Fall 2008: **COT 5407** Intro. to Algorithms
[Programming Assignment 1; Due Dec 2 at start of class]

General submission guidelines and policies: Add the following signed statement. Without this statement, your homework will not be graded.

> I have adhered to the collaboration policy for this class. In other words, everything written down in this submission is my own work. For problems where I received any help, I have cited the source, and/or named the collaborator.

Read the handout on **Homework guidelines and collaboration policy**.

Problem Description

Implement a program that reads in a list of n numbers and outputs the **median** value in the list. Note that the median is defined as the item with rank $\lfloor \frac{n}{2} \rfloor$.

Your program should use the following algorithms to figure out the answer. If you like, you may devise more algorithms. You will earn extra credit if the additional algorithm(s) have some innovative ideas in terms of the design or the experiments performed. The program should output the median value and the time it takes using each of the methods. Watch out! There may be duplicates in the input.

The algorithms are as follows:

Algorithm 1 (Naive) For every item in the list, compute its rank. Stop as soon as you have found the item with the correct rank. The rank can be computed by adding one to the count of the number of items smaller than it. However, you have to figure out how to deal with duplicates, and explain this in your report.

Algorithm 2 (Smart1) Sort the input list and report the output.

Algorithm 3 (Smart2) Use the Randomized Select algorithm we discussed in class.

Algorithm 4 (Smart3) Use the Improved Select algorithm we discussed in class.

You have to run the program on the data file `data.txt` for different values of n. Choose $n = 16, 32, 64, 128, 256, 512, 1024$ and 2048 (and higher, if you like). For the Randomized-Select algorithm, run the algorithm on 10 different choices of a random pivot (perhaps by picking a different seed for the random number generator). Write down the time taken by the three algorithms. In the case of the Randomized Select algorithm, take the average of 10 runs. Instead of time, it may also be reasonable to count the number of comparisons and plot it. To compute the time taken, you may ignore the time taken to read in the data, since this is common to all the methods.

Finally, you need to write a report discussing your results, i.e., explain what you learnt from the experiments and why. Compare the performances of the algorithms implemented.

Notes

Which sorting algorithm should you use in Algorithm 2? Something to think about! You could plot the run times for the algorithms as a graph. Why is it convenient to use powers of two for n? You could also plot the curves n, $n \lg n$ and n^2 for each value of n on the same plot. How will that help? Do the run times confirm what you know from your theoretical analysis?

You may use C++, C, or Java to do your task (any other languages will require my permission). Make sure that your algorithm works correctly even if the input array has duplicates; in particular it should work correctly even if there are duplicate entries for the median.

You are given one data file with a large number of integers in it. For different values of n, simply use the first n values from the same file. Document your program well and print out the source code and the output for submission.

Write a short (at most 4 pages) report summarizing your conclusions from the study. This summary is a very important part of your project.

You may not work in teams. This is a solo project.

Report

Your report should contain a brief description of the algorithms. It should mention any special ideas and/or tricks used in your program. It should report known bugs. You should provide the specifications of the computer that was used for the execution times mentioned in the report. Your discussion should not merely specify the results. Instead, it should discuss the observations. Are your observations consistent with the theoretical predictions, and why? Were there any surprises? Can you provide possible explanations for them?