PROBABILISTIC METHODS IN COMBINATORICS
MIT 18.226 (FALL 2022)
PROBLEM SET

https://yufeizhao.com/pm/

A. Introduction and linearity of expectations

A1. Verify the following asymptotic calculations used in Ramsey number lower bounds:

(a) For each k, the largest n satisfying $\binom{n}{k}2^{1-\binom{k}{2}} < 1$ has $n = \left(\frac{1}{e\sqrt{2}} + o(1)\right) k2^{k/2}$.

(b) For each k, the maximum value of $n - \binom{n}{k}2^{1-\binom{k}{2}}$ as n ranges over positive integers is $\left(\frac{1}{e} + o(1)\right) k2^{k/2}$.

(c) For each k, the largest n satisfying $e\left(\binom{k}{2}\binom{n}{k-2} + 1\right) 2^{1-\binom{k}{2}} < 1$ satisfies $n = \left(\frac{\sqrt{2}}{e} + o(1)\right) k2^{k/2}$.

A2. Prove that, if there is a real $p \in [0,1]$ such that

$$\binom{n}{k}p^{\binom{k}{2}} + \binom{n}{t}(1-p)^{\binom{t}{2}} < 1$$

then the Ramsey number $R(k,t)$ satisfies $R(k,t) > n$. Using this show that

$$R(4,t) \geq c\left(\frac{t}{\log t}\right)^{3/2}$$

for some constant $c > 0$.

ps1 A3. Let G be a graph with n vertices and m edges. Prove that K_n can be written as a union of $O(n^2(\log n)/m)$ isomorphic copies of G (not necessarily edge-disjoint).

A4. Prove that there is an absolute constant $C > 0$ so that for every $n \times n$ matrix with distinct real entries, one can permute its rows so that no column in the permuted matrix contains an increasing subsequence of length at least $C\sqrt{n}$. (A subsequence does not need to be selected from consecutive terms. For example, $(1,2,3)$ is an increasing subsequence of $(1,5,2,4,3)$.)

A5. *Generalization of Sperner's theorem.* Let $\mathcal{F}$ be a collection of subset of $[n]$ that does not contain $k+1$ elements forming a chain: $A_1 \subsetneq \cdots \subsetneq A_{k+1}$. Prove that $\mathcal{F}$ is no larger than taking the union of the k levels of the Boolean lattice closest to the middle layer.

A6. Let G be a graph on $n \geq 10$ vertices. Suppose that adding any new edge to G would create a new clique on 10 vertices. Prove that G has at least $8n - 36$ edges.

Hint in white:

A7. Let $k \geq 4$ and H a k-uniform hypergraph with at most $4^{k-1}/3^k$ edges. Prove that there is a coloring of the vertices of H by four colors so that in every edge all four colors are represented.

ps1 A8. Given a set $\mathcal{F}$ of subsets of $[n]$ and $A \subseteq [n]$, write $\mathcal{F}|_A := \{S \cap A : S \in \mathcal{F}\}$ (its *projection* onto A). Prove that for every n and k, there exists a set $\mathcal{F}$ of subsets of $[n]$ with $|\mathcal{F}| = O(k2^k \log n)$ such that for every k-element subset A of $[n]$, $\mathcal{F}|_A$ contains all 2^k subsets of A.

ps1 A9. Let $A_1, \ldots, A_m$ be r-element sets and $B_1, \ldots, B_m$ be s-element sets. Suppose $A_i \cap B_i = \emptyset$ for each i, and for each $i \neq j$, either $A_i \cap B_j \neq \emptyset$ or $A_j \cap B_i \neq \emptyset$. Prove that $m \leq (r+s)^{r+s}/(r^r s^s)$.

ps1⋆ A10. Show that in every non-2-colorable n-uniform hypergraph, one can find at least $\frac{n}{2}\binom{2n-1}{n}$ unordered pairs of edges with each pair intersecting in exactly one vertex.

A11. Let A be a measurable subset of the unit sphere in $\mathbb{R}^3$ (centered at the origin) containing no pair of orthogonal points.

ps1 (a) Prove that A occupies at most 1/3 of the sphere in terms of surface area.

ps1⋆ (b) Prove an upper bound smaller than 1/3 (give your best bound).

ps1⋆ A12. Prove that every set of 10 points in the plane can be covered by a union of disjoint unit disks.

