Homework 6

Due: 30 October, 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

The following are a few short questions.

(*a*) (1 $\checkmark$) Is it possible for a uniformity tester to have 1-sided error? If so, give a construction. If not, give a proof why not.

(*b*) (1 $\checkmark$) Show that a discrete distribution over $[n]$ can be learnt to within ϵ error in Hellinger distance using $O(n/\epsilon^4)$ samples, with constant probability of success. (Note: it can in fact be learnt in $O(n/\epsilon^2)$ samples, but is more "non-trivial" to show.)

Problem 2

(2 $\checkmark$s)

A *symmetric* property $\mathcal{P}$ (over the domain $[n]$, say) is a set of distributions such that, if a distribution $\mathbf{p}$ is in $\mathcal{P}$, then for any permutation ρ on $[n]$, $\mathbf{p} \circ \rho$ is also in $\mathcal{P}$.

Given a set of k samples, the *empirical fingerprint* is a function $h : [k] \to [n]$ that encodes that there are $h(j)$ distinct support elements that appear exactly j times in the set.

Show that an ϵ-tester for a symmetric property can, without loss of generality, depends only on the empirical fingerprint instead of the set of samples it sees.

Problem 3

Consider the following formulation of (unweighted, undirected, simple) graph testing for the property of degree regularity.

Graph access You can only sample edges uniformly (and independently) at random, and you get the set of 2 endpoints $\{u, v\}$

Distance from regularity Given a graph G over n vertices, consider its length-n sequence of degrees d_i. Also let the average degree be $\bar{d}$. The distance of G from regularity is defined as

$$\frac{1}{n\bar{d}} \sum_i |d_i - \bar{d}|$$

(*a*) (2 ✓s) Give and prove the correctness/efficiency of a tester using $O(\sqrt{n}/\epsilon^2)$ samples.

(*b*) (2 ✓s) Show that any tester must require $\Omega(\sqrt{n})$ samples even for constant ϵ.

Problem 4

(3 ✓s, **Extra credit**)

Consider the *monotonicity* property over the domain $[n]$: a distribution $\mathbf{p}$ is monotonic/non-decreasing if $p_i \leq p_j$ for all $i \leq j \in [n]$.

Show an $\Omega(\frac{\sqrt{n}}{\epsilon^2})$ sample complexity lower bound by reducing from uniformity testing. You may assume without proof that uniformity testing has an $\Omega(\frac{\sqrt{n}}{\epsilon^2})$ lower bound.

(Hint: a distribution is uniform if and only if it is monotone and the distribution with its domain reversed is also monotone.)