Please submit solutions only to the problems, not to exercises. **Please collaborate in groups of 2 (or at most 3). Write your own solutions, no sharing of written content. Put down names of your collaborator(s) on the front page and also in the last problem.** Submissions will be via gradescope, and the link will appear on the course webpage and on Brightspace. Also, changes, corrections, and clarifications will also appear on the Ed discussion board, so please check it regularly.

Exercises

1. **(2-Coloring Hypergraphs via Lovász Local Lemma.)** A hypergraph $\mathcal{H} = (V, E)$ consists of a set of vertices V and a collection of edges E, where each edge is a subset of V. We are interested in a 2-coloring of the vertices, say $\chi : V \to \{\text{Red, Blue}\}$, such that no edge is monochromatic (i.e., every edge contains at least one Red vertex and at least one Blue vertex). If such a coloring exists, the hypergraph is called 2-colorable.

   We consider k-uniform hypergraphs, where every edge $E_i \in E$ has size $|E_i| = k$.

   (a) Consider a randomized coloring where each vertex $v \in V$ independently chooses its color (Red or Blue) with probability 1/2 each. For a fixed edge E_i, calculate the probability that E_i is monochromatic.

   (b) (Warm-up: Standard Probabilistic Method) Let m be the total number of edges in $\mathcal{H}$. Show that if $m < 2^{k-1}$, then $\mathcal{H}$ is 2-colorable.

   (c) The condition in (b) works regardless of how the edges intersect. If the intersections are sparse, we can do better. Let's define the degree of an edge E_i, denoted $\deg(E_i)$, as the number of other edges E_j $(j \neq i)$ such that $E_i \cap E_j \neq \varnothing$. Suppose the maximum degree is bounded by d, i.e., $\deg(E_i) \leq d$ for all i.

   Let A_i be the "bad" event that edge E_i is monochromatic. We want to understand the dependencies between these events. Explain why the event A_i is mutually independent of the set of all events $\{A_j : E_i \cap E_j = \varnothing\}$.

   (d) Use the Symmetric Lovász Local Lemma (LLL) to prove the following classical theorem (Erdős-Lovász, 1975): Any k-uniform hypergraph in which every edge intersects at most d other edges is 2-colorable if $e \cdot (d+1) \leq 2^{k-1}$. (Here $e \approx 2.718$ is the base of the natural logarithm).

   Reminder (Symmetric LLL): Let $A_1, \ldots, A_m$ be events. If each event A_i has probability $\mathbf{Pr}[A_i] \leq p$ and each event is mutually independent of all other events except for at most d events (the dependency degree), and if $e \cdot p \cdot (d+1) \leq 1$, then $\mathbf{Pr}[\bigcap_{i=1}^{m} \overline{A_i}] > 0$.

2. **(Discrepancy of Sparse, Uniform Set Systems.)** Let $U = \{1, \ldots, n\}$ be a universe of elements, and let $\mathcal{S} = \{S_1, \ldots, S_m\}$ be a collection of subsets of U. We seek a coloring $\chi : U \to \{-1, +1\}$ that minimizes the discrepancy:

$$\operatorname{disc}(\chi) = \max_{i \in [m]} \left| \sum_{j \in S_i} \chi(j) \right|.$$

We consider a special case (related to the Beck-Fiala setting) where the set system satisfies the following locality conditions for some $k \geq 2$:

- (Sparsity) Each element $j \in U$ belongs to at most k subsets.
- (Uniformity) Each subset S_i contains at most k elements, i.e., $|S_i| \leq k$.

Use the Symmetric Lovász Local Lemma (LLL) to prove that there exists a coloring χ such that $\mathrm{disc}(\chi) = O(\sqrt{k \log k})$.

Hint: Consider a randomized coloring and use the Chernoff bound for the sum of N independent Rademacher variables: $\mathbf{Pr}(|\sum X_j| \geq t) \leq 2e^{-t^2/(2N)}$.

3. **(Vertex-exposure McDiarmid for $\chi(G(n,p))$.)** Let $G \sim G(n,p)$ be an Erdős-Rényi random graph. We analyze the concentration of the chromatic number $\chi(G)$ using the vertex-exposure approach and McDiarmid's inequality.
   (a) Show that $\chi(G)$ is 1-Lipschitz with respect to vertex modifications. That is, if G and G' differ only in the edges incident to a single vertex v, then $|\chi(G) - \chi(G')| \leq 1$.
   (b) Apply McDiarmid's inequality using an appropriate independent vertex exposure to conclude that
   $$\mathbf{Pr}(|\chi(G) - \mathbb{E}[\chi(G)]| \geq \lambda) \leq 2\exp\left(-\frac{2\lambda^2}{n}\right).$$

4. **(Talagrand for edge-disjoint H-packings).** Fix a constant graph H with $e_H = |E(H)| \geq 1$. Let $G \sim G(n,p)$ and $f(G) = \max\{\text{number of edge-disjoint copies of } H \text{ in } G\}$.
   (a) (1-Lipschitz) Show that toggling a single edge changes f by at most 1.
   (b) (Certificate) Prove f is h-certifiable with $h(s) = e_H s$.
   (c) (Talagrand) Using the inequality from the notes, for all $\lambda > 0$,
   $$\mathbf{Pr}(|f - M_f| \geq \lambda) \leq 4\exp\left(-\frac{\lambda^2}{4e_H(M_f + \lambda)}\right),$$
   where M_f is the median of f. Conclude $f(G) = M_f \pm O(\sqrt{M_f \log n})$ w.h.p.
   (d) (Mean vs median) Using boundedness $f \leq m/e_H$, $m = \binom{n}{2}$, derive $f(G) = E[f(G)] \pm O(\sqrt{E[f(G)] \log n} + 1)$ w.h.p.

