Design and Analysis of Algorithms I

Linear-Time Selection

Randomized Selection (Analysis)

Running Time of RSelect

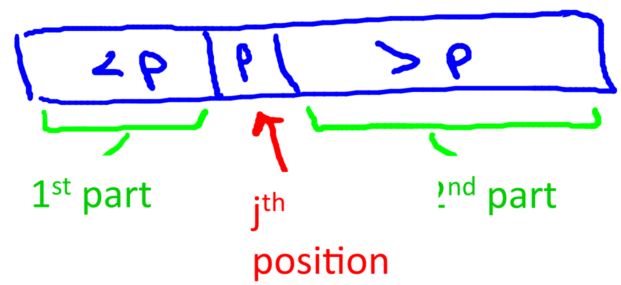
Rselect Theorem : for every input array of length n, the average running time of Rselect is O(n)

-- 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 <= cn operations outside of recursive call [for some constant c > 0] [from partitioning]

Notation : Rselect is in phase j if current array size between $(\frac{3}{4})^{j+1} \cdot n$ and $(\frac{3}{4})^{j} \cdot n$

-X_j = number of recursive calls during phase j

# of phase j subproblems

<= array size during phase j

Note : running time of RSelect $\leq \sum_{phases\ j} X_j \cdot c \cdot \boxed{(\frac{3}{4})^j \cdot n}$

Work per phase j subproblem

Tim Roughgarden

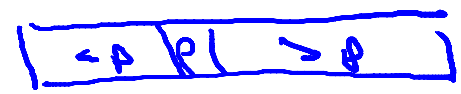
both RHS and LHS are r.v. for they all depend on the pivot chosen

Proof II: Reduction to Coin Flipping

X_j = # of recursive calls during phase j

Size between $(\frac{3}{4})^{j+1} \cdot n$ and $(\frac{3}{4})^{j} \cdot n$

Note : if Rselect chooses a pivot giving a 25 – 75 split (or better) then current phase ends !
(new subarray length at most 75 % of old length)



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

So : $E[X_j]$ <= 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")

memoryless

Note : $E[N] = 1 + (1/2)*E[N]$

1st coin flip

Probability of tails

# of further coin flips needed in this case

Solution : $E[N] = 2$ (Recall $E[X_j] <= E[N]$)

Putting It All Together

Expected running time of RSelect

$$\le E\left[cn \sum_{phase\ j} \left(\frac{3}{4}\right)^j X_j\right] \qquad (*)$$

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

= E[# of coin flips N] = 2

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

geometric sum,
<= 1/(1-3/4) = 4

$$\le 8cn = O(n)$$

Q.E.D.