CS60029: Practice Problem Set 2

September 13, 2026

---

1. Consider the following variant for approximating the DNF counting problem. The t-th trial of this algorithm first picks a random clause C_t. where the probability of choosing a clause is proportional to the number of satisfying truth assignments for it. Next, it selects a random satisfying truth assignment a for the chosen clause. Define the random variable $X_t = 1/|\text{cov(a)}|$, where cov(a) denotes the set of clauses that are satisfied by the truth assignment a. The estimator for the number of satisfying assignments is the random variable

$$Y = \eta \times \sum_{t=1}^{N} \frac{X_t}{N}$$

   where η denotes the sum of the sizes of the coverage sets over all possible truth assignments. Prove that Y is an (ε, δ)-approximation for the number of satisfying assignments when $N = O\left(\frac{m}{\varepsilon} \ln \frac{1}{\delta}\right)$.

2. Show that a k-universal hash family is also a $(k-1)$-universal hash family.

3. Generalize the construction of the 2-universal hash family discussed in the class and give an example of a k-universal hash family with its proof of k-universality.

4. Give an example of a k-universal hash family that is not a $(k+1)$-universal hash family for any integer $k \geqslant 3$. Prove your claim.

5. Suppose a stream of length m over an universe of size n is distributed over k servers. Explain how using the Count-Min Sketch algorithm, you can compute the approximate frequencies of the elements with polylogarithmic communication complexity (the number of bits communicated). What is the communication complexity of your algorithm?

6. Prove that there is no algorithm with polylogarithmic (in n and m) space complexity that outputs the frequency of any element queried at the end of an m length stream over a universe of size n.

7. Recall the definition of Cuckoo hashing discussed in the class, where we maintain an array H of size b and two hash functions h_1 and h_2 where $h_1, h_2 : U \to \{0, 1, \ldots, b-1\}$. Also recall the definition of Cuckoo graph G.

   (a) Let $E_{i,j,\ell}$ be the event such that the shortest from vertex i to vertex j is ℓ. Then,

$$\mathbb{P}[E_{i,j,\ell}] \leqslant 3^{-\ell}/b.$$

   Using this, show that, $\mathbb{P}[G \text{ contains a cycle}] \leqslant 1/2$.

   (b) Show that, the expected insertion time is $O(1)$ if there are no rehashes.

   (c) Show that, the expected insertion time is $O(1)$ in Cuckoo hashing.

8. Let $X_1, \ldots, X_n$ be independent and identically distributed geometric random variables with parameter $p \in (0, 1)$. Prove that

$$\mathbb{E}\left[\max_{i=1}^{n} X_i\right] = O\left(\frac{1}{p} \log n\right).$$

9. Prove that, the Johnson-Lindenstrauss map not only approximately preserves the interpoint distances, it also approximately preserves angles between points.

10. In the proof of Johnson-Lindenstrauss lemma presented in the class, we only proved the upper bound. Modify that proof to prove the other direction. Specifically, for i.i.d. standard Gaussians $X_1, \ldots, X_n$ and $\alpha = (1-\varepsilon)^2 n$, prove that

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i^2 < \alpha\right] \leqslant e^{-cn\varepsilon^2},$$

where $c > 0$ is constant.