Design and Analysis of Algorithms I

QuickSort

Analysis I: A Decomposition Principle

Necessary Background

Assumption: you know and remember (finite) sample spaces, random variables, expectation, linearity of expectation. For review:

- Probability Review I (video)
- Lehman-Leighton notes (free PDF)
- Wikibook on Discrete Probability

Average Running Time of QuickSort

<u>QuickSort Theorem</u> : for every input array of length n, the average running time of QuickSort (with random pivots) is $O(n\log(n))$.

<u>Note</u> : holds for every input. [no assumptions on the data]

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

Preliminaries

Fix input array A of length n

Sample Space Ω = all possible outcomes of random choices in QuickSort (i.e., pivot sequences)

Key Random Variable : for $\sigma \in \Omega$

$C(\sigma)$ = # of comparisons between two input elements made by QuickSort (given random choices σ)

Lemma: running time of QuickSort dominated by comparisons.

There exist constant c s.t. for all $\sigma \in \Omega$, $RT(\sigma) \leq c \cdot C(\sigma)$ (see notes)

Remaining goal : E[C] = O(nlog(n))

Tim Roughgarden

Building Blocks

<u>Note</u> can't apply Master Method [random, unbalanced subproblems]

$\boxed{6 \mid 8 \mid 10 \mid 2}$

$z_2 \quad z_3 \quad z_4 \quad z_1$

[A = final input array]

<u>Notation</u> : z_i = i^{th} smallest element of A

For $\sigma \in \Omega$, indices i< j

$X_{ij}(\sigma)$ = # of times z_i,z_j get compared in QuickSort with pivot sequence σ

Ter
ver

Fix two elements of the input array. How many times can these two elements get compared with each other during the execution of QuickSort?

- 1
- 0 or 1
- 0, 1, or 2
- Any integer between 0 and $n-1$

Reason : two elements compared only when one is the pivot, which is excluded from future recursive calls.

Thus : each X_{ij} is an "indicator" (i.e., 0-1) random variable

A Decomposition Approach

So : $C(\sigma)$ = # of comparisons between input elements

$X_{ij}(\sigma)$ = # of comparisons between z_i and z_j

Thus : $\forall \sigma, C(\sigma) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}(\sigma)$

complicated

simple

By Linearity of Expectation : $E[C] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[X_{ij}]$

Since $E[X_{ij}] = 0 \cdot Pr[X_{ij} = 0] + 1 \cdot Pr[X_{ij} = 1] = Pr[X_{ij} = 1]$

Thus : $E[C] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Pr[z_i, z_j \text{ get compared}] \quad (*)$

Next video

Tim Roughgarden

A General Decomposition Principle

1. Identify random variable Y that you really care about

2. Express Y as sum of indicator random variables :

$$Y = \sum_{l=1}^{m} X_e$$

3. Apply Linearity of expectation :

"just" need to understand these!

$$E[Y] = \sum_{l=1}^{m} Pr[X_e = 1]$$