Second Mandatory Project in DM851 (2016)

This project is meant to connect empirical results in combinatorics to analytical results.

1. Read Chapter 1 (focus on quicksort) and Chapter 3.3 of [1].

2. Infer the OGF
$$C(z) = \frac{2}{(1-z)^2} \ln \frac{1}{1-z}$$
and
$$[z^N]C(z) = 2(N+1)(H_{N+1} - 1)$$
for the average number of comparisons used by quicksort. Start from Equation (1) on page 109 (no need for a write-up of this part, make sure you understand the individual steps, no written documentation needed). Note, that the harmonic number H_N can be approximated : $H_N \approx \ln N + 0.57721$.

3. Read Chapter 3.9 of [1], make sure you understand how to infer the expected value for a uniform distribution (all number $0, 1, \ldots, n-1$ being equally likely) and its variance by applying Theorem 3.7 (no written documentation needed, page 130).

4. Back to Quicksort: Given the Probability GF
$$Q_N(u) = \frac{1}{N} \sum_{1 \le k \le N} u^{N+1} Q_{k-1} Q_{N-k}(u)$$
for the number of comparisons used by quicksort, on can via BGFs infer the variance $\approx N^2 \cdot (7 - \frac{2\pi}{3})$ for the number of comparisons used by quicksort (see page 138).

   (a) Implement Quicksort in a programming language of your choice (see Program 1.2 in [1]).

   (b) Sort $k \in \{10000, 20000, \ldots, 100000\}$ integers and count the numbers of comparisons you needed, repeat this experiment 100 times for each k and compute the mean as well as the variance. Make a plot similar to Fig. 1.1 in [1], that depicts the predicted number of comparisons, as well as the empirical mean of number of comparisons (use the approximation for H_N given above).

   (c) Sort $N = 2000$ integers and repeat this at least 10000 times. Make a histogram similar to Fig. 1.4 in [1] (the shaded area and the distribution plot is not necessary). What is the empirical percentage of trials that are max one standard deviation from the mean? What is the analytical percentage of trials that are max one standard deviation from the mean?

Submit your code and a small report (minimum requirements are, that the plots for Fig. 4b and 4c as well are given and the questions in 4c are answered). Submit before December 22nd, 10am.

References

[1] Sedgewick, R. and Flajolet P., An Introduction to the Analysis of Algorithms, 2nd edition, Addison-Wesley, 2013.