

CPS130: Homework 4
Due: Thursday, March 10, 2005

Do not be intimidated by this homework, there are only three problems here (plus one extra credit) - it looks longer than it is.

1. One of the ways we can improve randomized Quicksort is to use a median of three partitioning scheme. With this method, we pull three randomly selected items out of the list, and we use the median of the three values as our pivot. Note that this removes the two worst cases for Quicksort, where the minimum or maximum is chosen to be the pivot. Assuming that there are at least three elements in the list and they all have distinct values.
 - a) If we let x be the chosen pivot and $A'[1..n]$ be the sorted list, what is the probability we choose $x = A'[i]$? Denote this value p_i .
 - b) How much have we increased the likelihood of choosing the median as the pivot element, as compared to the normal implementation? What is the limit as $n \rightarrow \infty$?
 - c) If we define a "good" split to be a choice of pivot in the range $n/3 \leq i \leq 2n/3$, how much have we increased the likelihood of getting such a split? (*Consider approximating the sum with an integral*)
 - d) Argue that this method does not change the $\Omega(n \log n)$ lower bound on Quicksort, only the constant factor involved.
2. Suppose that at each run of Quicksort, the splits are in the proportion $1 - \alpha$ to α , for some $0 < \alpha < 1/2$ is a constant. Show that the minimum depth of a leaf in the recursion tree is approximately $-\log n / \log \alpha$ and the maximum depth is approximately $-\log n / \log(1 - \alpha)$.
3. Suppose we have a hash table with n slots where collisions are resolved via chaining, and suppose we insert exactly n items into the hash table. Each key is equally likely to be hashed to any given cell of the table. Let M be the maximum number of keys in any slot after all the keys have been inserted. We wish to find an upper bound on the expected value of M .

- a) Argue that the probability Q_k that exactly k keys hash to a particular slot is given by:

$$Q_k = \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k} \binom{n}{k}$$

- b) Let P_k be the probability that M equals k , that is, the probability that the slot containing the most keys contains k keys. Show $P_k \leq nQ_k$.
- c) Use Stirling's approximation to show that $Q_k < e^k / k^k$.
- d) Show that there exists a constant $c > 1$ such that $Q_{k_0} < 1/n^3$ for $k_0 = c \log n / \log \log n$. Conclude that $P_k < 1/n^2$ for some $k \geq k_0$.
- e) Argue that this is true:

$$E[M] \leq n \Pr \left\{ M > \frac{c \log n}{\log \log n} \right\} + \Pr \left\{ M \leq \frac{c \log n}{\log \log n} \right\} \frac{c \log n}{\log \log n}$$

And conclude that $E[M] \in O(\log n / \log \log n)$.

Extra Credit: Suppose we several decks of normal playing cards, all shuffled together. You are going to draw cards, one at a time, until you have all 13 of the Spades. For the sake of convenience, let's assume that we have so many decks of cards that at any given time, the probability of drawing any of the 52 possible cards is identical, and we can never draw so many of a certain card that it changes this probability. In this extremely hypothetical situation, how many cards do you expect to draw so that you get all 13 Spades? What if we extend this so that you want all of the Spades and all of the Diamonds? A complete deck?