A13. Let $\boldsymbol{r} = (r_1, \ldots, r_k)$ be a vector of nonzero integers whose sum is nonzero. Prove that there exists a real $c > 0$ (depending on $\boldsymbol{r}$ only) such that the following holds: for every finite set A of nonzero *reals*, there exists a subset $B \subseteq A$ with $|B| \geq c|A|$ such that there do not exist $b_1, \ldots, b_k \in B$ with $r_1 b_1 + \cdots + r_k b_k = 0$.

ps1 A14. Prove that every set A of n nonzero integers contains two disjoint subsets B_1 and B_2, such that both B_1 and B_2 are sum-free, and $|B_1| + |B_2| > 2n/3$.

ps1 A15. Let G be an n-vertex graph with pn^2 edges, with $n \geq 10$ and $p \geq 10/n$. Prove that G contains a pair of vertex-disjoint and isomorphic subgraphs (not necessarily induced) each with at least cp^2n^2 edges, where $c > 0$ is a constant.

ps1⋆ A16. Prove that for every positive integer r, there exists an integer K such that the following holds. Let S be a set of rk points evenly spaced on a circle. If we partition $S = S_1 \cup \cdots \cup S_r$ so that $|S_i| = k$ for each i, then, provided $k \geq K$, there exist r congruent triangles where the vertices of the i-th triangle lie in S_i, for each $1 \leq i \leq r$.

ps1⋆ A17. Prove that $[n]^d$ cannot be partitioned into fewer than 2^d sets each of the form $A_1 \times \cdots \times A_d$ where $A_i \subsetneq [n]$.

B. Alteration method

B1. Using the alteration method, prove the Ramsey number bound

$$R(4, k) \geq c(k/\log k)^2$$

for some constant $c > 0$.

B2. Prove that every 3-uniform hypergraph with n vertices and $m \geq n$ edges contains an independent set (i.e., a set of vertices containing no edges) of size at least $cn^{3/2}/\sqrt{m}$, where $c > 0$ is a constant.

B3. Prove that every k-uniform hypergraph with n vertices and m edges has a transversal (i.e., a set of vertices intersecting every edge) of size at most $n(\log k)/k + m/k$.

ps2 B4. *Zarankiewicz problem.* Prove that for every positive integers $n \geq k \geq 2$, there exists an $n \times n$ matrix with $\{0, 1\}$ entries, with at least $\frac{1}{2}n^{2-2/(k+1)}$ 1's, such that there is no $k \times k$ submatrix consisting of all 1's.

ps2 B5. Fix k. Prove that there exists a constant $c_k > 1$ so that for every sufficiently large $n > n_0(k)$, there exists a collection $\mathcal{F}$ of at least c_k^n subsets of $[n]$ such that for every k distinct $F_1, \ldots, F_k \in \mathcal{F}$, all 2^k intersections $\bigcap_{i=1}^k G_i$ are nonempty, where each G_i is either F_i or $[n] \setminus F_i$.

B6. *Acute sets in* $\mathbb{R}^n$. Prove that, for some constant $c > 0$, for every n, there exists a family of at least $c(2/\sqrt{3})^n$ subsets of $[n]$ containing no three distinct members A, B, C satisfying $A \cap B \subseteq C \subseteq A \cup B$.

Deduce that there exists a set of at least $c(2/\sqrt{3})^n$ points in $\mathbb{R}^n$ so that all angles determined by three points from the set are acute.

Remark. The current best lower and upper bounds for the maximum size of an "acute set" in $\mathbb{R}^n$ (i.e., spanning only acute angles) are $2^{n-1} + 1$ and $2^n - 1$ respectively.

ps2⋆ B7. *Covering complements of sparse graphs by cliques*

(a) Prove that every graph with n vertices and minimum degree $n - d$ can be written as a union of $O(d^2 \log n)$ cliques.

(b) Prove that every bipartite graph with n vertices on each side of the vertex bipartition and minimum degree $n - d$ can be written as a union of $O(d \log n)$ complete bipartite graphs (assume $d \geq 1$).

ps2⋆ B8. Let $G = (V, E)$ be a graph with n vertices and minimum degree $\delta \geq 2$. Prove that there exists $A \subseteq V$ with $|A| = O(n(\log \delta)/\delta)$ so that every vertex in $V \setminus A$ contains at least one neighbor in A and at least one neighbor not in A.

ps2⋆ B9. Prove that every graph G without isolated vertices has an induced subgraph H on at least $\alpha(G)/2$ vertices such that all vertices of H have odd degree. Here $\alpha(G)$ is the size of the largest independent set in G.

C. Second moment method

ps2 C1. *Threshold for k-APs.* Let $[n]_p$ denote the random subset of $\{1, \ldots, n\}$ where every element is included with probability p independently. For each fixed integer $k \geq 3$, determine the threshold for $[n]_p$ to contain a k-term arithmetic progression.

C2. Show that, for each fixed positive integer k, there is a sequence p_n such that

$$\mathbb{P}(G(n, p_n) \text{ has a connected component with exactly } k \text{ vertices}) \to 1 \quad \text{as } n \to \infty.$$

Hint in white:

ps2 C3. *Poisson limit.* Let X be the number of triangles in $G(n, c/n)$ for some fixed $c > 0$.

