CS 473: Algorithms, Spring 2021
HW 4 (due Wednesday, March 3rd 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 up to three members. Only one solution for each group needs to be submitted. Follow the submission instructions carefully.

1. 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 $\Pr[|X_1 - n| > c\sqrt{n}] \le \epsilon$ for suitable choice of c for a given ϵ. What is the dependence of c on ϵ?
    - Use Chernoff bound to show that $\Pr[|X_1 - n| > c\sqrt{n}] \le \epsilon$. 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 $\Pr[X_1 - X_2 > c\sqrt{n}] < \epsilon$.
    - **Not to submit:** 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 $\mathbb{E}[|Z_n|] = O(\sqrt{n})$.

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 uniform rooted tree of height h (every leaf is at distance h from the root). The root, as well as any internal node, has 3 children. Each leaf has a boolean value associated with it. Each internal node returns the value returned by the majority of its children. The

evaluation problem consists of determining the value of the root; at each step, an algorithm can choose one leaf whose value it wishes to read.

(a) Show that for any deterministic algorithm, there is an instance (a set of boolean values for the leaves) that forces it to read all $n = 3^h$ leaves. (*Hint*: Consider an adversary argument, where you provide the algorithm with the minimal amount of information as it request bits from you. In particular, one can devise such an adversary algorithm.).

(b) Consider the recursive randomized algorithm that evaluates two subtrees of the root chosen at random. If the values returned disagree, it proceeds to evaluate the third sub-tree. If they agree, it returns the value they agree on.

Write an explicit exact formula for the expected number of leaves being read, for a tree of height $h = 1$, and height $h = 2$.

(c) Using (b), prove that the expected number of leaves read by the algorithm on any instance is at most $n^{0.9}$.