$\boxed{\text{Selection}}$

What is the selection problem?

It is the problem of finding the kth smallest key in an array of keys.

– smallest key: $k = 1$

– largest key: $k = n$

– median key: $k = n/2$

Solutions

† Find smallest, Find second smallest, ..., Find k'th smallest
$O(kn)$ time $\Longrightarrow$ total: $O(n^2)$

† (1) Sort the array
(2) find the key in the k'th place
$\Longrightarrow O(n\log n)$

Selection by partition

$Select(X, L, R, k)$

begin
    if $L < R$ *then*
        $M :=$ $partition(X, L, R)$
        if $(M - L + 1) > k$ *then*
            $Select(X, L, M - 1, k)$
        else if $(M - L + 1) < k$ *then*
            $Select(X, M + 1, R, k - (M - L + 1))$
end

Call Select$(X, 1, n, k)$
The kth smallest key is in the kth place of X

Complexity: worst case

$$T(n) = O(n^2)$$

Complexity: average case

$$T(n) = n - 1 + \frac{1}{n}\sum_{i=1}^{n} T(i-1)$$

With an analysis similar to that of quicksort, we can show that $T(n) = O(n)$.

Note: If $T(n) = n - 1 + \frac{2}{n} \sum_{i=1}^{n} T(i-1)$, then $T(n) = O(n\mathrm{log}n)$.

Do we really need to use sorting?

Can we do better than $O(n \log n)$ in worst case?

A selection algorithm

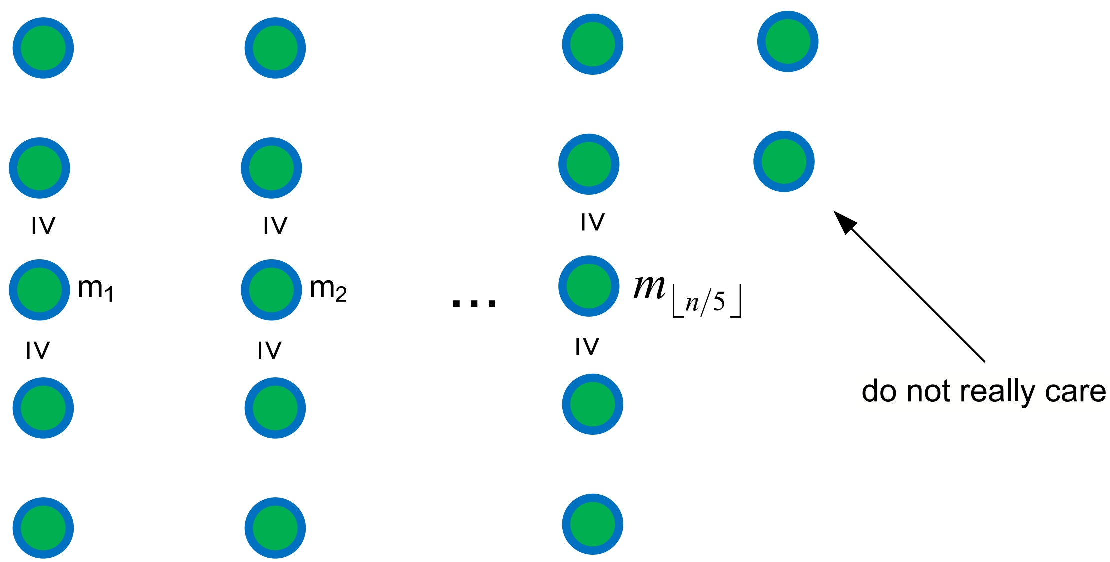
Input: S: a set of n keys

k: an integer such that $1 \le k \le n$

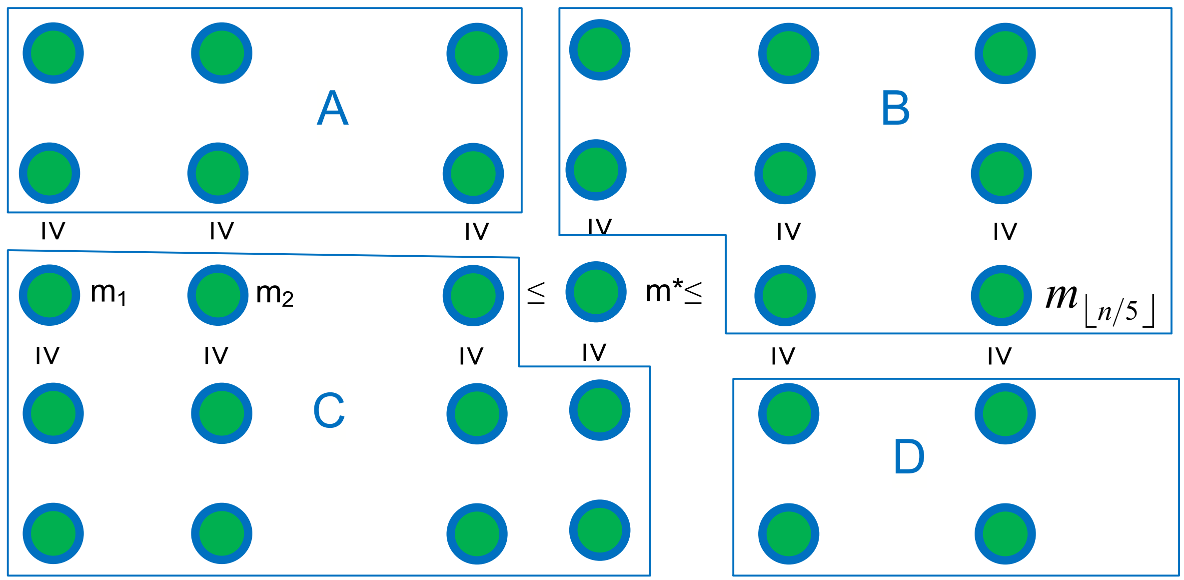
Output: the kth smallest key in S.