(a) For every nonnegative integer k, determine the limit of $\mathbb{E}\binom{X}{k}$ as $n \to \infty$.

(b) Let $Y \sim \text{Binomial}(n, \lambda/n)$ for some fixed $\lambda > 0$. For every nonnegative integer k, determine the limit of $\mathbb{E}\binom{Y}{k}$ as $n \to \infty$, and show that it agrees with the limit in (a) for some $\lambda = \lambda(c)$.

We know that Y converges to the Poisson distribution with mean λ. Also, the Poisson distribution is determined by its moments.

(c) Compute, for fixed every nonnegative integer t, the limit of $\mathbb{P}(X = t)$ as $n \to \infty$.

(In particular, this gives the limit probability that $G(n, c/n)$ contains a triangle, i.e., $\lim_{n\to\infty} \mathbb{P}(X > 0)$. This limit increases from 0 to 1 continuously when c ranges from 0 to $+\infty$, thereby showing that the property of containing a triangle has a coarse threshold.)

ps2 C4. *Central limit theorem for triangle counts.* Find a real (non-random) sequence a_n so that, letting X be the number of triangles and Y be the number of edges in the random graph

$G(n, 1/2)$, one has

$$\operatorname{Var}(X - a_n Y) = o(\operatorname{Var} X).$$

Deduce that X is asymptotically normal, that is, $(X - \mathbb{E}X)/\sqrt{\operatorname{Var} X}$ converges to the normal distribution.

(You can solve for the minimizing a_n directly similar to ordinary least squares linear regression, or first write the edge indicator variables as $X_{ij} = (1 + Y_{ij})/2$ and then expand. The latter approach likely yields a cleaner computation.)

C5. *Isolated vertices.* Let $p_n = (\log n + c_n)/n$.

(a) Show that, as $n \to \infty$,

$$\mathbb{P}(G(n, p_n) \text{ has no isolated vertices}) \to \begin{cases} 0 & \text{if } c_n \to -\infty, \\ 1 & \text{if } c_n \to \infty. \end{cases}$$

(b) Suppose $c_n \to c \in \mathbb{R}$, compute, with proof, the limit of LHS above as $n \to \infty$, by following the approach in C3.

ps2⋆ C6. Is the threshold for the bipartiteness of a random graph coarse or sharp?

(You are not allowed to use any theorems that we did not prove in class/notes.)

ps2 C7. *Triangle packing.* Prove that, with probability approaching 1 as $n \to \infty$, $G(n, n^{-1/2})$ has at least $cn^{3/2}$ edge-disjoint triangles, where $c > 0$ is some constant.

Hint in white:

ps3 C8. *Nearly perfect triangle factor.* Prove that, with probability approaching 1 as $n \to \infty$,

(a) $G(n, n^{-2/3})$ has at least $n/100$ vertex-disjoint triangles.

(b) *Simple nibble.* $G(n, Cn^{-2/3})$ has at least $0.33n$ vertex-disjoint triangles, for some constant C.

Hint in white:

C9. *Permuted correlation.* Recall that the *correlation* of two non-constant random variables X and Y is defined to be $\operatorname{corr}(X, Y) := \operatorname{Cov}[X, Y]/\sqrt{(\operatorname{Var} X)(\operatorname{Var} Y)}$.

Let $f, g \in [n] \to \mathbb{R}$ be two non-constant functions. Prove that there exist permutations π and τ of $[n]$ such that, with Z being a uniform random element of $[n]$,

$$\operatorname{corr}(f(\pi(Z)), g(Z)) - \operatorname{corr}(f(\tau(Z)), g(Z)) \geq \frac{2}{\sqrt{n-1}}.$$

Furthermore, show that equality can be achieved for even n.

Hint in white:

ps3 C10. Let $v_1 = (x_1, y_1), \ldots, v_n = (x_n, y_n) \in \mathbb{Z}^2$ with $|x_i|, |y_i| \leq 2^{n/2}/(100\sqrt{n})$ for all $i \in [n]$. Show that there are two disjoint sets $I, J \subseteq [n]$, not both empty, such that $\sum_{i \in I} v_i = \sum_{j \in J} v_j$.

ps3⋆ C11. Prove that there is an absolute constant $C > 0$ so that the following holds. For every prime p and every $A \subseteq \mathbb{Z}/p\mathbb{Z}$ with $|A| = k$, there exists an integer x so that $\{xa : a \in A\}$ intersects every interval of length at least $Cp/\sqrt{k}$ in $\mathbb{Z}/p\mathbb{Z}$.

ps3⋆ C12. Prove that there is a constant $c > 0$ so that every hyperplane containing the origin in $\mathbb{R}^n$ intersects at least c-fraction of the 2^n closed unit balls centered at $\{-1, 1\}^n$.

