Problem Set 6 CS265, Winter 2022

Due: Wednesday February 16, 11:30am (before class).

Please follow the homework policies on the course website.

---

1. **(6 pt.)**

   Suppose that there is a population of n fish, belonging to one of two types (**A**nglefish and **B**etta fish). There are also a bunch of fish tanks that these fish could live in. These fish are somewhat picky: each one has a list of at least $1 + \log_2(n)$ fish tanks that it would be happy in (e.g., the Anglefish Annie likes the tank with the mini pirate ship, or the tank with the colorful rocks, but she is not into the tank with the fake ferns). The fishes' lists may be different, and may overlap.

   Unfortunately, these two types of fish don't get along well: when an Anglefish and a Betta fish are put in the same tank as each other, they will fight.

   Use the probabilistic method to show that it is possible to put all of the fish into fishtanks in such a way that each fish is happy with their tank, and none of them fight with each other. That is, define a probability distribution for putting fish into tanks so that with positive probability everyone is happy and there is no fighting.

   [**HINT:** *If you put each fish in a random tank that they like, the resulting scheme does* ***not*** *seem to work...you'll have to come up with a more interesting random distribution.*]

2. **(7 pt.) [On the existence of good error correcting codes]**

   For this question, we'll need the following two definitions:

   - Let $H(p) := p\log_2(\frac{1}{p}) + (1-p)\log_2(\frac{1}{1-p})$ for $0 \le p \le 1$ denote the binary entropy function. (Here, $0\log_2 0 = 0$ by convention).
   - Define the *Hamming distance* between two binary strings $x, y \in \{0,1\}^n$ as the number of coordinates at which they differ $d(x,y) := \sum_{i=1}^{n} 1_{x(i)\neq y(i)}$. The *Hamming ball* of radius δn about $x \in \{0,1\}^n$ is the set $B^n(x, \delta n) = \{y \in \{0,1\}^n \,:\, d(x,y) \le \delta n\}$.

   We'll also need the following fact about the relationship between the number of points in a Hamming ball and the entropy function:

   FACT: *For* $\delta \in (0, \frac{1}{2})$, $|B^n(x, \delta n)| = 2^{nH(\delta) \pm o(n)}$.

   Okay, now with the definitions and facts out of the way, on with the question.

   It is a major open question in the theory of *error correcting codes* whether, for all large enough n, there exists some constant $0 < \delta < 1/2$ and set $S \subseteq \{0,1\}^n$ so that

   - $|S| \ge 2^{(1-H(\delta)+0.001)n}$ and
   - $\forall x, y \in S$, we have $d(x,y) > \delta n$.

   (But you don't need to know anything about error correcting codes to do this problem!)

(a) **(2 pt.)** Your friend has just learned the LLL, and they think they can show the existence of such a set. Their strategy is the following:

i. Let $S \subseteq \{0,1\}^n$ be a random set where each $x \in \{0,1\}^n$ is included independently with probability $2^{-n(H(\delta)-0.001)+1}$ (we choose δ with $H(\delta) > 0.001$ and assume that n is sufficiently large so that the probability is less than 1). Then $\mathbb{E}|S| = 2^{(1-H(\delta)+0.001)n+1}$, and by a Chernoff bound $|S|$ is at least $\mathbb{E}|S|/2 = 2^{(1-H(\delta)+0.001)n}$ with overwhelming probability.

ii. For any $x, y \in \{0,1\}^n$ such that $d(x,y) \leq \delta n$, let A_{xy} denote the bad event that both x and y are contained in S. Clearly, $\Pr[A_{xy}] = 2^{-2n(H(\delta)-0.001)+2}$. Then we can use the LLL to show that no bad events happen.

iii. A_{xy} is only dependent with

- A_{xz} for z such that $d(x,z) \leq \delta n$.
- A_{zy} for z such that $d(z,y) \leq \delta n$.

By the FACT above, there are $2 \cdot 2^{n(H(\delta)\pm o(1))} = 2^{n(H(\delta)\pm o(1))}$ of these dependent events.

iv. Hence when we use the LLL, we can set "d" (the number of dependent events) to be $2^{n(H(\delta)\pm o(1))}$.

v. We can apply the LLL whenever $4pd < 1$ which is equivalent to $4 \cdot 2^{-2n(H(\delta)-0.001)+2} \cdot 2^{n(H(\delta)\pm o(1))} < 1$. This holds for many positive constants δ, such as $\delta = 0.25$.

