Sublinear Time and Space Algorithms 2020B – Problem Set 1

Robert Krauthgamer

Due: May 11, 2020

**General instructions:** Please keep your answers short and easy to read. You can use results, calculations or notation seen in class without repeating them, unless asked explicitly to redo them.

1. Design a streaming algorithm that at any point \( m \) (not known in advance) receives a query \( S \subset [n] \) and outputs an estimate what fraction of items in the stream belong to \( S \) within additive error \( \epsilon \). Note that \( S \) is given only at query time (not in advance).

   Hint: Maintain \( O(1/\epsilon^2) \) random samples and use them to estimate the fraction in \( S \).

2. Suppose we are guaranteed that some item in the stream \( \sigma_1, \ldots, \sigma_m \) appears more than half the time, i.e., there exists (unknown) \( i \in [n] \) with frequency \( x_i > m/2 \). Design a streaming algorithm with space complexity \( O(\log n) \) bits that finds this item \( i \). Next, extend your algorithm to output also a \( (1 + \epsilon) \)-approximation to its frequency \( x_i \). Make sure to clearly state the space complexity of your algorithms.

   Hint (for the basic version): Store only two items.