution template.]*

1. Random Bits:

   (a) (4 points) Given access to coins whose probability of landing heads is $1/5$, describe a scheme for generating a uniformly random bit. In other words, how can one simulate a flip of a coin that lands heads with probability exactly $1/2$?

   (b) (4 points) Given access to coins whose probability of landing heads is $1/2$, describe a scheme for simulating the flip of a coin whose probability of heads is $1/3$. For full credit, your scheme should, *in expectation*, flip at most two coins.

   (c) (4 points) In one or two sentences, explain why the above problems are impossible if you were required to write a protocol that flips the given coins at most some fixed number, k, times.

2. (Linearity of Drunk Pigeons) Suppose there are 100 pigeons, and 100 pigeon-holes, with each pigeon-hole belonging to a pigeon. One crazy night, all the pigeons get rather drunk, and each one selects a pigeon-hole to return to *uniformly at random* .

   (a) (4 points) What is the expected number of pigeons that end up returning to their proper hole? (Hint: use linearity of expectation!)

   (b) (4 points) What is the expected number of pigeons that will end up in a hole with exactly one other pigeon? (Hint: use linearity of expectation!)

   (c) (Bonus 1 point) Suppose we have n pigeons, and n holes. Show that in the limit as $n \to \infty$, the expected fraction of empty holes after the drunken night approaches $1/e$.

3. Suppose we have n students in the class, and exactly $k \leq n$ are infected with the wait-until-Monday-to-start-homework illness. Fortunately, there is a blood test for identifying this, though it is very expensive. Suppose we take a sample of blood from each student. There are two options: 1) test a drop from each of the n samples, 2) randomly partition the n samples into groups of m, for each group, mix together a drop from each of the m samples in the group and test that mixture; if that mixture tests positive, then individually test each of the m students in that group. For the following problems, feel free to assume that n is a multiple of m if that simplifies calculations.

(a) (4 points) If we use the second strategy, as a function of n, k and m, what is the expected number of combined samples (i.e. groups) that will test positive? (Hint: linearity of expectation!)

(b) (4 points) If we use the second strategy, what is the expected total number of tests that we will need to perform to identify all of the k infected students, as a function of $n, k,$ and m? When is this significantly better than using the first strategy (testing all n students separately)? Justify your answer with one or two sentences.

(c) (4 points) If we use the second strategy, what is a near-optimal group size, m, as a function of n and k, and what is the expected number of total tests that will be conducted using this value of m?

(d) (Double Bonus—this part is not graded.) Strategy 2 is a two-step scheme—can you design an optimal multistep scheme? What changes if the test is not 100% accurate, but only accurate with probability $3/4$? Some variants of these questions are still open research problems!

4. Improving the random min-cut algorithm seen in class. Recall that the algorithm proceeds by iteratively choosing a random edge to "contract" until the graph has only 2 vertices, and then outputs the corresponding cut. (In class we showed that the probability that this scheme outputs a minimum cut is at least $\frac{2}{n(n-1)} \approx 2/n^2$.)

(a) (4 points) Consider the following variation: starting with a graph on n vertices, first contract the graph down to k vertices, and now run the randomized algorithm on this smaller graph ℓ (independent) times and output the smallest cut found in any of the ℓ runs. Determine the number of edge contractions and bound the probability of finding a minimum cut, as functions of n, k, ℓ.

(b) (4 points) What are near-optimal values of k, ℓ, (as functions of n) which maximize the probability of finding a minimum cut, while using at most $2n$ total edge contractions? [Feel free to just give the asymptotics—i.e. "$k = \Theta(n^{blah})$, in which case the probability of success is $\Theta(n^{-blah'})$". For comparison, if we were to repeat the algorithm from class twice, and output the better of the two cuts found, we would end up using $2(n-2) \approx 2n$ contractions, and have probability of success $\geq 1 - (1 - \frac{2}{n(n-1)})^2 \approx \frac{4}{n^2}$.]

(c) (Bonus 2 points) Note that one could extend the above improvements by having a multi-stage algorithm with $k_1, k_2, \ldots,$ and corresponding $\ell_1, \ell_2, \ldots$ (as opposed to the two-stage algorithm considered above). Describe an (optimal) scheme that requires at most $\tilde{O}(n^2) = O(n^2 \text{ polylog } n)$ edge contractions to achieve a probability of success $> 1/2$. For this part, feel free to ignore all poly-logarithmic factors of n. [For comparison, the scheme in class would need to be repeated $O(n^2)$ times to achieve success probability $1/2$, and hence the total number of edge contractions would be $\Theta(n^3)$.]

(d) (Bonus 1 point) Prove that it is impossible for any scheme of the above form to achieve success probability greater than $1/2$ using fewer than $\Theta(n^2)$ edge contractions.