Problem Set 1

- Due Date: **23 Sep, 2021**
- The points for each problem is indicated on the side. The total for this set is **100** points.
- The problem set has a fair number of questions so please do not wait until close to the deadline to start on them. Try and do one question every couple of days.
- Turn in your problem sets electronically (PDF; either LaTeXed or scanned etc.) on Acadly.
- Collaboration is encouraged, but all writeups must be done individually and must include names of all collaborators.
- Referring to sources other than the text book and class notes is strongly discouraged. But if you do use an external source (eg., other text books, lecture notes, or any material available online), ACKNOWLEDGE all your sources (including collaborators) in your writeup. This will not affect your grades. However, not acknowledging will be treated as a serious case of academic dishonesty.
- Be clear in your writing.

1. **[Derandomising Turán's theorem]** (3 + 7)

   Let $G = (V, E)$ be an undirected graph. For a vertex $v \in V$, let $d(v)$ denote the degree of the vertex v in G. Let $d_{\text{avg}} = 2|E|/|V|$ denote the average degree.

   (a) Show that any such graph G has an independent set (a subset of vertices such that no two of them are connected) of size at least

   $$\sum_{v \in V} \frac{1}{d(v)+1} \geq \frac{|V|}{d_{\text{avg}+1}}$$

   [Hint: Consider the set of vertices in a random order and pick an independent set greedily. What size do you get on expectation? AM-HM should be helpful for the inequality.]

   (b) Come up with a deterministic polynomial time algorithm to compute an independent set of size of the above size.

2. **[Some candidate constructions of pairwise independent hash families]** (10)

   Which of the following family of functions of the form $\{h : \{0,1\}^n \to \{0,1\}^n\}$ constitute a pairwise independent hash family? Support your answer with a proof of pairwise independence (if yes), or provide a counter-example (if no).

   (a) $\mathcal{H} = \left\{h_A(x) = Ax \ : \ A \in \mathbb{F}_2^{n \times n}\right\}$. That is, each hash function is specified by a matrix A and the hash function is just matrix-vector multiplication (over $\mathbb{F}_2$).

   A random function from the family is chosen by picking the matrix A uniformly at random.

Base version: (2021-09-19 13:16:16 +0530) , b4f1bb5

(b) $\mathcal{H} = \left\{h_{A,b}(x) = Ax + b \ : \ A \in \mathbb{F}_2^{n \times n} \ , \ b \in \mathbb{F}_2^n\right\}$. That is, each hash function is given by multiplication by a matrix A followed by adding b (again, over $\mathbb{F}_2$).

A random function from the family is chosen by picking the matrix A and vector b uniformly at random.

3. **[Lower bounds for pairwise independent hash families]** (5 + 7 + 8)

Let $\mathcal{H} = \{h : [N] \to [M]\}$ be a pairwise independent hash family.

(a) If $N \geq 2$, show that $|\mathcal{H}| \geq M^2$.

(b) If $M = 2$, show that $|\mathcal{H}| \geq N + 1$.

[Hint: Based on $\mathcal{H}$, try to construct some orthogonal vectors in $\{\pm 1\}^{|\mathcal{H}|}$.]

(c) More generally, prove that for arbitrary M, we have $|\mathcal{H}| \geq N \cdot (M-1) + 1$.

[Hint: For each $x \in [N]$, construct $M-1$ linearly independent vectors $v_{x,y} \in \mathbb{R}^{|\mathcal{H}|}$ such that $v_{x,y} \perp v_{x',y'}$ if $x \neq x'$.]

4. **[Lower bound for k-wise independent families]** (10)

For this problem, we will only consider families of the form $\mathcal{H} = \{h : [n] \to \{0,1\}\}$. Each such $h : [n] \to \{0,1\}$ can be thought of as just a string in $\{0,1\}^n$ and hence $\mathcal{H}$ is just some (multi-)set of strings in $\{0,1\}^n$.

