Design and Analysis of Algorithms I

QuickSort

Analysis II: The Key Insight

Average Running Time of QuickSort

Quicksort Theorem: for every input array of length n, the average running time of QuickSort (with random pivots) is O(n log n).

Note: holds for every input. [no assumptions on the data]

- recall our guiding principles!
- "average" is over random choices made by the algorithm (i.e., pivot choices)

The Story So Far

$C(\sigma)$ = # of comparisons QuickSort makes with pivots σ

$X_{ij}(\sigma)$ = # of times z_i & z_j get compared

i^{th}, j^{th} smallest entries in array

Recall: $E[C] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Pr[X_{ij}=1]$ = Pr[z_i, z_j get compared]

Key Claim: $\forall i < j$, Pr[z_i, z_j get compared] $= \frac{2}{(j-i+1)}$

Proof of Key Claim

$Pr[z_i, z_j \text{ get compared}] = \frac{2}{(j-i+1)}$

Fix z_i, z_j with $i < j$.

Consider the set $z_i, z_{i+1}, \ldots, z_{j-1}, z_j$.

Inductively: as long as none of these are chosen as a pivot, all are passed to the same recursive call.

Consider the first among $z_i, z_{i+1}, \ldots, z_{j-1}, z_j$ that gets chosen as a pivot.

1. if z_i or z_j gets chosen first, then z_i and z_j get compared
2. if one of $z_{i+1}, \ldots, z_{j-1}$ gets chosen first, then z_i & z_j are never compared [split into different recursive calls]

key insight

Tim Roughgarden

Proof of Key Claim (con'd)

(1) z_i or z_j chosen first $\Rightarrow$ they get compared

(2) One of $z_{i+1}, \ldots, z_{j-1}$ chosen first $\Rightarrow$ z_i, z_j never compared

Note: since pivots always chosen uniformly at random, each of $z_i, z_{i+1}, \ldots, z_{j-1}, z_j$ is equally likely to be the first.

$$\Rightarrow \Pr[z_i, z_j \text{ get compared}] = \frac{2}{(j-i+1)}$$

choices that lead to case (1)

total # of choices

QED!

So: $E[C] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{(j-i+1)}$ (*)

[still need to show this $O(n \log n)$]