D. Chernoff bound

D1. Prove that with probability $1 - o(1)$ as $n \to \infty$, every bipartite subgraph of $G(n, 1/2)$ has at most $n^2/8 + 10n^{3/2}$ edges.

ps3 D2. *Unbalancing lights.* Prove that there is a constant C so that for every positive integer n, one can find an $n \times n$ matrix A with $\{-1, 1\}$ entries, so that for all vectors $x, y \in \{-1, 1\}^n$, $|y^\intercal Ax| \le Cn^{3/2}$.

ps3 D3. Prove that there exists a constant $c > 1$ such that for every n, there are at least c^n points in $\mathbb{R}^n$ so that every triple of points form a triangle whose angles are all less than $61°$.

ps3 D4. *Planted clique.* Give a deterministic polynomial-time algorithm for the following task so that it succeeds over the random input with probability approaching 1 as $n \to \infty$.

Input: some unlabeled n-vertex G created as the union of $G(n, 1/2)$ and a clique on $t = \lfloor 100\sqrt{n \log n} \rfloor$ vertices.

Output: a clique in G of size t.

D5. *Weighing coins.* You are given n coins, each with one of two known weights, but otherwise indistinguishable. You can use a scale that outputs the combined weight of any subset of the coins. You must decide in advance which subsets $S_1, \dots, S_k \subseteq [n]$ of the coins to weigh. We wish to determine the minimum number of weighings needed to identify the weight of every coin. (Below, X and Y represent two possibilities for which coins are of the first weight.)

ps3⋆ (a) Prove that if $k \le 1.99n/\log_2 n$ and n is sufficiently large, then for every $S_1, \dots, S_k \subseteq [n]$, there are two distinct subsets $X, Y \subseteq [n]$ such that $|X \cap S_i| = |Y \cap S_i|$ for all $i \in [k]$.

(There is a neat solution to part (a) using information theory, though here you are explicitly asked to solve it using the Chernoff bound.)

ps3⋆ (b) Show that there is some constant C such that (a) is false if 1.99 is replaced by C. (What is the best C you can get?)

E. Lovász local lemma

ps3 E1. Show that it is possible to color the edges of K_n with at most $3\sqrt{n}$ colors so that there are no monochromatic triangles.

E2. Prove that it is possible to color the vertices of every k-uniform k-regular hypergraph using at most $k/\log k$ colors so that every color appears at most $O(\log k)$ times on each edge.

ps3⋆ E3. *Hitting thin rectangles.* Prove that there is a constant $C > 0$ so that for every sufficiently small $\epsilon > 0$, one can choose exactly one point inside each grid square $[n, n+1) \times [m, m+1) \subset \mathbb{R}^2$, $m, n \in \mathbb{Z}$, so that every rectangle of dimensions ϵ by $(C/\epsilon)\log(1/\epsilon)$ in the plane (not necessarily axis-aligned) contains at least one chosen point.

ps4 E4. *List coloring.* Prove that there is some constant $c > 0$ so that given a graph and a set of k acceptable colors for each vertex such that every color is acceptable for at most ck neighbors of each vertex, there is always a proper coloring where every vertex is assigned one of its acceptable colors.

E5. Prove that, for every $\epsilon > 0$, there exist ℓ_0 and some $(a_1, a_2, \dots) \in \{0, 1\}^{\mathbb{N}}$ such that for every $\ell > \ell_0$ and every $i > 1$, the vectors $(a_i, a_{i+1}, \dots, a_{i+\ell-1})$ and $(a_{i+\ell}, a_{i+\ell+1}, \dots, a_{i+2\ell-1})$ differ in at least $(\frac{1}{2} - \epsilon)\ell$ coordinates.

ps4 E6. *Avoiding periodically colored paths.* Prove that for every Δ, there exists k so that every graph with maximum degree at most Δ has a vertex-coloring using k colors so that there is no path of the form $v_1 v_2 \dots v_{2\ell}$ (for any positive integer ℓ) where v_i has the same color as $v_{i+\ell}$ for each $i \in [\ell]$. (Note that vertices on a path must be distinct.)

ps4 E7. Prove that every graph with maximum degree Δ can be properly edge-colored using $O(\Delta)$ colors so that every cycle contains at least three colors.

(An edge-coloring is *proper* if it never assigns the same color to two edges sharing a vertex.)

ps4⋆ E8. Prove that for every Δ, there exists g so that every bipartite graph with maximum degree Δ and girth at least g can be properly edge-colored using $\Delta + 1$ colors so that every cycle contains at least three colors.

ps4⋆ E9. Prove that for every positive integer r, there exists C_r so that every graph with maximum degree Δ has a *proper* vertex coloring using at most $C_r \Delta^{1+1/r}$ colors so that every vertex has at most r neighbors of each color.

