1. Graph Theory

Let m be a positive integer and let G be a $2m$-colorable graph. Show that G contains a bipartite subgraph H such that $|E(H)| \geq \frac{m}{2m-1}|E(G)|$.

Solution: Let H be a bipartite subgraph of G with maximum number of edges. Note that H may be chosen spanning. Let $V_1, \ldots, V_{2m}$ denote the color classes of a $2m$-coloring of G.

Consider a partition of the color classes into $\mathcal{A} = \{V_{i_1}, \ldots, V_{i_m}\}$ and $\mathcal{B} = \{V_{i_{m+1}}, \ldots, V_{i_{2m}}\}$, where $\{i_1, \ldots, i_{2m}\} = \{1, \ldots, 2m\}$. Let $H(\mathcal{A}, \mathcal{B})$ denote the bipartite subgraph of G with bipartition $(\mathcal{A}, \mathcal{B})$ and edge set $\{uv \in E(G) : u \in V_s \in \mathcal{A}, v \in V_t \in \mathcal{B}\}$. Then each edge of G is counted $2\binom{2m-2}{m-1}$ times in

$$\sum_{(\mathcal{A},\mathcal{B})} |E(H(\mathcal{A}, \mathcal{B}))|.$$

So there exists some $H := H(\mathcal{A}, \mathcal{B})$ with at least

$$\frac{2\binom{2m-2}{m-1}}{\binom{2m}{m}}|E(G)|$$

edges. A simple calculation establishes $|E(H)| \geq \frac{m}{2m-1}|E(G)|$.

2. Probability

With the help of the Strong Law of Large Numbers, find

$$\lim_{n\to+\infty}\int_0^1\cdots\int_0^1 f\left(\frac{x_1+\cdots+x_n}{n}\right)dx_1\cdots dx_n,$$

where f is a continuous and bounded function from $\mathbb{R}$ to $\mathbb{R}$.

Solution: Let $X_1, X_2, \ldots, X_n, \ldots$ be iid uniform r.v. on $[0,1]$, then the SLLN tells us that almost surely

$$\lim_{n\to+\infty}\frac{X_1+\cdots+X_n}{n}=\mathbb{E}X_1=\int_0^1 x\,dx=\frac{1}{2}\,.$$

As $f:\mathbb{R}\to\mathbb{R}$ is continuous, we also have

$$\lim_{n\to+\infty} f\left(\frac{X_1+\cdots+X_n}{n}\right)=f\left(\lim_{n\to+\infty}\left(\frac{X_1+\cdots+X_n}{n}\right)\right)=f\left(\frac{1}{2}\right),$$

with the convergence holding almost surely. Now since f is bounded, i.e., $|f|\le C$, it also follows that

$$\left|f\left(\frac{X_1+\cdots+X_n}{n}\right)\right|\le C.$$

Hence by dominated convergence,

$$\lim_{n\to+\infty}\mathbb{E}f\left(\frac{X_1+\cdots+X_n}{n}\right)=\mathbb{E}f\left(\frac{1}{2}\right)=f\left(\frac{1}{2}\right).$$

But,

$$\begin{aligned}\mathbb{E}f\left(\frac{X_1+\cdots+X_n}{n}\right)&=\int_\Omega f\left(\frac{X_1+\cdots+X_n}{n}\right)d\mathbb{P}\\&=\int_{\mathbb{R}^n} f\left(\frac{x_1+\cdots+x_n}{n}\right)\mathbb{P}_{(X_1,\ldots,X_n)}(dx_1,\ldots,dx_n)\\&=\int_{\mathbb{R}^n} f\left(\frac{x_1+\cdots+x_n}{n}\right)\mathbb{P}_{X_1}(dx_1)\cdots\mathbb{P}_{X_n}(dx_n)\\&=\int_{[0,1]^n} f\left(\frac{x_1+\cdots+x_n}{n}\right)dx_1\cdots dx_n,\end{aligned}$$

where we have used the fact that $X_1,\ldots,X_n$ are independent, indentically distributed and have common uniform law $\mathbb{P}_{X_1}(dx)=\mathbf{1}_{[0,1]}(x)dx$.