Method: Select(S, k)

1. Divide the keys in S into sets of five each and find the median of each set.



2. Let M be the set of all medians.
Find the median m^* of M. (How? Recursively!)
Select $(M, \lceil |M|/2 \rceil)$ $\qquad |M| = n/5$



$\Longrightarrow$ all keys in $C \leq m^*$ and all keys in $B \geq m^*$

Number of keys in $A \cup B \cup D \leq$

$$\leq \frac{2}{5}n + (\frac{3}{5}n)/2 = \frac{2}{5}n + \frac{3}{10}n = \frac{7}{10}n.$$

Similarly, number of keys in $A \cup C \cup D \leq \frac{7}{10}n$.

3. Construct

$$S_1 = C \cup \{x|\ x \in A \cup D \text{ and } x \leq m^*\} \cup \{m^*\}$$
$$S_2 = B \cup \{x|\ x \in A \cup D \text{ and } x > m^*\}$$

4.

If $k = |S_1|$ then

m^* is the kth smallest key.

else if $k < |S_1|$ *then*

Select(S_1, k) $\qquad$ $(|S_1| \leq \frac{7}{10}n)$

else $(k > |S_1|)$

Select$(S_2, k - |S_1|)$ $\qquad$ $(|S_2| \leq \frac{7}{10}n)$

Complexity (worst case)

Step 1) $\frac{6}{5}n$ key comparisons

Step 2) $T(\frac{n}{5})$ key comparisons

Step 3) $\frac{2}{5}n$ key comparisons

Step 4) $T(\frac{7}{10}n)$ key comparisons

Time complexity:

$$T(n) \leq 6, \quad n \leq 5$$

$$T(n) = \frac{6}{5}n + T(\frac{1}{5}n) + \frac{2}{5}n + T(\frac{7}{10}n), \quad n > 5$$

We can solve it by guessing: $T(n) \leq 16n$

Proof:

$n \leq 5$, obvious.

Assume it is true for all integers less than n.

$$T(n) = \frac{6}{5}n + T(\frac{1}{5}n) + \frac{2}{5}n + T(\frac{7}{10}n)$$
$$\leq 1.2n + 0.2n \times 16 + 0.4n + 0.7n \times 16$$
$$= 1.6n + 0.2n \times 16 + 0.7n \times 16$$
$$= 16n$$

Find the median of a set of five keys (with 6 comparisons)

x_1, x_2, x_3, x_4, x_5

(1) $X_1 : X_2$

$$\begin{array}{c} X_1 \\ \geq \\ X_2 \end{array}$$

(2) $X_3 : X_4$

$$\begin{array}{c} X_3 \\ \geq \\ X_4 \end{array}$$

(3) $X_1 : X_3$

$$\begin{array}{c} X_1 \\ \geq \\ X_3 \end{array} \Longrightarrow \begin{array}{lll} X_1 & \geq & X_2 \\ \geq & & \\ X_3 & & \\ \geq & & \\ X_4 & & \end{array} \Longrightarrow X_1 \text{ is out}$$

(4) $X_2 : X_5$

$$\begin{array}{c} X_2 \\ \geq \\ X_5 \end{array}$$

(5) $X_2 : X_3$

$$\begin{array}{c} X_3 \\ \geq \\ X_2 \end{array} \Longrightarrow \begin{array}{lll} X_3 & \geq & X_4 \\ \geq & & \\ X_2 & & \\ \geq & & \\ X_5 & & \end{array} \Longrightarrow X_3 \text{ is out}$$

(6) $X_2 : X_4$

$$\begin{array}{c} X_2 \\ \geq \\ X_4 \end{array} \Longrightarrow \begin{array}{lll} X_2 & \geq & X_4 \\ \geq & & \\ X_5 & & \end{array} \Longrightarrow X_2 \text{ is median}$$

OR

$$\begin{array}{c} X_4 \\ \geq \\ X_2 \\ \geq \\ X_5 \end{array} \Longrightarrow X_4 \text{ is median}$$