Sorting and selection – handout 1

March 17, 2014

In class (March 10 and March 17) we discussed the deterministic sorting algorithm *insertion sort* that uses roughly $n \log n$ comparisons (which is nearly best possible) and the randomized sorting algorithm *quicksort*, for which we showed a proof that the expected number of comparisons is roughly $2n \log n$.

We presented a simple randomized algorithm for selection the median that makes $O(n)$ comparisons in expectation, and a more complicated one that makes roughly $\frac{3n}{2}$ comparisons. In passing we encountered some important principles for probabilistic analysis, namely, linearity of expectation, and concentration bounds for independent random variables.

We also presented a deterministic median selection algorithm that uses $O(n)$ comparisons, and showed that every such algorithm needs to make at least $\frac{3n}{2}$ comparisons.

Other principles encountered are *divide and conquer algorithms* (the use of the splitting item for quicksort), and *Yao's principle* for lower bounding the expected running time of randomized algorithms.

No class on March 24, due to the *Israel CS Theory Day* in the Open University. (Participation is free, registration is requested. See http://www.openu.ac.il/theoryday2014/.)

Homework - and in (either in Hebrew or English) by Monday March 31.

1) We have seen in class a deterministic algorithm for selecting the median based on partitioning the items into groups of size 5. The number of comparisons it uses is at most $20n$ (up to low order terms). Design a similar algorithm based on partitioning the items into groups of size 7, and prove an upper bound better than $20n$ on the number of comparisons that it makes. (If you are not making wasteful comparisons, you should be able to prove a bound of roughly $16n$.)

2) We have seen a proof that every deterministic comparison-based median selection algorithm needs to make at least $3n/2$ comparisons (up to low order terms) in the worst case. Prove for some $\delta > 0$ of your choice (say, $\delta = \frac{1}{10}$) that for every deterministic median selection algorithm, the expected number of comparisons made when the input is a random permutation is at least $(1 + \delta)n$ (up to low order terms). (Rather that fixing a random permutation in advance, it may be more convenient for the proof to construct the random permutation "on the fly". Whenever the algorithm makes a comparison that involves an item that has not been involved in any previous comparison, at that point the item is given a random value in $\{1, 2, \ldots, n\}$, among the values not given to previous items.)