Rephrasing the definition of k-wise independent in this setting, we have that for any distinct $i_1, \ldots, i_k \in [n]$ and (not necessarily distinct) $a_1, \ldots, a_k \in \{0,1\}$,

$$\Pr_{x \in \mathcal{H}}[x_{i_1} = a_1 \ , \ \cdots \ , \ x_{i_k} = a_k] = \frac{1}{2^k}.$$

For any $T \subseteq [n]$, define $\chi_T : \{0,1\}^n \to \mathbb{R}$ as $\chi_T(x) = (-1)^{\sum_{i \in T} x_i}$.

(a) Suppose $\mathcal{H}$ was a k-wise independent (multi-)set. Consider the following collection of vectors in $\mathbb{R}^{|\mathcal{H}|}$:

$$\{(\chi_T(x) \ : \ x \in \mathcal{H})\}_{T \subseteq [n] \ , \ |T| \leq (k/2)}$$

That is, there is a vector for each $T \subseteq [n]$ of size at most $k/2$, and each such vector consists of the evaluation of χ_T on the points in $\mathcal{H}$.

Show that the above set of vectors are linearly independent over $\mathbb{R}$.

(b) Conclude that $|\mathcal{H}| \geq \sum_{i=0}^{k/2} \binom{n}{i}$.

5. **[Better tail bounds with higher independence]** (12 + 3)

Suppose $X_1, \ldots, X_t$ are random variables taking values in $[0,1]$ and let $X = X_1 + \cdots + X_t$. Let $\mu_i = \mathbb{E}[X_i]$, and $\mu = \sum \mu_i = \mathbb{E}[X]$. Suppose that these random variables are 4-wise independent, i.e. for any set of 4-distinct indices i_1, i_2, i_3, i_4 and any *events* $A_1, A_2, A_3, A_4 \subseteq [0,1]$, we have

$$\Pr\left[X_{i_1} \in A_1 \ , \ \ldots \ , \ X_{i_4} \in A_4\right] = \prod_{j=1}^{4} \Pr[X_{i_j} \in A_j].$$

(a) Prove that $\mathbb{E}[(X-\mu)^4] \leq O(t+t^2)$

[Hint: Rewrite $(X-\mu)^4 = (Y_1 + \cdots + Y_t)^4$ where $Y_i = X_i - \mu_i$. What happens to terms that have Y_i with a single power (i.e. not terms of the form $Y_1^2 Y_2^2$, but terms such as $Y_1 Y_2^3$)?]

(b) Conclude that $\Pr[|X-\mu| \geq t\varepsilon] \leq O\left(\frac{1}{t^2\varepsilon^4}\right)$ in the 4-wise independent case.

(c) **[extra credit]** Can you generalise this to k-wise independence (for even k)? That is, show that if $X_1, \ldots, X_t$ are k-wise independent and $X = \sum X_i$, then

$$\Pr[|X-\mu| > t\varepsilon] \leq O\left(\frac{k^k}{t^{k/2}\varepsilon^k}\right)$$

[Hint: Once again, expand out $\mathbb{E}[(Y_1 + \cdots + Y_t)^k]$ as earlier and argue that the only terms that matter are those where each Y_i in that terms appears at least with an exponent of 2. Use this to show $\mathbb{E}[(Y_1 + \cdots + Y_t)^k] \leq O(k^k \cdot t^{k/2})$.]

6. **[Almost pairwise independent distributions]** (10 + 5)

Recall that a family of hash functions $\mathcal{H} = \{h : \{0,1\}^n \to \{0,1\}^m\}$ is a pairwise independent family if for any $x \neq x' \in \{0,1\}^m$ and $y, y' \in \{0,1\}^m$, we have

$$\Pr_{h \in \mathcal{H}}[h(x) = y \ , \ h(x') = y'] = \frac{1}{M^2}$$

where $M = 2^m$.

We will call a family of hash functions to be an ε-almost pairwise independent family if the above equation is replaced by the following slightly weaker guarantee: for all $x \neq x' \in \{0,1\}^n$ and $y, y' \in \{0,1\}^m$ we have

$$\Pr_{h \in \mathcal{H}}[h(x) = y \ , \ h(x') = y'] \leq \frac{1}{M^2} + \varepsilon.$$

As you can imagine, many applications of pairwise independence might work more or less the same if we replace them with ε-almost pairwise independence.

