CS 473: Algorithms, Fall 2016
HW 4 (due Tuesday, September 27th at 8pm)

This homework contains three problems. **Read the instructions for submitting homework on the course webpage**.

Collaboration Policy: For this home work, each student can work in a group with upto three members. Only one solution for each group needs to be submitted. Follow the submission instructions carefully.

1. Consider the following variant of Quick Sort. Given an array A of n numbers (which we assume are distinct for simplicity) the algorithm picks a pivot x uniformly at random from A and computes the rank of x. If the rank of x is between $n/4$ and $3n/4$ (call such a pivot a good pivot) it behaves like the normal Quick Sort in partitioning the array A and recursing on both sides. If the rank of x does not satisfy the desired property (the pivot picked is not good) the algorithm simply repeats the process of picking the pivot until it finds a good one. Note that in priniciple the algorithm may never terminate!

    - Write a formal description of the algorithm.
    - Prove that the expected run time of this algorithm is $O(n \log n)$ on an array of n numbers.
    - **Extra Credit:** Prove that the algorithm terminates in $O(n \log n)$ time with high probability. Does this immediately imply that the expected run time is $O(n \log n)$?

2. Let σ be a uniformly random permutation of $\{1, \ldots, n\}$. That is $\sigma(1), \sigma(2), \ldots, \sigma(n)$ is a permutation and it is chosen uniformly from one of the $n!$ permutations. We say that position i is a peak in σ if $\sigma(i)$ is the maximum number amongst $\sigma(1), \sigma(2), \ldots, \sigma(i)$. For instance if σ is the permutation $3, 4, 1, 2, 5$ then positions $1, 2, 5$ are peaks and positions 3 and 4 are not. Note that position 1 is always a peak. Let σ be a uniform random permutation of $\{1, 2, \ldots, n\}$.

    - What is the probability that position i is a peak in σ?
    - What is the expected number of peaks in σ?

3. Consider a balls and bins experiment with $2n$ balls but only two bins. Each ball is thrown independently into a bin chosen uniformly at random. Let X_1 be the random variable for the number of balls in bin 1 and X_2 for bin 2. It is easy to see that $\mathbb{E}[X_1] = \mathbb{E}[X_2] = n$. We would like to have a handle on the difference $X_1 - X_2$. Our goal is to prove that for any fixed $\epsilon > 0$ there is a fixed constant $c > 0$ such that $\Pr[X_1 - X_2 > c\sqrt{n}] < \epsilon$. By symmetry we can then argue that $\Pr[|X_1 - X_2| > c\sqrt{n}] < 2\epsilon$. *Hint:* The parts below will help you answer this question using two bounds and compare them.

    - Compute the variance of X_1. Then use Chebyshev bound to show that $\Pr[|X_1 - n| > c\sqrt{n}] \leq \epsilon$ for suitable choice of c for a given ϵ. What is the dependence of c on ϵ?
    - Use Chernoff bound to show that $\Pr[|X_1 - n| > c\sqrt{n}] \leq \epsilon$. You need to use the bound separately for computing $\Pr[X_1 > n + c\sqrt{n}]$ and for $\Pr[X_1 < n - c\sqrt{n}]$. What is the dependence of c on ϵ?

- Using the preceding show that $\Pr[X_1 - X_2 > c\sqrt{n}] < \epsilon$.
- **Extra Credit:** A one-dimensional random walk on the integer line starts at position 0 on the number line. In each step we move from the current position one unit step to the left or one unit step to the right with equal probability (independent of the previous choices). Let Z_n be the position of the walk after n steps (it is an integer in the range $[-n, n]$). Using a simple connection to the problem of throwing balls into two bins show that for any fixed ϵ, there is a c that depends only on ϵ such that $\Pr[|Z_n| > c\sqrt{n}] < \epsilon$. Also derive that $\mathbb{E}[|Z_n|] = O(\sqrt{n})$.

The remaining problems are for self study. Do *NOT* submit for grading.

- Markov's inequality states that if X is a non-negative random variable then for $t \geq 1$, $\Pr[X \geq t\,\mathbb{E}[X]] \leq 1/t$. We claim the following corresponding inequality for the lower tail: for non-negative X and $t \geq 1$, $\Pr[X < \mathbb{E}[X]/t] \leq 1/t$. Prove the inequality, or disprove via a counter example.

- See Jeff's home work problems on randomized algorithms from last semester. `https://courses.engr.illinois.edu/cs473/sp2016/hw/hw3.pdf` (Ignore the last problem in this) and `https://courses.engr.illinois.edu/cs473/sp2016/hw/hw4.pdf`

- Jeff's notes have several nice problems including some in the home work.

- Suppose you are presented with a very large set S of real numbers, and you would like to approximate the median of these numbers by sampling. You may assume all the members in S are distinct. Let $n = |S|$. We say that a number x is an ϵ-approximate median of S if at least $(1/2 - \epsilon)n$ numbers in S are less than x, and at least $(1/2 - \epsilon)n$ numbers in S are greater than x. Consider an algorithm that works as follows. You select a subset $S' \subseteq S$ uniformly at random, compute the median of S', and return this as an approximate median of S. Show that there is an absolute constant c, independent of n, so that if you apply this algorithm with a sample S' of size c, then with probability at least .99, the number returned will be a (.05)-approximate median of S. (You may consider either the version of the algorithm that constructs S' by sampling with replacement, so that an element of S can be selected multiple times, or one without replacement.)

- Consider the following experiment with balls and bins. The experiment proceeds in rounds. In the beginning there are n balls and n bins. At the start of a round each remaining ball is thrown into a bin independently and uniformly into one of the n bins. After the balls are thrown any ball that is *alone* (that is, it is the only ball in its bin) is removed from the experiment. This finishes a round. The remaining balls participate in the next round. The experiment stops when there are no balls remaining after a round.
  - What is the probability that a specific ball i remains after the first round?
  - Prove that the experiment finishes in $c \log n$ rounds with probability at least $1 - 1/n$ for some appropriate choice of constant c. *Hint:* This is similar to the high probability analysis of Quick Sort.
  - **Not to submit:** Also show that the expected number of rounds for the experiment to finish is $O(\log n)$.
  - **Extra Credit:** Prove that the expected number of rounds for the experiment to finish is $O(\log \log n)$.