CSE 5311 Notes 6: Medians/Selection

(Last updated 10/4/15 12:48 PM)

MINIMUM AND MAXIMUM IN NO MORE THAN $3\left\lfloor \frac{n}{2} \right\rfloor$ COMPARISONS (CLRS 9.1)

Ordinary method takes $2n - 3$ comparisons:

1. Find minimum with $n - 1$ comparisons.
2. Remove minimum from further consideration.
3. Find maximum with $n - 2$ comparisons.

Faster Method:

1. Pair up numbers. Put smallest in minimum table, largest in maximum table.

   If n is odd, then leftover number goes in both tables.

   ($\left\lfloor \frac{n}{2} \right\rfloor$ comparisons)

2. Find ordinary minimum for minimum table. ($\left\lceil \frac{n}{2} \right\rceil - 1$ comparisons)
3. Find ordinary maximum for maximum table. ($\left\lceil \frac{n}{2} \right\rceil - 1$ comparisons)

FINDING ELEMENT OF ARBITRARY RANK IN $\Theta(n)$ <u>EXPECTED</u> TIME (CLRS 9.2)

n elements, element of rank k (kth smallest) is desired.

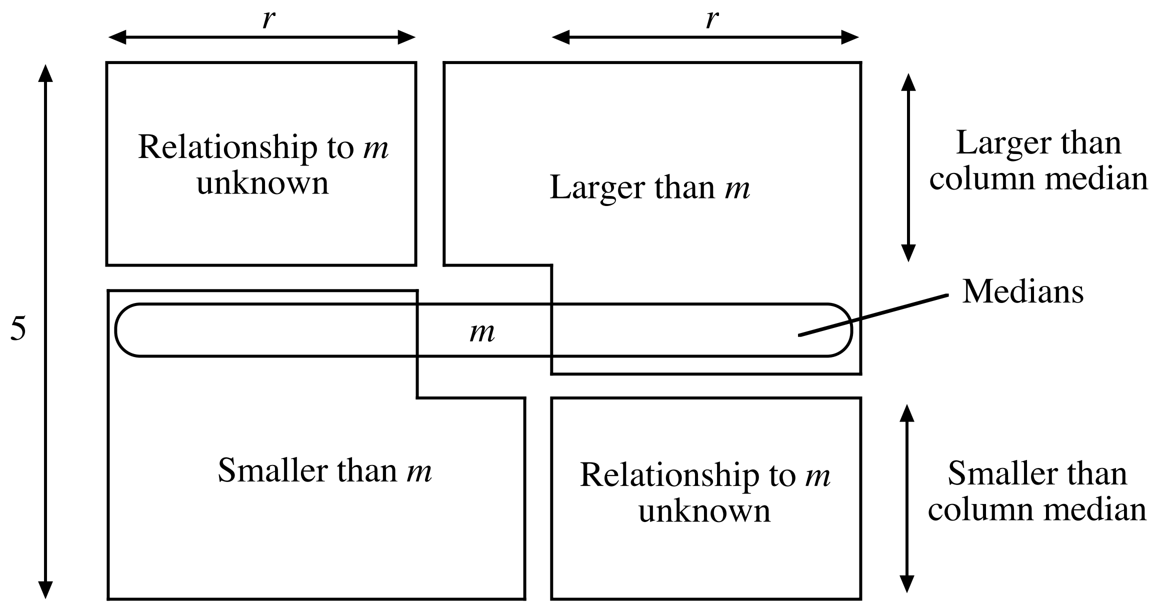
Use PARTITION from QUICKSORT (random pivot or assume random data) to determine the pivot's rank.

Throw away subarray without desired element and continue.

$\Theta(n^2)$ worst-case time.

(CSE 2320 Notes 8.B has analysis. Expected number of comparisons is bounded by $4n$)

FINDING ELEMENT OF ARBITRARY RANK IN $\Theta(n)$ WORST-CASE TIME



n elements, element of rank k (kth smallest) is desired.

1. Group in columns of five ($2r + 1$ columns, $n = 10r + 5$).

2. Find each column's median, along with two elements smaller and two elements larger.

3. Recursively find median-of-medians (call it m).

4. Place groups by their medians' relationship (e.g. smaller/larger) to m.

5. Unknown sections (NW and SE) are split by comparing each element with m.

6. Two subsets from 5. are distributed to "Smaller than m" (SW) and "Larger than m" (NE).

   $3r + 3 \leq \text{rank}(m) \leq n - 3r - 2$ (extremes are based on $4r$ elements in 5. going one way or the other)

7. If rank(m) = k then done. Otherwise, continue with appropriate subset from 6.

   if rank(m) > k
   - $n' = \text{rank}(m) - 1$
   - $k' = k$

   else
   - $n' = n - \text{rank}(m)$
   - $k' = k - \text{rank}(m)$

Analysis (different from CLRS 9.3):

$W(n)$ = maximum number of key comparisons to find element of arbitrary rank among n values

Assume $n = 5(2r + 1)$ for some r. ($r = .1n - .5 \leq .1n$)

The processing required for 1. and 2. takes 6 comparisons for each group of 5: $6(n/5) = 1.2n$

(Trivial to do with 10 comparisons using insertion sort. This raises c below to at least 24. Knuth, *TAOCP*, Vol. 3, 5.3.1 goes into this deeply)

Steps 3. and 4. take $W(n/5) = W(.2n)$.

Steps 5. and 6. take $2 \bullet 2 \bullet r \leq .4n$ comparisons

"Recursive" call for step 7. takes no more than $W(7r + 2) = W(.7n - 1.5) \leq W(.7n)$

$W(n) \leq 1.2n + W(.2n) + .4n + W(.7n) = 1.6n + W(.2n) + W(.7n)$

Use substitution method to show $W(n) \leq cn$.

$W(n) \leq 6$ for $n \leq 5$

$W(n) \leq 1.6n + W(.2n) + W(.7n)$ for $n > 5$

Assume that some c gives linear upper bound for $k < n$, i.e. $W(k) \leq ck$.

$W(n) \leq 1.6n + .2cn + .7cn = 1.6n + .9cn$

$= cn - .1cn + 1.6n$

$\leq cn$ if $c \geq 16$