University of Edinburgh

INFR11156: Algorithmic Foundations of Data Science (2019)

Homework 6

Problem 1: We discussed in class an efficient construction of a family of k-wise independent hash functions h such that it holds for any x that $\mathbf{E}[h^i(x)] = 1$ if $i \geq 1$ is an even number, and $\mathbf{E}[h^i(x)] = 0$ otherwise. In this question, you need to construct a family of k-wise independent hash functions g such that it holds for any x that $\mathbf{E}[h^i(x)] = 1$ if i is divisible by 3, and 0 otherwise.

Problem 2: 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. 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 in Algorithm 1 below. Prove that the returned value $Z^3/6$ is an unbiased estimator of the number of triangles in G, i.e.,

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

Algorithm 1 Approximate number of triangles

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

Problem 3: 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.