E10. *Vertex-disjoint cycles in digraphs.* (Recall that a directed graph is *k-regular* if all vertices have in-degree and out-degree both equal to k. Also, cycles cannot repeat vertices.)

ps4 (a) Prove that every k-regular directed graph has at least $ck/\log k$ vertex-disjoint directed cycles, where $c > 0$ is some constant.

ps4⋆ (b) Prove that every k-regular directed graph has at least ck vertex-disjoint directed cycles, where $c > 0$ is some constant.

Hint in white:

E11. (a) *Generalization of Cayley's formula.* Using Prüfer codes, prove the identity

$$x_1 x_2 \cdots x_n (x_1 + \cdots + x_n)^{n-2} = \sum_T x_1^{d_T(1)} x_2^{d_T(2)} \cdots x_n^{d_T(n)}$$

where the sum is over all trees T on n vertices labeled by $[n]$ and $d_T(i)$ is the degree of vertex i in T.

(b) Let F be a forest with vertex set $[n]$, with components having $f_1, \dots, f_s$ vertices so that $f_1 + \cdots + f_s = n$. Prove that the number of trees on the vertex set $[n]$ that contain F is exactly $n^{n-2} \prod_{i=1}^{s} (f_i / n^{f_i - 1})$.

(c) *Independence property for uniform spanning tree of K_n.* Show that if H_1 and H_2 are vertex-disjoint subgraphs of K_n, then for a uniformly random spanning tree T of K_n, the events $H_1 \subseteq T$ and $H_2 \subseteq T$ are independent.

ps4⋆ (d) *Packing rainbow spanning trees.* Prove that there is a constant $c > 0$ so that for every edge-coloring of K_n where each color appears at most cn times, there exist at least cn edge-disjoint spanning trees, where each spanning tree has all its edges colored differently.

(In your submission, you may assume previous parts without proof.)

The next two problems use the lopsided local lemma.

ps4 E12. *Packing two copies of a graph.* Prove that there is a constant $c > 0$ so that if H is an n-vertex m-edge graph with maximum degree at most cn^2/m, then one can find two edge-disjoint copies of H in the complete graph K_n.

ps4⋆ E13. *Packing Latin transversals.* Prove that there is a constant $c > 0$ so that every $n \times n$ matrix where no entry appears more than cn times contains cn disjoint Latin transversals.

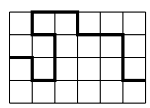
F. Correlation inequalities

F1. Let $G = (V, E)$ be a graph. Color every edge with red or blue independently and uniformly at random. Let E_0 be the set of red edges and E_1 the set of blue edges. Let $G_i = (V, E_i)$ for each $i = 0, 1$. Prove that

$$\mathbb{P}(G_0 \text{ and } G_1 \text{ are both connected}) \leq \mathbb{P}(G_0 \text{ is connected})^2.$$

F2. A set family $\mathcal{F}$ is *intersecting* if $A \cap B \neq \emptyset$ for all $A, B \in \mathcal{F}$. Let $\mathcal{F}_1, \ldots, \mathcal{F}_k$ each be a collection of subsets of $[n]$ and suppose that each $\mathcal{F}_i$ is intersecting. Prove that $\left|\bigcup_{i=1}^k \mathcal{F}_i\right| \leq 2^n - 2^{n-k}$.

F3. *Percolation.* Let $G_{m,n}$ be the grid graph on vertex set $[m] \times [n]$ (m vertices wide and n vertices tall). A *horizontal crossing* is a path that connects some left-most vertex to some right-most vertex. See below for an example of a horizontal crossing in $G_{7,5}$.



Let $H_{m,n}$ denote the random subgraph of $G_{m,n}$ obtained by keeping every edge with probability $1/2$ independently.

Let $\mathsf{RSW}(k)$ denote the following statement: there exists a constant $c_k > 0$ such that for all positive integers n, $\mathbb{P}(H_{kn,n} \text{ has a horizontal crossing}) \geq c_k$.

ps5 (a) Prove $\mathsf{RSW}(1)$.

ps5 (b) Prove that $\mathsf{RSW}(2)$ implies $\mathsf{RSW}(100)$.

(c) (Challenging) Prove $\mathsf{RSW}(2)$.

F4. Let A and B be two *independent* increasing events of independent random variables. Prove that there are two *disjoint* subsets S and T of these random variables so that A depends only on S and B depends only on T.

