

22C : 031 Algorithms
Homework 7
Due Thursday, December 11

You are expected to do the homework assignments on your own without consulting human and non-human sources (like web pages) for the solutions.

In this assignment, you should implement the lossy counting algorithm of Manku and Motwani that we covered in Lecture 18 (see lecture notes). You can also find the algorithm described in Section 4.2 of their paper. (See lecture notes page for a link.) Note that the authors use s to denote what we denoted by Θ .

You should choose $\Theta = 1/1000$ and $\varepsilon = 1/10000$ in your implementation. The stream is a sequence of numbers from a file – your program should obtain the stream by reading these numbers one by one. Your program should have one argument which is the name of the file being read. Assume that this file is in the same directory as your program.

Three test files, each consisting of 10^7 integers in the range $[0, 10^6]$, have been posted.

Your program should output two things: (a) the frequently occurring elements at the end of the entire stream – recall that this is not all elements stored in the data structure at the end but only a subset; (b) the maximum number of elements stored in the data structure at any point in the processing of the stream.

Besides the source code, submit three files each of which contains the output of your program on each of the three test files. Zip all the files you are turning in and submit a single zipped file into ICON.