Homework 5 – Due Friday, February 21, 2020 at noon

Submit solutions to all problems on separate sheets. They will be graded by different people.

Page limit You can submit **at most** 1 sheet of paper 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 3 of Mitzenmacher-Upfal.

Problems

1. (**Random hats**) Suppose n people come to a theater performance wearing hats. When they leave, they get back a uniformly random hat. That is, each of $n!$ assignments of n hats to n people is equally likely.

   (a) We say that a pair of people (let's call them Alice and Bob) exchanged their hats if Alice got Bob's hat and Bob got Alice's hat. Let X denote the number of pairs of people that exchanged their hats. Calculate the expectation of X.

   Hint: Write X as a sum of indicator random variables.

   (b) Calculate $\mathrm{Var}[X]$.

   Hint: Use the same sum as in part (a); remember that the indicators are not independent.

2. (**MaxCut**) In the problem MaxCut, we are given an undirected graph $G = (V, E)$ and asked to find a cut of maximum size in G. (Recall that a cut in G is a partition of the vertex set V into two parts; the size of the cut is the number of edges with one endpoint in each part of the partition.) In contrast to the seemingly very similar problem MinCut discussed in class (recall Karger's algorithm), MaxCut is a famous NP-hard problem, so we do not expect to find an efficient algorithm that solves it exactly. Here is a simple linear-time randomized algorithm that gives a pretty good approximation:

   - Randomly and independently color each vertex $v \in V$ red or blue with probability $1/2$.
   - Output the cut defined by the red/blue partition of vertices.

   (a) Let random variable X denote the size of the cut output by the algorithm. Compute $E[X]$ as a function of the number of edges in G, and deduce that $E[X] \geq OPT/2$, where OPT is the size of a maximum cut in G.

   Hint: Write X as a sum of indicator random variables.

   (b) Let p denote the probability that the cut output by the algorithm has size at least $0.49\,OPT$. Show that $p \geq 1/51$.

   Hint: Applying Markov's inequality to X will not work here. Try applying Markov's inequality to a different random variable.

(c) Now compute the variance $\mathrm{Var}[X]$.

Hint: Again write X as the sum of indicators, as in part (a).

(d) Let p be the probability defined in part (b). Use Chebyshevs inequality together with part (c) to show that $p = 1 - O(1/|E|)$.

(Note how Chebyshev's inequality gives us a better bound here than Markov's.)

3. (**Generalization of Randomized Median Algorithm**) In this problem, you are asked to generalize the randomized median-finding algorithm from class (Section 3.5 of the MU book), so that it finds an element of rank k (that is, the kth smallest element) in an array of n distinct elements, for any given $k \in [4n^{3/4}, n - 4n^{3/4}]$. You may ignore rounding issues in your algorithm and analysis.

(a) Explain how to modify lines $3, 4, 6$ and 8 of the algorithm in the book.

(b) Analyze the running time of the modified algorithm.

(c) We will follow the same analysis outline as in class (and in the book). Change the definitions of events $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_{3,1}$ and $\mathcal{E}_{3,2}$, so that they apply for general k.

(d) Write $\Pr[\mathcal{E}_1]$ and $\Pr[\mathcal{E}_{3,1}]$ as probability expressions involving the tail of a suitable binomial random variable.

Do *not* repeat the rest of the analysis (which is essentially the same as in the case of finding the median).