F5. Let U_1 and U_2 be increasing events and D a decreasing event of independent Boolean random variables. Suppose U_1 and U_2 are independent. Prove that $\mathbb{P}(U_1 | U_2 \cap D) \leq \mathbb{P}(U_1 | U_2)$.

ps5 F6. *Coupon collector.* Let $s_1, \ldots, s_m$ be independent random elements in $[n]$ (not necessarily uniform or identically distributed; chosen with replacement) and $S = \{s_1, \ldots, s_m\}$. Let I and J be disjoint subsets of $[n]$. Prove that $\mathbb{P}(I \cup J \subseteq S) \leq \mathbb{P}(I \subseteq S)\mathbb{P}(J \subseteq S)$.

ps5⋆ F7. Prove that there exist $c < 1$ and $\epsilon > 0$ such that if $A_1, \ldots, A_k$ are increasing events of independent Boolean random variables with $\mathbb{P}(A_i) \leq \epsilon$ for all i, then, letting X denote the number of events A_i that occur, one has $\mathbb{P}(X = 1) \leq c$. (Give your smallest c. It is conjectured that any $c > 1/e$ works.)

ps5⋆ F8. *Disjoint containment.* Let $\mathcal{S}$ and $\mathcal{T}$ each be a collection of subsets of $[n]$. Let $R \subseteq [n]$ be a random subset where each element is included independently (not necessarily with the same probability). Let A be the event that $S \subseteq R$ for some $S \in \mathcal{S}$. Let B be the event that $T \subseteq R$ for some $T \in \mathcal{T}$. Let C denote the event there exist *disjoint* $S, T \subseteq R$ with $S \in \mathcal{S}$ and $T \in \mathcal{T}$. Prove that $\mathbb{P}(C) \leq \mathbb{P}(A)\mathbb{P}(B)$.

G. Janson inequalities

ps5 G1. *3-AP-free probability.* Determine, for all $0 < p \le 0.99$ (p is allowed to depend on n), the probability that $[n]_p$ does not contain a 3-term arithmetic progression, up to a constant factor in the exponent. (The form of the answer should be similar to the conclusion in class about the probability that $G(n,p)$ is triangle-free. See C1 for notation.)

G2. Prove that with probability $1 - o(1)$, the size of the largest subset of vertices of $G(n, 1/2)$ inducing a triangle-free subgraph is $\Theta(\log n)$.

G3. *Nearly perfect triangle factor, again.* Using Janson inequalities this time, give another solution to Problem C8 in the following generality.

ps5 (a) Prove that for every $\epsilon > 0$, there exists $C_\epsilon > 0$ such that such that with probability $1 - o(1)$, $G(n, C_\epsilon n^{-2/3})$ contains at least $(1/3 - \epsilon)n$ vertex-disjoint triangles.

(b) (Optional) Compare the the dependence of the optimal C_ϵ on ϵ you obtain using the method in Problem C8 versus this problem (don't worry about leading constant factors).

ps5⋆ G4. *Threshold for extensions.* Show that for every constant $C > 16/5$, if $n^2p^5 > C \log n$, then with probability $1 - o(1)$, every edge of $G(n, p)$ is contained in a K_4.

Be careful, this event is not increasing, and so it is insufficient to just prove the result for one specific p.

G5. *Lower tails of small subgraph counts.* Fix graph H and $\delta \in (0, 1]$. Let X_H denote the number of copies of H in $G(n, p)$. Prove that for all n and $0 < p < 0.99$,

$$\mathbb{P}(X_H \le (1-\delta)\mathbb{E}X_H) = e^{-\Theta_{H,\delta}(\Phi_H)} \quad \text{where } \Phi_H := \min_{H' \subseteq H : e(H') > 0} n^{v(H')} p^{e(H')}.$$

Here the hidden constants in $\Theta_{H,\delta}$ may depend on H and δ (but not on n and p).

ps5⋆ G6. *List chromatic number of a random graph.* Show that the list chromatic number of $G(n, 1/2)$ is $(1 + o(1))\frac{n}{2\log_2 n}$ with probability $1 - o(1)$.

The *list-chromatic number* (also called *choosability*) of a graph G is defined to the minimum k such that if every vertex of G is assigned a list of k acceptable colors, then there exists a proper coloring of G where every vertex is colored by one of its acceptable colors.

H. Concentration of measure

ps5 H1. *Sub-Gaussian tails.* For each part, prove there is some constant $c > 0$ so that, for all $\lambda > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \ge \lambda\sqrt{\operatorname{Var} X}) \le 2e^{-c\lambda^2}.$$

(a) X is the number of triangles in $G(n, 1/2)$.

(b) X is the number of inversions of a uniform random permutation of $[n]$ (an *inversion* of $\sigma \in S_n$ is a pair (i, j) with $i < j$ and $\sigma(i) > \sigma(j)$).

H2. Prove that for every $\epsilon > 0$ there exists $\delta > 0$ and n_0 such that for all $n \ge n_0$ and $S_1, \dots, S_m \subset [2n]$ with $m \le 2^{\delta n}$ and $|S_i| = n$ for all $i \in [m]$, there exists a function $f\colon [2n] \to [n]$ so that $(1 - e^{-1} - \epsilon)n \le |f(S_i)| \le (1 - e^{-1} + \epsilon)n$ for all $i \in [m]$.

