Order Statistics

We begin our discussion by considering a particular problem in identifying a specific element in an unordered list based on its position in a corresponding ordered list.

Definition: The ith order statistic of a set of n elements is the ith smallest element in the set.

Definition: The minimum of a set is the first order statistic.

Definition: The maximum of a set is the nth order statistic.

Definition: The median of a set is the midpoint in the ordered set such that there are as many elements less than the median as there are greater than the median.

Given our definition of order statistics, the problem we will solve here is called the "selection" problem.

Definition: The selection problem is defined as follows:

- Input: A set $\mathbf{A}$ of n distinct elements and a number i where $1 \leq i \leq n$.
- Output: The element that is greater than exactly $i - 1$ other elements in $\mathbf{A}$.

As a naïve approach, the selection problem can be solved in $O(n \lg n)$ time by first sorting the elements in $\mathbf{A}$ and then selecting the ith element in the sorted set. This appears to be pretty good, but we will consider algorithms that improve on this bound.