Randomized Algorithms 2023A – Problem Set 3

Robert Krauthgamer and Moni Naor

Due: Feb. 13, 2023

1. We saw in class a randomized algorithm (call it Algorithm $E+$) with success probability $3/4$ for counting DNF solutions. (Here, success means that the output is within factor $1 \pm \varepsilon$ of $|S|$, the number of satisfying assignments of the input DNF formula.)

   Consider now Algorithm $E++$, which repeats that Algorithm $E+$ independently $t = O(\log \frac{1}{\delta})$ times and then reports the *median* of their outputs. Prove that Algorithm $E++$ has success probability $1 - \delta$ (again, success means that the output is within factor $1 \pm \varepsilon$ of $|S|$).

   Hint: Use concentration bounds to count the number of "successes".

2. Prove the following three lemmas (mentioned in the lecture notes), from the construction seen in class of a strong coreset Y of a set $X \subset \mathbb{R}^d$ (for 1-median) via Importance Sampling. Recall that Y is a multiset of m iid samples from this distribution, and $\hat{Z}$ is the Importance Sampling estimator (without repetitions), i.e., $\hat{Z} = \frac{\|y-c\|}{q(y)}$ for one sample $y \in Y$ and an arbitrary $c \in \mathbb{R}^d$.

   Lemma 5: $\hat{Z} \le S(X) \cdot \mathbb{E}[\hat{Z}]$ with probability 1.

   Hint: It is quite immediate.

   Lemma 6: The success probability in Lemma 4 can be improved to $1 - \delta$ by using $m \ge L\varepsilon^{-2} \log \frac{1}{\delta}$ for a suitable constant $L > 0$, i.e.,

   $$\forall c \in \mathbb{R}^d, \qquad \Pr\left[\frac{1}{m} \sum_{y \in Y} \frac{\|y - c\|}{q(y)} \in (1 \pm \varepsilon) \sum_{x \in X} \|x - c\|\right] \ge 1 - \delta.$$

   Hint: Use a concentration bound and Lemma 5.

   Lemma 8: If $m \ge L\varepsilon^{-2} \log \frac{1}{\delta}$, then $w(Y) = \frac{1}{m} \sum_{y \in Y} \frac{1}{q(y)}$ satisfies

   $$\Pr[w(Y) \in (1 \pm \varepsilon)n] \ge 1 - \delta.$$

   Hint: Show that all $x \in X$ satisfy $\frac{1}{q(x)} \le O(n)$, then use a concentration bound.

3. Show that in the Cuckoo Graph when $r = 2n$ it is not likely that there are components of size larger than $O(\log n)$.

4. Recall that we used the method of compression to analyze Cuckoo Hashing. Use the method to show that with high probability, a random graph $G_{n,1/2}$ oes not contain a clique or independent set of size larger than $O(\log n)$.