University of Edinburgh

INFR11156: Algorithmic Foundations of Data Science (2025)

Homework 3

Problem 1: Prove that the medium of the returned values from $\Theta(\log(1/\delta))$ independent copies of the BJKST algorithm gives an (ε, δ)-approximation of F_0.

Problem 2: Let $Y_1, \ldots, Y_n$ be independent random variables with $\mathbb{P}[Y_i = 0] = \mathbb{P}[Y_i = 1] = 1/2$. Let $Y := \sum_{i=1}^{n} Y_i$ and $\mu := \mathbb{E}[Y] = n/2$. Apply the uniform Chernoff Bound to prove it holds for any $0 < \lambda < \mu$ that

$$\mathbb{P}[Y \geq \mu + \lambda] \leq \mathrm{e}^{-2\lambda^2/n}.$$

Problem 3: For any undirected graph $G = (V, E)$ with n vertices, we say three vertices u, v, w form a triangle if there are three edges connecting u, v, w respectively. This problem is to analyse a streaming algorithm for approximately computing the number of triangles in an undirected graph G. To describe the proposed algorithm, let $\mathcal{H}$ be a family of 12-wise independent hash functions, where every $h \in \mathcal{H}$ is of the form $h : V \to \{-1, 1\}$. Let Z be our estimator, which is set to be 0 initially. The algorithm is described as follows:

Algorithm 1 Approximate the number of triangles in G

Pick a function h uniformly at random from $\mathcal{H}$;
$Z \leftarrow 0$;
while an edge $\{u, v\}$ arrives **do**
    $Z \leftarrow Z + h(u) \cdot h(v)$;
end while
Return $Z^3/6$.

You need to prove that the returned value $Z^3/6$ is an unbiased estimator of the number of triangles in G, i.e.,

$$\mathbb{E}\left(\frac{Z^3}{6}\right) = \text{number of triangles in } G.$$

Hence, the number of triangles can be approximately counted by running Algorithm 1 above multiple times in parallel and returning the medium of the returned values.

Problem 4: We are given two independent streams of elements from $\{1, \ldots, n\}$, and we only consider the cash register model. Let $A[1, \ldots, n]$ and $B[1, \ldots, n]$ be the number of occurrences of item i in two streams, respectively. Design a streaming algorithm to estimate $X = \sum_{i=1}^{n} A[i]B[i]$ with additive error $\varepsilon \cdot \|A\|_1 \cdot \|B\|_1$. You need to analyse the space complexity of your proposed algorithm, and analyse the correctness of your algorithm.