CS:5360 Fall 2018 Homework 3

Due: Thu, 10/4

Notes: (a) It is possible that solutions to some of these problems are available to you via text-books on randomized algorithms or on-line lecture notes, etc. If you use any such sources, please acknowledge these in your homework fully *and* present your solutions in your own words. You will benefit most from the homework, if you seriously attempt each problem on your own first, before seeking other sources. (b) As mentioned in the syllabus, it is okay to form groups of two in solving and submitting homework solutions. But, my advice from (b) still applies: you will benefit most from the homework, if you seriously attempt each problem on your own first, before seeking help from your group partner. (c) Discussing these problems with any of your classmates is okay, provided you and your classmates are not being too specific about solutions. In any case, make sure that you take no written material away from these discussions *and* (as in (b)) you present your solutions in your own words. When discussing homework with classmates please be aware of guidelines on "Academic Dishonesty" as mentioned in the course syllabus.

1. This problem is on the median finding algorithm that uses random sampling.

   (a) In the analysis of this algorithm, Chebyshev's inequality was used to derive the $1/n^{1/4}$ upper bound on the probability that the algorithm will fail to yield a median. By adjusting the parameters in this algorithm, namely t and t', can you reduce the error probability to $1/n^c$ for some $c > 1/4$ (e.g., $c = 2/5$)? For your answer, you can either show how to obtain the $1/n^c$ upper bound on the failure probability for some $c > 1/4$ *or* explain why no matter what values of t and t' are used, it is just not possible to improve the $1/n^{1/4}$ error probability upper bound to $1/n^c$ for some $c > 1/4$.

   (b) In this algorithm we used a sample S of size $n^{3/4}$. Suppose we use a much smaller sample of size $O(\log n)$. What other appropriate changes to the algorithm (e.g., in the definitions of d and u) could we make to get an algorithm that runs in $O(n)$ time with failure probability bounded above by $1/n^c$ for a constant $c > 0$? Your answer should either be a restatement of the algorithm (with appropriate changes to parameter values) followed by a modified analysis *or* an explanation of why a sample size of $O(\log n)$ is too small.

2. n balls are thrown uniformly at random into n bins. You can assume that $n \geq 2$. Let Z_i be the indicator variable indicating if bin i is empty. Let $Z = \sum_{i=1}^{n} Z_i$ be the number of empty bins. Show that for $i \neq j$, $Cov(Z_i, Z_j) \leq \frac{c}{n}$ for a constant c. Using this show that $Var(Z) = O(n)$.
   Notes: The inequality $1 + x \leq e^x$ for all real x is useful here. In addition, the following inequality is also useful:
   $$e^t\left(1 - \frac{t^2}{n}\right) \leq \left(1 + \frac{t}{n}\right)^n,$$
   for all real n and t such that $n \geq 1$ and $|t| \leq n$.

3. A *fixed point* of a permutation $\pi : [1, n] \to [1, n]$ is a value for which $\pi(x) = x$. Find the variance in the number of fixed points of a permutation chosen uniformly at random from all permutations.

Hint: Let $X_i = 1$ if $\pi(i) = i$, so that $X = \sum_{i=1}^{n} X_i$ is the number of fixed points. Calculate $E[X_i]$ and $E[X_i \cdot X_j]$ in order to calculate $E[X]$ and $E[X^2]$ and use this to calculate $Var(X)$.

4. In class we showed that the *expected* number of F-light edges in G is at most n/p. Let X be the random variable denoting the number of F-light edges in G. Can you use Chebyshev's Inequality to show that $Pr(X \geq c \cdot n/p) \leq 1/n$ for some constant $c \geq 1$? Like Problem (1) this a bit open-ended and your answer can either be a proof using Chebyshev's Inequality or an explanation about why Chebyshev's Inequality may not be strong enough to show the desired probability bound.

   Note: You can take as given the following fact: if Y is a random variable that has the negative binomial distribution with parameters n and p, then $Var[Y] = n(1-p)/p^2$.