Here is a template for a construction of ε-almost pairwise independent hash families.

- Fix a field $\mathbb{F}$ of size 2^k for a suitably chosen k (you need to figure out what k needs to be chosen).
- Fix a standard pairwise independent hash family $\mathcal{H} = \left\{h : \{0,1\}^k \to \{0,1\}^m\right\}$.
- The new family of ε-almost pairwise independent hash family is

$$\widetilde{\mathcal{H}} = \{H_{a,h} : \{0,1\}^n \to \{0,1\}^m \ : \ h \in \mathcal{H} \ , \ a \in \mathbb{F}\}$$

where the function $H_{a,h}$ is defined as follows:

> Think of the n-bit input as $f_0, f_1, \cdots, f_{n-1} \in \{0,1\}$. Return the value $h(\alpha)$ where $\alpha = f_0 + f_1 a + f_2 a^2 + \cdots + f_{n-1}a^{n-1} \in \mathbb{F}$ interpretted as an element of $\{0,1\}^k$.
> That is, think of the input as the coefficients of a polynomial, evaluate that polynomial on a, and then apply the "inner" hash function h on the result.

(a) Choose a suitable value for k (possibly depending on m, n and ε) that guarantees that the family $\widetilde{\mathcal{H}}$ is an ε-almost pairwise independent family. Prove your claim formally.

[Hint: The analysis of the ε-biased space construction we saw in class will be useful. If $f \neq g$ are two "inputs" to the hash function $H_{a,b}$, think about cases where $f(a) = g(a)$ and $f(a) \neq g(a)$. What can you say in each of these cases?]

(b) By instantiating the "inner" hash function with whatever we did in class, how many bits do you need to specify a function from $\widetilde{\mathcal{H}}$?

A nice application of this is that if you are given a graph G on N vertices and M edges, and your goal is to find a cut of size $\left(\frac{1}{2} - o(1)\right) M$ edges, there is a deterministic algorithm with running time $M \cdot \text{polylog}N$ that achieves this (basically, use the above almost pairwise independent family and run over all possible random choices). The standard method of using usual pairwise independence usually requires $M \cdot \text{poly}(N)$ time.

7. **[Lower bounds for ε-biased spaces]** (2 + 4 + 7 + 7)

For this problem, you may assume the following fact:

> Any distribution $\mathcal{D}$ on $\{0,1\}^n$ that is ε-biased, for $\varepsilon < \frac{1}{2^{n/2}}$, must have support at least $\frac{1}{2} \cdot 2^n$.

(a) If X is a random variable taking value in $\{0,1\}$, then show that

$$\Pr[X = 0] \in \frac{1}{2} \pm \varepsilon \quad \iff \quad \left|\mathbb{E}[(-1)^X]\right| \leq 2\varepsilon.$$

(b) Suppose $\mathcal{D}$ on $\{0,1\}^n$ is an ε-biased space for an arbitrary ε. Define the distribution $\mathcal{D}^t$ on $\{0,1\}^n$ given by the following sampling procedure:

- Pick t independent samples $x_1, \ldots, x_t \in \{0,1\}^n$ from $\mathcal{D}$.
- Return $x = x_1 \oplus \cdots \oplus x_t$, the bit-wise XOR (or addition in $\mathbb{F}_2^n$).

Show that if the support of $\mathcal{D}$ is s, then the support of $\mathcal{D}^t$ is at most $\binom{s+t}{t}$.

(It may be useful to use the bound $\binom{s+t}{t} \leq \left(\frac{e(s+t)}{t}\right)^t \leq \left(\frac{2es}{t}\right)^t$ (assuming $t \leq s$).)

(c) Show that if $\mathcal{D}$ is an ε-biased distribution, then $\mathcal{D}^t$ is an ε^t-biased distribution.

[Hint: Try and use the "expectation version" of bias from part (a), and the linearity of the linear function you are considering.]

(d) Using the above parts, and the fact given at the beginning, conclude that any ε-biased distribution $\mathcal{D}$ must have support at least

$$\Omega\left(\frac{n}{\varepsilon^2 \log(1/\varepsilon)}\right).$$