Hence by the LLL your friend concludes that such a set S exists.

As awesome as it would be if your friend had solved this problem, unfortunately there's a problem with the proof strategy above. **What is the flaw in this reasoning?**

(b) **(5 pt.)** Use the LLL to show that for any constant $\delta \in (0, 1/2)$ and for large enough n, there exists a set S of size $|S| \geq 2^{n(1-H(\delta)-0.001)}$ so that $\forall x, y \in S$, we have $d(x,y) > \delta n$.

[**HINT:** *Choose a random multi-set $S \in \{0,1\}^n$ by choosing $2^{n(1-H(\delta)-0.001)}$ randomly, independently with replacement.*]

[**HINT:** *Try and justify that any set S constructed via the previous hint that has the desired distance properties must also have the desired size property.*]

(c) **(0 pt.)** [**Optional: This part will not be graded.**]

For any constant $\delta \in (0, 1/2)$, let S be a random subspace of $\{0,1\}^n$ of dimension $n(1 - H(\delta) - 0.001)$, where we think of $\{0,1\}^n$ as a vector space with addition mod 2. Note that $|S| = 2^{n(1-H(\delta)-0.001)}$. Show that with probability at least 0.99, the Hamming distance between any two distinct strings in S is greater than δn.

Note that sampling a random subspace can be done computationally efficiently, which is why this is a more interesting result than the one from part (b).

(d) **(0 pt.)** [**Optional: This part will not be graded.**]

Your friend is pretty excited about part (b), since they claim that it means that the constructive LLL gives an efficient Las Vegas algorithm to find a set as in part (b). Your friend heard that it is also a major open problem to find such a set with a Las Vegas algorithm in expected time poly(n), and they are looking forward to winning the Shannon award for this. What is your friend missing? (They are correct that this is a major open problem).

Note: The constructive LLL will only be covered in class on Monday, 2/14.

3. **(10 pt.)** Recall that $G(n,p)$ is the distribution on graphs with n vertices, where each of the $\binom{n}{2}$ possible edges is present independently with probability p.

Suppose the $p = c/n$ for some constant c, and let $G \sim G(n,p)$ be a random graph. Let T be the number of triangles in G.

(a) In this part we will show that $\lim_{n\to\infty} \mathbb{P}(T=0) = e^{-c^3/6}$. In order to do this, we'll use a result called *Janson's inequality.*

Theorem 1. *(Janson's inequality) Let Ω be a finite set, and let $A_i \subseteq \Omega$ for $i = 1, \ldots, t$. Choose a random subset $R \subseteq \Omega$ by including each $\omega \in \Omega$ independently with probability p_ω. For $i = 1, \ldots, t$, let B_i be the event that $A_i \subseteq R$. For $i, j \in \{1, \ldots, t\}$ with $i \neq j$, say that "$i \sim j$" if and only if $A_i \cap A_j \neq \emptyset$. Define*

$$M = \prod_{i=1}^{t} \mathbb{P}[\overline{B_i}] \quad \textit{and} \quad \Delta = \sum_{i \sim j} \mathbb{P}[B_i \wedge B_j],$$

where Δ is over all ordered *pairs $i \sim j$ (so $\frac{1}{2}\Delta$ is the sum over unordered pairs). If $\mathbb{P}[B_i] \leq \epsilon$ for $i = 1, \ldots, t$, then*

$$M \leq \mathbb{P}\left[\bigwedge_{i=1}^{t} \overline{B_i}\right] \leq M \exp\left(\frac{\Delta}{2(1-\epsilon)}\right).$$

Use Janson's inequality to show that $\mathbb{P}[T=0] \to e^{-c^3/6}$ as $n \to \infty$ (and c remains fixed).

Note: What is Janson's inequality saying? If the events B_i were independent, then we could bound $\mathbb{P}[\bigwedge_{i=1}^{t} \overline{B_i}] = \prod_{i \in I} \mathbb{P}[\overline{B_i}] = M$. Unfortunately, in the set-up of Janson's inequality, these events are *not* independent since the A_i might overlap. However, if it's the case the *most* of the pairs A_i, A_j don't overlap (or, if they do overlap, than $\mathbb{P}(B_i \wedge B_j)$ is not too big), we'd still hope for this to approximately hold, and that's exactly what Janson's theorem says.

(b) **(Optional, this will not be graded)** You can also get an upper bound on $\mathbb{P}[T=0]$ using the second moment method. Work this out. How does it compare to the result from part (a)?