IE531: Algorithms for Data Analytics
Spring, 2017
Programming Assignment 2: Randomized Version of the Algorithm that picks the k-th Smallest Element in an Unordered List of Numbers
Due Date: February 24, 2017

This assignment is an easy one. You are going to modify the code on Compass for the *Deterministic Selection Algorithm* (i.e. the one that uses the Median-of-Medians etc.) to pick the k-th smallest element in an ordered list of numbers. I want you to compare your running time with the *QuickSort Based Algorithm* (i.e. you sort the list and pick the k-th smallest from the sorted-list) with that of the Randomized Version. It has to take three command-line inputs, (1) the length (n) of the unordered-list of numbers[1] (2) the number (m) of trials (because the randomized-version will have different running times for each trial) and (3) the name of the CSV output file which contains the runtime data for the m-many trials of picking the $\lceil \frac{n}{2} \rceil$-th smallest number among the n-many numbers.

In your numerical experiments you will vary n from $100,000$ to $500,000$ in five steps. For each choice of n, you will pick the $\lceil \frac{n}{2} \rceil$-th smallest number using the randomized-version of the deterministic-select (i.e. median-of-medians) algorithm. This will be done m-many times (in my experiments I picked $m = 1000$). You will plot the histograms of the running-times for $n = 100,000, 200,000, \ldots, 5000,000$ (cf. figure 1). You will also plot the mean-running-time as a function of n (cf. figure 2).

In addition to your C++ code, you will send me two plots (along the lines of figures 1 and 2). Watch the video instructions on Compass for additional details of how this can be done in MATLAB.

[1] Set $k = \lceil \frac{n}{2} \rceil$ for your numerical experiments, after you know what n is.

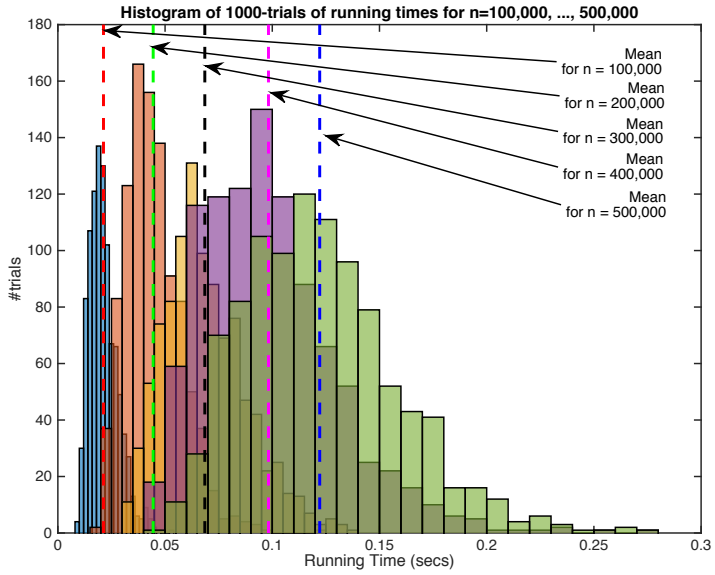

Figure 1: Histogram of 1000-many runs of the randomized algorithm for $n = 100,000, 200,000, \ldots, 500,000$.)

.

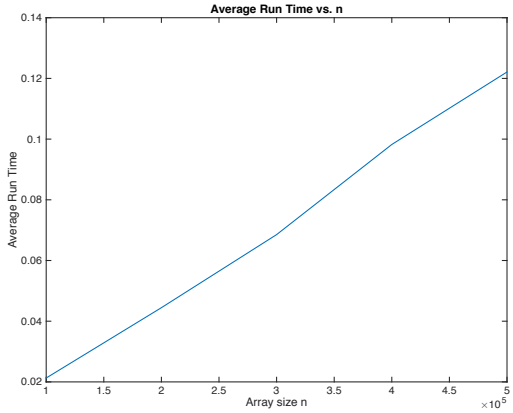

Figure 2: Average Run-Time vs. n ($n == 100,000, 200,000, \ldots, 500,000$). Notice the linear-relationship, as predicted by theory.

.