9

Medians and Order Statistics

i-th order statistic of a set of n elements: *i-th smallest element.*

minimum:	$i = 1$
median:	$i = \lfloor (n+1)/2 \rfloor$ or $\lceil (n+1)/2 \rceil$
maximum:	$i = n$

***The selection problem*:**

(For simplicity, assume that all elements in the input set are distinct.)

Input: $A[1..n]$ and i
Output: the i-th order statistic of A

9.1 Minimum (Maximum)

9-1x

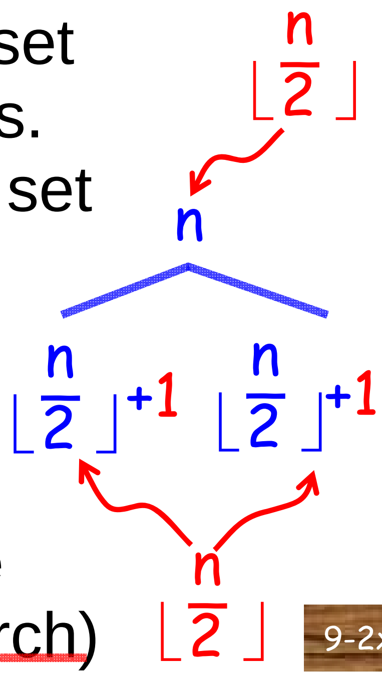
Minimum(A)
    min $\leftarrow A[1]$
    for $i \leftarrow 2$ **to** n **do**
        if $A[i]$<min **then** min $\leftarrow A[i]$
    return min

- $T(n)=O(n)$ (exactly n-1 comparisons)
- n-1 is optimal (every element loses at least once)

- Simultaneous minimum and maximum:

  Step 1: Perform $\lfloor n/2 \rfloor$ disjoint pairwise comparison.
  Step 2: Find minimum among the set containing the smaller elements.
  Step 3: Find maximum among the set containing the larger elements.

  $T(n)$: at most $3\lfloor n/2 \rfloor$ comparisons.

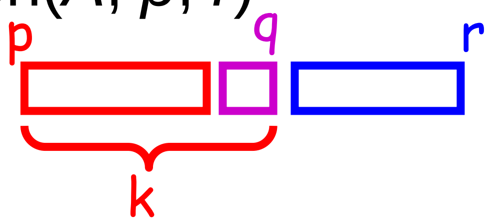


9.2 Selection in expected linear time
(divide & conquer, or prune-and-search)

9-2x

Randomized-Select(A, p, r, i)
  if $p=r$ **then return** $A[p]$
  $q \leftarrow$ Randomized-Partition(A, p, r)
  $k \leftarrow q-p+1$
  if $i = k$ **then return** $A[q]$
  elseif $i < k$
    then return Randomized-Select(A, p, $q-1$, i)
  else return Randomized-Select(A, $q+1$, r, $i-k$)

*assume $1 \le i \le r-p+1$



- Worst case: $T(n) = O(n) + T(n-1)$
  $= O(n^2)$.

9-2z

9-3x

the i-th one is always on the larger side

n-1 ⋮ 2 1 0   0 ⋮ n-3 n-2 n-1

- Average case:

$$E(n) = O(n) + \frac{1}{n} \sum_{1 \le k \le n} \underline{E(\max\{k-1, n-k\})}$$

$$= O(n) + \frac{2}{n} \sum_{\lfloor n/2 \rfloor \le k \le n-1} E(k)$$

$$= O(n)$$

(Prove it using substitution method by yourself.)