Problems

Please write short and clear solutions to each of these problems. Use the language of probability to your advantage. Be clear what the events are, what probabilities and expectations you are reasoning about. **If you use any concentration bounds, please clearly make sure you argue that the conditions are satisfied.**

1. **(Frugal Vertex Coloring.)** Let $G = (V, E)$ be a graph with maximum degree Δ. A vertex coloring $\chi : V \to C$ is called *β-frugal* if for every vertex v, no color appears more than β times in its neighborhood $N(v)$. That is,
   $$|\{u \in N(v) : \chi(u) = c\}| \leq \beta \quad \text{for all } v \in V, c \in C.$$
   A coloring is *proper* if $\chi(u) \neq \chi(v)$ for all edges $\{u, v\} \in E$.

Show that for any constant integer $\beta \geq 1$, there exists a β-frugal coloring of G using $Q = O(\Delta^{1+1/\beta})$ colors.

In fact, a stronger statement is true: there is a coloring that is both proper and β-frugal. But we only require to prove the above weaker statement.

2. **(Concentration for Euclidean MST.)** Let $X_1, \ldots, X_n$ be n points chosen independently and uniformly at random from the unit square $[0,1]^2$. Let $L(X_1, \ldots, X_n)$ denote the total length of the Minimum Spanning Tree (MST) on these points, using Euclidean distances.

   Let $\mu = \mathbb{E}[L(X_1, \ldots, X_n)]$ be the expected length of the MST. Prove that for any $\epsilon > 0$, the probability of deviating from the mean by ϵn is exponentially small in n. Specifically, show that:

$$\mathbf{Pr}(|L - \mu| \geq \epsilon n) \leq 2 \exp\left(-\frac{\epsilon^2 n}{25}\right)$$

   Hint: You may use the following fact without proof.

   Fact: ~~Any~~ There exists a Euclidean MST on points in the 2D plane (using the L_2 norm) has a maximum vertex degree of at most 5.

3. **(A randomized algorithm for k-SAT).** Consider a satisfiable k-CNF Φ on n variables. One try of the algorithm: start at uniform $x_0 \in \{0,1\}^n$; for T steps $t = 0, 1, \ldots, T-1$, if x_t satisfies Φ return x_t, else pick an unsatisfied clause C, choose a uniform random literal $\ell \in C$ and flip its variable to obtain x_{t+1} from x_t. If no solution within T steps, restart. Fix a satisfying assignment $x^\star$ and let $D_t = \|x_t - x^\star\|_1$.

   (a) Show that whenever $D_t > 0$, $\mathbf{Pr}[D_{t+1} = D_t - 1 \mid x_t] \geq 1/k$.

   (b) If $D_0 = d$, prove $\mathbf{Pr}[\text{hit } 0 \text{ within } d \text{ steps}] \geq (1/k)^d$ (via d consecutive decreases). (For this problem, you should assume a unique satisfying assignment.)

   (c) For x_0 uniform, $D_0 \sim \mathrm{Bin}(n, 1/2)$. Show $\mathbf{Pr}[\text{success in one try}] \geq \left(\frac{k+1}{2k}\right)^n$.

   (d) Argue that $T = n$ suffices to capture the event in (b), and conclude the expected time $\tilde{O}\left((\frac{2k}{k+1})^n\right)$; specialize to $k = 3$ as $\tilde{O}\left((\frac{3}{2})^n\right)$.

   Remark (Schöning's bound). If in (b) you instead bound $\mathbf{Pr}[\text{ever hit } 0 \mid D_0 = d] \geq (1/(k-1))^d$ using a biased random-walk/gambler's-ruin argument with step -1 w.p. $1/k$ and $+1$ w.p. $1 - 1/k$, then averaging as in (c) yields per-try success $\left(\frac{k}{2(k-1)}\right)^n$ and expected time $\tilde{O}\left((2 - \frac{2}{k})^n\right)$ (e.g., $\tilde{O}\left((\frac{4}{3})^n\right)$ for 3-SAT).

4. **(The Long(est) Path Home.)** Given a graph $G = (V, E)$, you want to find long simple paths in the graph in polynomial time.

   (a) (Algorithm 1: Dead easy.) Show that you can find a path of length k (if such a path exists) in time $n\Delta^k$, where Δ is the maximum degree of G.

   (b) (Easy.) If the graph were directed and acyclic (i.e., a DAG), then show that you can deterministically find the longest path in G in time $O(m + n)$. Here, and in general, $m = |E|$ and $n = |V|$.

(c) (Algorithm 2:) Consider running the following procedure n times, and outputing the longest path found in these n tries.

> Take a random permutation of the vertices, and direct each edge from the lower endpoint to the higher endpoint to create a DAG $\vec{G}$. Find a longest path in $\vec{G}$.

Show that for $k = c\frac{\log n}{\log \log n}$ for some constant $c > 0$, Algorithm 2 will find a path of length k (if it exists) with probability at least $1/2$.

(d) Now, consider a slight extension of this idea. Suppose you have a graph G, and you color the vertices using k colors (neighbors need not have different color). A path is called *polychromatic* if has $\ell \leq k$ vertices, and all the ℓ vertices have different colors.

i. Show that you can find a polychromatic path of length k in time that is $\text{poly}(n, k)2^k$. (So, this is polynomial time for $k = O(\log n)$).

ii. (Algorithm 3:) Consider running the following procedure n times, and outputing the longest path found in these n tries.

> Take a random coloring of the vertices using k colors, and find the polychromatic path of length at most k in G.

Show that for $k = c \log n$ for some constant $c > 0$, Algorithm 3 will find a path of length k (if it exists) with probability at least $1/2$. (Hint: Use Stirling's approximation.)