Your name
CS 261, Spring 2021
Homework 4
Due Monday, April 26

1. Many sites for reporting coronavirus infection rates smooth out the numbers by reporting the average numbers over a window (typically a week) rather than day-to-day. Describe an algorithm that takes as input an array of n numbers (the daily infection rates, in the coronavirus application) and a number k (the window size), and outputs the sequence of the averages of the input numbers over windows of length k. Your algorithm should use only $O(1)$ extra space beyond the space for the input.

(Hints: Use the turnstile-model streaming algorithm for the average from the lecture notes. Your algorithm is allowed to read values from the input array in any order; it is not required for it to process the input as a stream. However, this array is read-only, and cannot be used to hold other information than the input. The output should be produced by a sequence of output commands; it is not allowed to store it in a temporary array and then return the array, because that would use too much extra space.)

■

2. If we use reservoir sampling to select a random sample of three numbers out of a stream of n numbers, what is the probability, in the limit for large n, that the median of the stream lies between the smallest and largest of the three sampled numbers?

■

3. Suppose we want to solve the following modified version of the majority problem:

- Process a cash register stream of elements
- If more than 60% of the elements have the same value, return that value
- If there is no value shared by 40% or more of the elements, return "None"
- In the remaining cases, any behavior is acceptable

Describe how to use the heavy hitters structure to solve this problem using only $O(1)$ words of memory.

■

4. The last part of this week's lecture notes describes the invertible Bloom filter, a somewhat complicated data structure that can report up to k stragglers (elements that were inserted and then never removed) in a turnstile data stream, for a given parameter k. At least of the other data structures in the lecture notes can solve the same problem much more simply for the special case $k = 1$. Which data structure, and how?

(If it helps, you can assume that the elements are numbers.)

■