H3. *Simultaneous bisections.* Fix Δ. Let $G_1, \dots, G_m$ with $m = 2^{o(n)}$ be connected graphs of maximum degree at most Δ on the same vertex set V with $|V| = n$. Prove that there exists a partition $V = A \cup B$ so that every G_i has $(1 + o(1))e(G_i)/2$ edges between A and B.

ps5⋆ H4. Prove that there is some constant $c > 0$ so that for every graph G with chromatic number k, letting S be a uniform random subset of V and $G[S]$ the subgraph induced by S, one has, for every $t \geq 0$,

$$\mathbb{P}(\chi(G[S]) \leq k/2 - t) \leq e^{-ct^2/k}.$$

ps5⋆ H5. Prove that there is some constant $c > 0$ so that, with probability $1 - o(1)$, $G(n, 1/2)$ has a bipartite subgraph with at least $n^2/8 + cn^{3/2}$ edges.

H6. Let $k \leq n/2$ be positive integers and G an n-vertex graph with average degree at most n/k. Prove that a uniform random k-element subset of the vertices of G contains an independent set of size at least ck with probability at least $1 - e^{-ck}$, where $c > 0$ is a constant.

ps6⋆ H7. Prove that there exists a constant $c > 0$ so that the following holds. Let G be a d-regular graph and $v_0 \in V(G)$. Let $m \in \mathbb{N}$ and consider a simple random walk $v_0, v_1, \ldots, v_m$ where each v_{i+1} is a uniform random neighbor of v_i. For each $v \in V(G)$, let X_v be the number times that v appears among $v_0, \ldots, v_m$. For that for every $v \in V(G)$ and $\lambda > 0$

$$\mathbb{P}\left(\left|X_v - \frac{1}{d}\sum_{w \in N(v)} X_w\right| \geq \lambda + 1\right) \leq 2e^{-c\lambda^2/m}$$

Here $N(v)$ is the neighborhood of v.

H8. Prove that for every k there exists a $2^{(1+o(1))k/2}$-vertex graph that contains every k-vertex graph as an induced subgraph.

ps6⋆ H9. *Tighter concentration of chromatic number*

(a) Prove that with probability $1 - o(1)$, every vertex subset of $G(n, 1/2)$ with at least $n^{1/3}$ vertices contains an independent set of size at least $c \log n$, where $c > 0$ is some constant.

(b) Prove that there exists some function $f(n)$ and constant C such that for all $n \geq 2$,

$$\mathbb{P}(f(n) \leq \chi(G(n, 1/2)) \leq f(n) + C\sqrt{n}/\log n) \geq 0.99.$$

ps6 H10. Show that for every $\epsilon > 0$ there exists $C > 0$ so that every $S \subseteq [4]^n$ with $|S| \geq \epsilon 4^n$ contains four elements with pairwise Hamming distance at least $n - C\sqrt{n}$ apart.

ps6 H11. *Concentration of measure in the symmetric group.* Let $U \subseteq S_n$ be a set of at least $n!/2$ permutations of $[n]$. Let U_t denote the set of permutations that can be obtained starting from some element of U and then applying at most t transpositions. Prove that

$$|U_t| \geq (1 - e^{-ct^2/n})n!$$

for every $t \geq 0$, where $c > 0$ is some constant.

Hint in white:

For the remaining exercises in this section, use Talagrand's inequality

H12. Let Q be a subset of the unit sphere in $\mathbb{R}^n$. Let $\boldsymbol{x} \in [-1, 1]^n$ be a random vector with independent random coordinates. Let $X = \sup_{\boldsymbol{q} \in Q} \langle \boldsymbol{x}, \boldsymbol{q} \rangle$. Let $t > 0$. Prove that

$$\mathbb{P}(|X - \mathbb{M}X| \geq t) \leq 4e^{-ct^2}$$

where $c > 0$ is some constant.

ps6 H13. *First passage percolation.* Prove that there are constants $c, C > 0$ so that the following holds. Let G be a graph, and let u and w be two distinct vertices with distance at most ℓ between them. Every edge of G is independently assigned some random weight in $[0,1]$ (not necessarily uniform or identically distributed). The weight of a path is defined to be the sum of the weights of its edges. Let X be the minimum weight of a path from u to w using at most ℓ edges. Prove that there is some $m \in \mathbb{R}$ so that

$$\mathbb{P}(|X - m| \geq t) \leq Ce^{-ct^2/\ell}.$$

ps6⋆ H14. *Second largest eigenvalue of a random matrix.* Let A be an $n \times n$ random symmetric matrix whose entries on and above the diagonal are independent and in $[-1,1]$. Show that the second largest eigenvalue $\lambda_2(A)$ satisfies

$$\mathbb{P}(|\lambda_2(A) - \mathbb{E}\lambda_2(A)| \geq t) \leq Ce^{-ct^2},$$

for every $t \geq 0$, where $C, c > 0$ are constants.

