Mediam and Order Statistics

- A computational problem that is similar to but "easier" than the sorting problem.
- Given n numbers, we are looking for the kth largest.
- If $k = 1$, or $k = n$, we are looking for the minimum or the maximum,
- If $k = \lfloor \frac{n}{2} \rfloor$, we are looking for the median.

Selection Problem

- INPUT: A set A of n (distinct) numbers and a number i.
- OUTPUT: The element $x \in A$, x is larger than exactly $i - 1$ other elements of A.
- Selection problem can be solved in $O(n \log n)$ time, since Sort and report.

Minimum and Maximum

- Find the Minimum or the maximum can be done in $\Theta(n-1)$ time.
- Lower bound is $\Omega(n)$ time.
- Simultaneous minimum and maximum, a naive approach $2n-2$, a better approach $3\lceil \frac{n}{2} \rceil$.

Selection in Expected Linear Time

A Divide and Conquer approach

Randomized_Select(A, p, r, i)
{
    $q =$ Partition(A, p, r);
    $k = q - p + 1$;
    if $i = k$ done;
    else if $(i \leq k)$
        then return Randomized_Select (A, p, q, i);
        else return Randomize_Select $(A, q + 1, r, i - k)$;
}

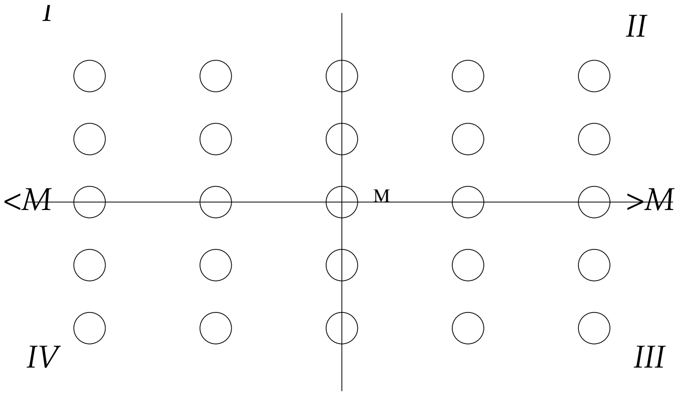
Selection in Expected Linear Time

- Worst case running time
  $T(n) = T(n-1) + n$,
- Best case running time, balance partition
  $T(n) = T(\lfloor \frac{n}{2} \rfloor) + n$

Selection in Worst Case Linear Time

1. We are looking for the kth largest.
2. Partition the n numbers into $\lceil \frac{n}{5} \rceil$ groups.
3. Find the median for each group.
4. Find the median of these medians, let it be M.
5. Rearange the groups so that the groups have medians $< M$ are to the left of M, the groups have medians $> M$ are to the right of M.

draw the figure


I
II
<M
M
>M
IV
III

1. There are 4 regions. We shall determine the region that **cannot** contains the answer (the kth largest one). Let $r = \text{Rank}(M)$.
2. Case 0, if $r = k$, found.
3. Case 1, if $r > k$, region III cannot have the answer.
4. Case 2, if $r < k$, region I cannot have the asnwer.
5. In either case, we can drop around $\frac{1}{4}n$.

1. Case 1, in the next iteration, we look for the kth largest in the set of $\frac{3}{4}n$ numbers.
2. Case 2, in the next iteration, we look for the $(k - r)$th largest in the set of $\frac{3}{4}n$ numbers.

$$T(n) = \begin{cases} \Theta(1), & \text{if } n \text{ is smaller than a given constant} \\ T(\frac{n}{5}) + T(\frac{3}{4}n) + n \end{cases}$$

- It is bounded by cn.