Find the Kth largest number

Special topics in Advanced Algorithms

-Slides prepared by Raghu Srinivasan

Problem

- Input: Unsorted set of numbers and an integer k
- Output: kth largest number from the given set of numbers

- Deterministic Solution – Using median of medians
  - Runtime: Linear i.e., O(n)

Randomization

- Let S be the set of numbers that are provided, of which we need to find the kth largest

  RandSelect(S, k)

  Step 1: Select a partitioning element s belonging to S at random

  Step 2: Partition S into {L} {s} {R}, where L <= s <= R

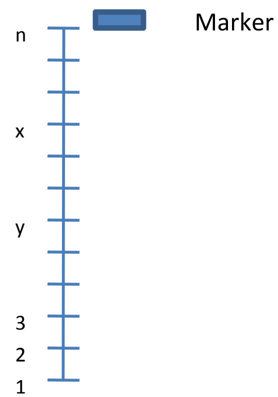
Randomization (Contd…)

Step 3: If |L| = k-1, then return s
  else if [L] >= k then return RandSelect(L,k)
  else return RandSelect(R, k-|L|-1)

What is the expected number of recursive calls?

Algorithms behavior simplified

- Consider this



Marker initially set to n

Pick a number at random such that it is less than n. E.g., x and move marker to that number

Pick a number at random such that it is less than x. E.g., y and move marker to that number

Continue this way till marker reaches 1

- Similar to the algorithm, here we don’t know how many steps it takes for the marker to reach ‘1’; could be anywhere between 1 and n

Intuition

- The randomly selected element, for most of the runs will divide the given set of numbers equally.

- From the above intuition, the estimated number of steps or recursive calls for most of the runs would be, $O(\log n)$

Proof by induction

- We will prove that the number of steps in the recursive process, $T(N) \le f(N)$

- Using summation we have: -

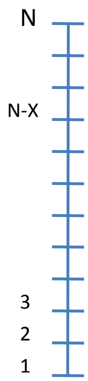
  $f(N) = \sum_{i=1}^{n} (1/(i/2))$,

    pleas read f(N) equals sigma, i ranging from 1 to n of, 1 over i by 2

- Note: Here, the above solution is already known and we are just verifying its correctness.

Proof by induction

- With the diagram on LHS as reference, please consider, that the marker was moved from N to N-X in the first move, therefore



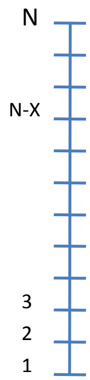
$$T(N) = 1 + E[\, f(N-X)]$$

$$T(N) = 1 + E\left[\, f(N) - \sum_{i=N-X}^{N} (1/(i/2))\right]$$

$$T(N) = 1 + E[f(N)] - E\left[\sum_{i=N-X}^{N} (1/(i/2))\right]$$

$$T(N) = 1 + f(N) - E\left[\sum_{i=N-X}^{N} (1/(i/2))\right]$$

Proof by induction



Since N is always >= I,

T(N) <= 1 + f(N) – E [sum_{i=N-X}^{N} (1/(N/2))]

T(N) <= 1 + f(N) – (E[X])/(N/2)

T(N) <= 1 + f(N) = (N/2)/(N/2)

Therefore T(N) <= f(N)
Hence the number of recursive calls is atmost O(log n)

Exercise: Prove that the total word done is linear in time

References

- Class notes on 01/30/2008
- Randomized Algorithms, by Rajeev Motwani and Prabhakar Raghavan