Hint in white:

H15. *Longest common subsequence.* Let $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_m)$ be two random sequences with independent entries (not necessarily identically distributed). Let X denote the length of the longest common subsequence, i.e., the largest k such that there exist $i_1 < \cdots < i_k$ and $j_1 < \cdots < j_k$ with $x_{i_1} = y_{j_1}, \ldots, x_{i_k} = y_{j_k}$. Show that, for all $t \geq 0$,

$$\mathbb{P}(X \geq \mathbb{M}X + t) \leq 2\exp\left(\frac{-ct^2}{\mathbb{M}X + t}\right) \quad \text{and} \quad \mathbb{P}(X \leq \mathbb{M}X - t) \leq 2\exp\left(\frac{-ct^2}{\mathbb{M}X}\right)$$

where $c > 0$ is some constant.

I. Entropy method

The problems in this section should be solved using entropy arguments or results derived from entropy arguments.

I1. *Submodularity.* Prove that $H(X,Y,Z) + H(X) \leq H(X,Y) + H(X,Z)$.

I2. Let $\mathcal{F}$ be a collection of subsets of $[n]$. Let p_i denote the fraction of $\mathcal{F}$ that contains i. Prove that

$$|\mathcal{F}| \leq \prod_{i=1}^{n} p_i^{-p_i}(1-p_i)^{-(1-p_i)}.$$

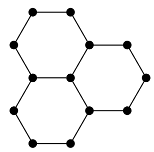
ps6⋆ I3. *Uniquely decodable codes.* Let $[r]^*$ denote the set of all finite strings of elements in $[r]$. Let A be a finite subset of $[r]^*$ and suppose no two distinct concatenations of sequences in A can produce the same string. Let $|a|$ denote the length of $a \in A$. Prove that

$$\sum_{a \in A} r^{-|a|} \leq 1.$$

ps6 I4. *Sudoku.* A $n^2 \times n^2$ *Sudoku square* (the usual Sudoku corresponds to $n = 3$) is an $n^2 \times n^2$ array with entries from $[n^2]$ so that each row, each column, and, after partitioning the square into $n \times n$ blocks, each of these n^2 blocks consist of a permutation of $[n^2]$. Prove that the number of $n^2 \times n^2$ Sudoku squares is at most

$$\left(\frac{n^2}{e^3 + o(1)}\right)^{n^4}.$$

$\boxed{\text{ps6}}$ I5. Prove Sidorenko's conjecture for the following graph.



$\boxed{\text{ps6}\star}$ I6. *Triangles versus vees in a directed graph.* Let V be a finite set, $E \subseteq V \times V$, and

$$\triangle = \left|\left\{(x,y,z) \in V^3 : (x,y),(y,z),(z,x) \in E\right\}\right|$$

(i.e., cyclic triangles; note the direction of edges) and

$$\wedge = \left|\left\{(x,y,z) \in V^3 : (x,y),(x,z) \in E\right\}\right|.$$

Prove that $\triangle \le \wedge$.

$\boxed{\text{ps6}\star}$ I7. *Box theorem.* Prove that for every compact set $A \subseteq \mathbb{R}^d$, there exists an axis-aligned box $B \subseteq \mathbb{R}^d$ with

$$\operatorname{vol} A = \operatorname{vol} B \qquad \text{and} \qquad \operatorname{vol} \pi_I(A) \ge \operatorname{vol} \pi_I(B) \quad \text{for all } I \subseteq [n].$$

Here π_I denotes the orthogonal projection onto the I-coordinate subspace.

(For the purpose of the homework, you only need to establish the case when A is a union of grid cubes. It is optional to give the limiting argument for compact A.)

I8. Let $\mathcal{G}$ be a family of graphs on vertices labeled by $[2n]$ such that the intersection of every pair of graphs in $\mathcal{G}$ contains a perfect matching. Prove that $|\mathcal{G}| \le 2^{\binom{2n}{2}-n}$.

I9. *Loomis–Whitney for sumsets.* Let A, B, C be finite subsets of some abelian group. Writing $A + B := \{a + b : a \in A, b \in B\}$, etc., prove that

$$|A + B + C|^2 \le |A + B|\,|A + C|\,|B + C|.$$

$\boxed{\text{ps6}\star}$ I10. *Shearer for sums.* Let X, Y, Z be independent random integers. Prove that

$$2H(X + Y + Z) \le H(X + Y) + H(X + Z) + H(Y + Z).$$