9-3a 9-3b

9.3 Selection in worst case linear time

(D&C, or prune and search)

Select(S, i)

9-3d

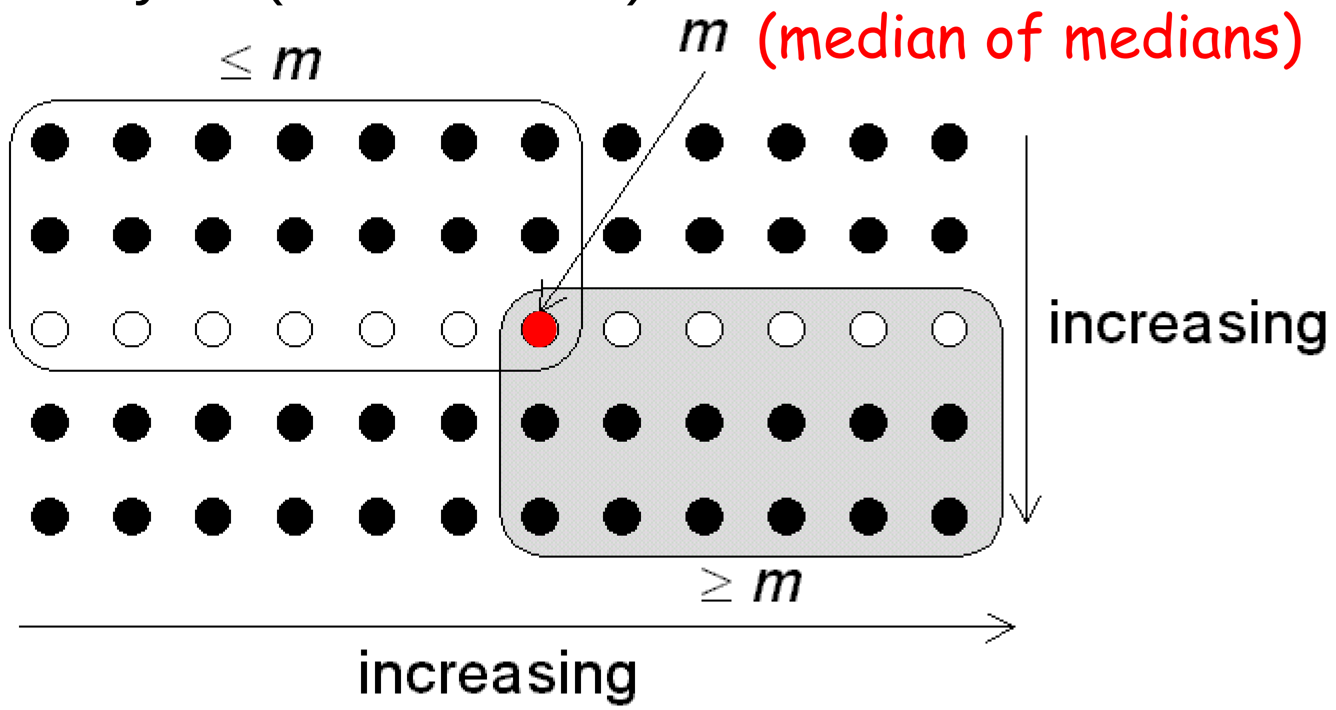
1. Divide S into n/r subsequences of r integers ($r \ge 5$, odd, constant).
   $= 5$
2. Sort every group. Let m_k be the median of the k-th subsequence.

   median of medians
3. Find the median m of m_k's. Select$(M, \lceil |M/2| \rceil)$
   (recursive call)
4. Partition S into three subsequences:
   $S_1=\{x|x<m\}$, $S_2=\{x|x=m\}$, $S_3=\{x|x>m\}$.
   Then, the i-th smallest element of S is located in one of the three subsequences.

recursive call

5. if $i \le |S_1|$ then Select(S_1, i)
   elseif $i \le (|S_1|+|S_2|)$ then return m
   else Select(S_3, $i-(|S_1|+|S_2|)$).

- Analysis (Assume r=5) 9-4a



At least one fourth of S is discard. ($|S_1| \le 3|S|/4$ and $|S_3| \le 3|S|/4$) (r = 5) 9-3d

Step	1.	2.	3.	4.	5.
Time	--	$O(n)$	$T(n/5)$	$O(n)$	$T(3n/4)$

$$T(n) \le T(n/5)+T(3n/4)+\Theta(n)$$

$$= O(n)$$ (Prove it by substitution method.) 9-4x

Homework: Ex. 9.1-1, 9.3-3, 9.3-6, 9.3-7, 9.3-8, 9.3-9, Pro. 9-1, 9-2, 9-3. Read Ch10 ~ 12.

*Every problem is worth studying! 9-4b