Homework 8

Due: 20 November, 2020

Each problem is graded on the coarse scale of $\checkmark^+$, $\checkmark$, $\checkmark^-$ and no $\checkmark$. It is also assigned a multiplier, denoting the relative importance of the problem. Both correctness and presentation are grading criteria.

Please read and make sure you understand the collaboration policy on the course missive. The problems for 2000-level credit are clearly marked below: students who are registered for 1000-level credit are welcome to solve and submit these problems for extra credit (see course missive for details).

Remember to prove all your (non-elementary) mathematical claims, unless stated otherwise.

Each pair of students should submit only 1 pdf to the corresponding Canvas assignment.

Problem 1

(*a*) (1 $\checkmark$) Show that a k-wise independent hash family is also $k-1$-wise independent, for $k > 1$.

(*b*) (1 $\checkmark$) Given a streaming algorithm that estimates some parameter α to accuracy ϵ with probability at least $\frac{2}{3}$ using s bits of space, how would you construct a new streaming algorithm that estimates α to accuracy ϵ with probability at least $1-\delta$? What is the space complexity of your new algorithm? Credit will only be given for solutions with a reasonable dependence on δ.

Problem 2

(3 $\checkmark$s)

Complete the proof of Theorem 17.6 in class, that Algorithm 17.5 is a streaming algorithm for the Count-Distinct problem using space $O(\frac{1}{\epsilon^2}\log n)$. You should analyse Algorithm 17.5 as using a pairwise independence hash family instead of making the simplifying assumption of using the uniformly random function family.

Feel free to modify the algorithm slightly to handle edge cases, if necessary.

(Hint: the interesting general case is when $\epsilon \geq \frac{1}{\sqrt{n}}$.)

Problem 3

(3 $\checkmark$s)

Consider an adversarially ordered stream σ consisting of m distinct elements in $[n]$. Give a log-space streaming algorithm that returns an element x such that the quantile of x in the stream is in $\alpha \pm \epsilon$ with probability at least $\frac{2}{3}$, for known parameters $\epsilon > 0$ and $\alpha > 0$. You may assume that α and ϵ are constants independent of $m, n \to \infty$, and in particular that you do not need to worry about rounding/indexing issues related to taking quantiles of a finite set.

Problem 4

(3 ✓s, **extra credit**)

Consider an arbitrary hash family $\mathcal{H} : U \to V$. In particular, we make absolutely no assumptions about $\mathcal{H}$, so no pairwise independence and no universality. Show that there must exist a pair of elements $x, y \in U$ such that

$$\mathbb{P}_{h \leftarrow \mathcal{H}}(h(x) = h(y)) \geq \frac{1}{|V|} - \frac{1}{|U|}$$

(Hint: what we learnt in the uniformity testing lecture might be helpful.)