K-th Order Statistics

- FIND-MIN is $\Omega(n)$ ➡ Select is $\Omega(n)$
- FIND-MIN is O(n) doesn't imply an upper bound on the general select problem.
- Theorem: The general select problem can be solved in linear time.

CS1409-1

Select(S,*k*)

- Choose *x* from S
- Let $S_1 = \{y \in S \mid y < x\}$
- Let $S_2 = \{y \in S \mid y > x\}$
- If $||S_1|| \geq k$ then return Select(S_1,*k*)
- Else if $||S_1|| = k-1$ then return *x*
- Else return Select(S_2, $k-||S_1||-1$)

CS1409-2

Analysis of Select(S,k)

In worst case x is min or max and we get

$$T(n) \leq T(n-1) + cn = \Theta(n^2)$$

CS1409-3

Suppose …

- Suppose we could choose *x* so that the recursive call is always on a set of size n/b, where b is constant.
- Then $T(n)=T(n/b)+cn$

$$= cn \Sigma_{i=0,\ldots,\log_b n} 1/b^i = O(n)$$

CS1409-4

To warm up …

Suppose we could choose x so that the recursive call is **typically** on a set of size *n/b*, where *b* is constant.

CS1409-5

"Average Case Analysis"

- Deterministic Algorithm: The expected running time when the input is chosen uniformly at random from all inputs of size *n*.
- Randomized Algorithm: The expected running time for worst-case input of size *n*.

CS1409-6

Random-Select(S,k)

- Choose x randomly from S
- Let $S_1 = \{y \in S \mid y < x\}$
- Let $S_2 = \{y \in S \mid y > x\}$
- If $\|S_1\| \geq k$ then return Select(S_1,k)
- Else if $\|S_1\| = k-1$ then return x
- Else return Select(S_2,k-$\|S_1\|$-1)

CS1409-7

Some useful facts

Assume WLOG that $S=\{1, \ldots, n\}$

- What is the probability that $x=i$?
- What is the probability that $\|S_1\|=i$?
- What is the probability that $\|S_2\|=i$?
- What is the probability that $\|S_1\|=i$ and $\|S_2\|=j$?
- What is the probability that a recursive call is made on an instance of size i?

CS1409-8

Analysis of Random-Select

$$
\begin{aligned}
E[T(n)] &\leq (1/n)\, \Sigma_{i=1,..n}\, T(\max(i-1,n-i)) + cn \\
&\leq (2/n)\, \Sigma_{i=\text{ceil}(n/2)\, \ldots\, n-1}\, T(i) + cn \\
&= O(n)
\end{aligned}
$$

CS1409-9