CS 13: Mathematical Foundations of Computer Science Fall 2023

Written Homework 03 (due December 4th, 2023)

Directions: *Write up carefully argued solutions to the following problems. Your solution should be clear enough that it should explain to someone who does not already understand the answer why it works. You may use results from lecture and previous homeworks without proof. Your solutions* must *be written in LaTeX using our homework template.* ***No solution to a single part may be more than one page.***

0. Hashing (20 points)

Suppose n names are chosen (with replacement) independently at random from a universe of $N \gg n$ possible names, then hashed into a hash table with b buckets. Suppose that N is a multiple of b, and that the hash function maps exactly N/b of the possible names into each of the b buckets.

(a) [10 Points] What is the probability that this results in *no collisions*, i.e., no two picks hash to the *same* bucket?

(b) [5 Points] For $b = 10000$, write a short program to *simulate* this process for *1000* trials when $n = 115, \ldots, 125$. Make sure to include your code, your results for each n, and the largest value of n such that this probability is greater than 0.5.

(c) [5 Points] Check that your theoretical answers match your empirical ones. Include a graph of your theoretical results.

1. Tournament Champions (30 points)

An *n-player tournament* consists of some set of $n \geq 2$ players and a specification, for every pair of players, of who won that game. That is, player p beats player q iff q does not beat p, for all players $p \neq q$.

(a) [5 Points] Rephrase the problem as a graph problem. What are the vertices? What are the edges? Is the graph directed? Is the graph weighted?

(b) [15 Points] Prove that, in every tournament, either there is a "champion" player that beats every other player, or there is a 3-cycle of players.

(c) [10 Points] A consistent ranking is a sequence $p_1, p_2, \ldots, p_n$ of all n players in the tournament such that p_i beats p_j iff $i < j$, for $i \geq 1$, $j \leq n$. Conclude that a tournament has no consistent ranking iff some subset of three of its players has no consistent ranking.

2. Six Color Theorem (25 points)

<u>New Definitions</u>

A graph is *planar* if there exists a drawing of it on a plane without any pair of edges crossing each other. A *face* of a planar graph is a region of space enclosed in edges of the graph. The region outside the graph is also considered a face.

<u>An Extra Theorem ("Euler's Formula")</u>

Let G be an arbitrary connected planar graph with v vertices, e edges, and f faces. Then $v - e + f = 2$.

(a) [25 Points] Show that every planar graph G with v vertices, e edges, and f faces can be colored with at most six colors by proving the following lemmas.

(i) Show that $2e \geq 3f$ in any *connected* planar graph with $v \geq 3$.

(ii) Use your bound to show that any planar graph must contain a vertex of degree at most five.

(iii) Prove the claim by graph induction on the number of vertices.

(b) [10000000 Points] **Extra Credit.** Prove the analogous theorem using only four colors.

3. Random Binary Search Trees (25 points)

Suppose we plan to insert N distinct values into an empty binary search tree. As we know, it's possible that the resulting tree is linear (aka a linked list). We'd like to avoid this if possible; so, before inserting our N items, we shuffle them into a random insertion order.

For the remainder of this problem, we make the additional assumption that the actual values are the numbers from 1 to N, because the values don't actually matter for analysis–only the order of them.

Finally, then, we say $x_1, x_2, \ldots, x_N$ is some random permutation of $[N]$ that we insert into a BST, in that order. Also, let $x \in [N]$ be arbitrary.

Let X be the length of the search path for x in the constructed tree. In this problem, we show that $\mathbb{E}[X] \in \mathcal{O}(\lg N)$. That is, the runtime of find(x) is, on average, $\mathcal{O}(\lg N)$.

(a) [5 Points]

Insight 1
Note that i, where $i < x$, is on the search path for x iff i is the first element from the set $\{i, i{+}1, ..., x{-}1, x\}$ added to the tree.

Explain why **Insight 1** must be true.

(b) [6 Points] Let X_i be an indicator random variable for the event that i is on the search path for x. Given part (a), find $\mathbb{E}[X_i]$. You can assume the analogous insight to **Insight 1** for $i > x$ - that is, i appears on the search path for x iff i is the first element added of $\{x, x+1, ..., i\}$.

(c) [7 Points] Recall that the harmonic numbers, H_n are defined as $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$. Find $\mathbb{E}[X]$ and prove your result.

(d) [7 Points] Finally, we need to show that $\ln n < H_n < (\ln n) + 1$. To do this, relate the integral of $\frac{1}{x}$ to the summation of $\frac{1}{x}$ and use those relations to upper and lower bound the summation. (For this part of this problem *only*, you may use your background knowledge of calculus.) Finally, conclude that, for some $a \in \mathbb{Q}_+$ and $b \in \mathbb{Z}$, $\mathbb{E}[X] \leq a \lg N + b$.
Hint: Note that one part of the problem uses $\lg = \log_2$ and the other uses $\ln$; you'll need to reconcile this for full credit.