6.856 — Randomized Algorithms

Spring Term, 2021

Feb. 24, 2021 — Problem Set 2, Due 3/3

Problem 1. 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.

(b) Show that there is a nondeterministic algorithm can determine the value of the tree by reading at most $n^{\log_3 2}$ leaves. In other words, prove that one can present a set of this many leaves from which the tree value can be determined.

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

Problem 2. MR 2.6. Use Yao's minimax principle to prove a lower bound on the expected running time of any Las Vegas algorithm for sorting n numbers that uses only comparisons. You might want to review deterministic sorting lower bounds from, e.g., the CLRS book.

Problem 3—This problem should be done without collaboration. Consider a sequence of n unbiased coin flips. Consider the length of the longest contiguous sequence of heads.

(a) Show that you are unlikely to see a sequence of length $c + \log_2 n$ for $c > 1$ (give a decreasing bound as a function of c).

(b) Show that with high probability you will see a sequence of length $\log_2 n - O(\log_2 \log_2 n)$. **Note:** this observation can be used to detect cheating. When told to fake a random sequence of coin tosses, most humans will avoid creating runs of this length under the mistaken assumption that they don't look random.

For both parts, use specifics of the problem, not the general deviation bounds.

Problem 4. When we studied the median finding algorithm in class, I showed that the median is probably between the two chosen "boundaries." But I merely asserted that not many items are found between these boundaries. Using the Chebyshev bound, prove that it is unlikely (probability $O(n^{-1/4})$) for the chosen boundaries to have many elements between them, for a suitable choice of "many."

Problem 5. The weak law of large numbers says that if $X_1, X_2, \ldots$ are independent, identically distributed random variables with mean μ and finite standard deviation, then for any constant $\epsilon > 0$:

$$\lim_{n\to\infty} \Pr\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) = 0.$$

In other words, the average of the random variables almost surely converges to the expectation. Use Chebyshev's inequality to prove the weak law.

Problem 6—Optional. Derive a distribution on input game trees that yields a tight lower bound (using Yao's principle) on the runtime of any randomized game tree evaluation algorithm.

Problem 7—Optional. MR 1.15. Prove that $NP \subseteq BPP$ implies $NP = RP$.