Design and Analysis of Algorithms I

Linear-Time Selection

Randomized Selection (Analysis)

Running Time of RSelect

RSelect Theorem: for every input array of length n, the average running time of RSelect is $O(n)$.

- holds for every input [no assumptions on data]
- "average" is over random pivot choices made by the algorithm

Randomized Selection

RSelect (array A, length n, order statistic i)

(0) if $n=1$ return $A[1]$

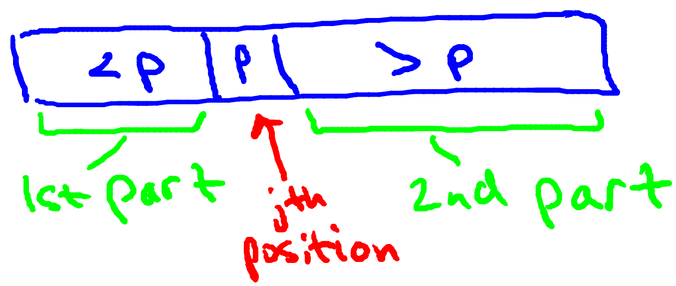
(1) choose pivot p from A uniformly at random

(2) partition A around p
let j = new index of p

(3) if $j=i$ return p

(4) if $j>i$ return RSelect (1st part of A, $j-1$, i)

(5) [if $j<i$] return RSelect (2nd part of A, $n-j$, $i-j$)



Proof I: Tracking Progress via Phases

Note: RSelect uses $\leq cn$ operations outside of the recursive call [for some constant $c > 0$] [from partitioning]

Notation: RSelect is in $\boxed{\text{phase } j}$ if current array size between $\left(\frac{3}{4}\right)^{j+1} n$ and $\left(\frac{3}{4}\right)^{j} n$ (*)

- X_j = number of recursive calls during phase j

Note: running time of RSelect

# of phase-j subproblems

$$\leq \sum_{\text{phases } j} X_j \cdot c \cdot \left(\frac{3}{4}\right)^j n$$

$\leq$ array size during phase j

$\leftarrow$ work per phase-j subproblem

Proof II: Reduction to Coin Flipping

X_j = # of recursive calls during phase j — size between $(3/4)^{j+1} n$ and $(3/4)^j n$

Note: if RSelect chooses a pivot giving a 25-75 split (or better) then current phase ends! | < p | p | > p |
(new subarray length at most 75% of old length)

Recall: probability of 25-75 split or better is 50%.

So: $E[X_j] \leq$ expected number of times you need to flip a fair coin to get one "heads".

(heads ≈ good pivot, tails ≈ bad pivot)

Tim Roughgarden

Proof III: Coin Flipping Analysis

Let N = number of coin flips until you get heads.

(a "geometric random variable")

Note: $$E[N] = \underset{\substack{\uparrow \\ \text{1st coin flip}}}{1} + \underset{\substack{\uparrow \\ \text{probability of tails}}}{\tfrac{1}{2}} \cdot \underbrace{E[N]}_{\text{\# of further coin flips needed in this case}}$$

Solution: $E[N] = 2$. (Recall $E[X_j] \le E[N]$)

Putting It All Together

expected running time of RSelect $\leq E\left[cn \sum_{\text{phase } j} \left(\frac{3}{4}\right)^j X_j\right]$ (*)

$= cn \sum_{\text{phase } j} \left(\frac{3}{4}\right)^j E[X_j]$ (LIN Exp)

$E[X_j] \leq E[\text{\# of coin flips } N] = 2$

$\leq 2cn \sum_{\text{phase } j} \left(\frac{3}{4}\right)^j$

geometric sum, $\leq \frac{1}{1 - 3/4} = 4$

$\leq 8cn = O(n)$

QED!

Tim Roughgarden