Homework 1

Due: 23 January, 2026

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. Extra credit problems are clearly marked below (see course missive for the details of grade calculations).

Remember to prove all your (non-elementary and not shown in class) mathematical claims, unless stated otherwise.

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

Problem 1

(3 $\checkmark$s)

The first motivating example discussed in class was testing whether a list of length n is all 0s, in the usual testing sense without any ϵ gap. In class, we claimed that any algorithm requires $\Omega(n)$ queries to succeed (with probability at least $\frac{2}{3}$, say). For this problem, show that a *deterministic* algorithm requires exactly n queries. That is, any deterministic algorithm that makes only $n-1$ queries must fail on some input list.

(Hint: What kind of list is hardest to distinguish from the all 0s list? Now suppose there is a tester algorithm $\mathcal{A}$ that makes at most $n-1$ queries, can you *construct* a list that $\mathcal{A}$ cannot distinguish from the all 0s list?)

Problem 2

We now try to generalise the $\Omega(n)$ lower bound from the previous problem to *randomised* algorithms, succeeding with probability $\geq \frac{2}{3}$. More formally, we aim to show that no randomised algorithm which always queries fewer than $\frac{2n}{3}$ locations in the list can succeed with probability at least $\frac{2}{3}$ with a *1-sided error*, always succeeding for the all 0s list. (For a 2-sided error algorithm, the lower bound is still $\Omega(n)$ but it's slightly more annoying to show.)

(a) (2 $\checkmark$) Show that, if there exists an algorithm that succeeds with probability $\frac{2}{3}$ after (always) making at most k (distinct) queries, then there exists an(other) algorithm that makes k uniformly and randomly drawn (distinct, without replacement) queries which also succeeds with $\frac{2}{3}$ probability.

(b) (2 $\checkmark$s, **Extra credit**) Using the previous part, show the desired lower bound on the randomised query complexity on the all 0s problem with 1-sided error.

Problem 3

Suppose $\mathbf{p}$ and $\mathbf{q}$ are two discrete distributions on the set $[n]$ $(= \{1..n\})$. The *total variation distance* $d_{\mathrm{TV}}(\mathbf{p}, \mathbf{q})$, or ($\frac{1}{2}$ times) the ℓ_1 *distance* between $\mathbf{p}$ and $\mathbf{q}$ is defined as

$$\frac{1}{2}||\mathbf{p} - \mathbf{q}||_1 = \frac{1}{2}\sum_{i\in[n]} |p_i - q_i|$$

(*a*) (1 ✓) Show that $d_{\mathrm{TV}}(\mathbf{p}, \mathbf{q})$ can be equivalently defined as

$$\sup_{A\subseteq[n]} \mathbf{p}(A) - \mathbf{q}(A) = \sup_{A\subseteq[n]} \sum_{i\in A} p_i - q_i$$

(*b*) (2 ✓s) Consider the following experiment: an adversary picks (adversarially, not randomly) either the distribution $\mathbf{p}$ or the distribution $\mathbf{q}$, and sends over one sample from the chosen distribution. Show that *no* (deterministic) algorithm or test can, on input this single sample, tell whether the sample came from $\mathbf{p}$ or $\mathbf{q}$ with probability better than $\frac{1}{2} + \frac{1}{2}d_{\mathrm{TV}}(\mathbf{p}, \mathbf{q})$. More formally, the claim is that no algorithm can be such that: 1) if the chosen distribution is $\mathbf{p}$, then the algorithm answers $\mathbf{p}$ with probability $> \frac{1}{2} + \frac{1}{2}d_{\mathrm{TV}}(\mathbf{p}, \mathbf{q})$, and 2) similarly for $\mathbf{q}$.

(*c*) (1 ✓, **Extra credit**, but slightly annoying or very annoying depending on how you prove it) Show the above impossibility result even for randomized algorithms/tests, namely, algorithms that, after seeing a sample, might flip a (sample-dependent biased) coin and answers $\mathbf{p}$ or $\mathbf{q}$ according to the coin flip.

(*d*) (1 ✓) Consider the above experiment again. Suppose that you, as the algorithm designer, have access to the complete descriptions of $\mathbf{p}$ and $\mathbf{q}$. Construct an algorithm or test that successfully distinguishes between a single sample from $\mathbf{p}$ versus $\mathbf{q}$, such that:

- $\mathbb{P}(\text{Answer} = \mathbf{p} \,|\, \mathbf{p}) - \mathbb{P}(\text{Answer} = \mathbf{p} \,|\, \mathbf{q}) \geq d_{\mathrm{TV}}(\mathbf{p}, \mathbf{q})$
- $\mathbb{P}(\text{Answer} = \mathbf{q} \,|\, \mathbf{q}) - \mathbb{P}(\text{Answer} = \mathbf{q} \,|\, \mathbf{p}) \geq d_{\mathrm{TV}}(\mathbf{p}, \mathbf{q})$

The results of this problem can be generalised to continuous distributions.

Problem 4

(3 ✓s)

Consider a sequence of k many *independent* Binomial random variables $X_i \leftarrow \mathrm{Bin}(n, \frac{1}{3})$. Use Chernoff bounds to prove a high-probability upper bound on the maximum of these k variables, assuming that $k \ll n$. Concretely, find "?" such that

$$\mathbb{P}\left(\max\{X_i\} > \frac{n}{3} + O(?)\right) \leq \frac{1}{k}$$

The "?" you find should depend sublinearly in n (so don't give some trivial upper bound like $2n/3$) and polylogarithmically in k. You should also write out the multiplicative constant in the big-O explicitly, and make sure that it is reasonably small (and not 10000).

Problem 5

(3 ✓s, **Extra credit**)

Let $X \leftarrow \mathrm{Poi}(\lambda)$ be a Poisson random variable for some mean λ. Prove the following tail bounds for X, where the function $h(u) \doteq 2\frac{(1+u)\ln(1+u)-u}{u^2}$: For any $x > 0$

$$\mathbb{P}(X > \lambda + x) \le e^{-\frac{x^2}{2\lambda}h(\frac{x}{\lambda})}$$

and for any $x \in (0, \lambda)$,

$$\mathbb{P}(X < \lambda - x) \le e^{-\frac{x^2}{2\lambda}h(-\frac{x}{\lambda})}$$

You may use without proof that the moment generating function of X is $M_X(t) = e^{\lambda(e^t-1)}$.