3. Analysis of Algorithms

A (d,c,α)-expander is a graph $G = (V,E)$ where each node has degree at most d, and every subset $S \subseteq V$ with at most cn nodes has $|N(S)| \geq \alpha|S|$, where $N(S)$ is the set of vertices in V that have a neighbor in S.

Starting with the set $V = \{1, 2, \ldots, n\}$ of nodes, add a random matching between the vertices as follows: (a) choose a permutation $v_1, v_2, ..., v_n$ of the nodes uniformly at random, and (b) add the edges (i, v_i) for all i. (We may have parallel edges and self-loops; that is fine.) Repeat this process $d = 600$ times; let G denote the resulting graph. Prove that G is a $(2d, 7/20, 3/2)$-expander with probability at least 1/2.

Solution: Fix a subset $S \subseteq V$ of size k. There are $\binom{n}{k}$ ways to pick S. If $T \subseteq V$ is of size αk, there are $\binom{n}{\alpha k}$ ways to pick T. In one iteration, the probablity that the image of S (which is of size k) lies in T is at most $\left(\frac{\alpha k}{n}\right)^k$. This is because if we were to allow each vertex of S to independently choose an image uniformly at random from n possibilities, the probability of the image falling in T is $\frac{\alpha k}{n}$. In our process where the vertices choose an image by fixing a permutation, the probability that all the images fall in T is smaller than if we allowed independent choices. Now, the probability that $N(S) \subseteq T$ is upper bounded by the probability that the image of S lies in T.

Let E_k denote the event that a subset of k vertices has fewer than αk neighbors. Then,

$$\mathbb{P}[E_k] \leq \binom{n}{k}\binom{n}{\alpha k}\left(\frac{\alpha k}{n}\right)^{dk}$$

Since $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$,

$$\mathbb{P}[E_k] \leq \left(\frac{ne}{k}\right)^k \left(\frac{ne}{\alpha k}\right)^{\alpha k} \left(\frac{\alpha k}{n}\right)^{dk} = \left[\left(\frac{k}{n}\right)^{d-\alpha-1} e^{1+\alpha}\alpha^{d-\alpha}\right]^k$$

Since $k \leq cn = 7/20n$,

$$\mathbb{P}[E_k] \leq \left[\left(\frac{7}{20}\right)^{d-\alpha-1} e^{\alpha+1}\alpha^{d-\alpha}\right]^k \leq \left[\left(\frac{7\alpha}{20}\right)^d \left(\frac{20}{7}e\right)^{\alpha+1}\right]^k$$

Since $\alpha = 3/2$ and $d = 600$,

$$\mathbb{P}[E_k] \quad \leq \quad \left[\left(\frac{21}{40}\right)^{600} \left(\frac{20}{7}e\right)^{\frac{5}{2}}\right]^k$$

By union bound, the probability that G is not a (2d,7/20,3/2) expander is at most

$$\sum_{k=1}^{cn-1} \mathbb{P}[E_k] \quad \leq \quad \sum_{k=1}^{cn-1} \left[\left(\frac{21}{40}\right)^{600} \left(\frac{20}{7}e\right)^{\frac{5}{2}}\right]^k \leq \frac{\beta}{1-\beta}$$

For $\beta = \left(\frac{21}{40}\right)^{600} \left(\frac{20}{7}e\right)^{\frac{5}{2}} < 1/3$, we have that $\frac{\beta}{1-\beta} < 1/2$, and hence G is an expander with probability at least 1/2.

4. Linear Programming

A discrete random variable ξ with zero mean takes values from the set $\{i/2005,\ i = 0, \pm 1, \pm 2, \ldots, \pm 2005\}$. What can be the largest possible expectation of $\exp\{2\xi\}$?

Solution. Denoting $a_i = i/2005$, $|i| \leq 2005$, the problem is, what is the optimal value in the linear programming program

$$\max_x \left\{ \sum_i \exp\{2a_i\} x_i : x_i \geq 0, \sum_i a_i x_i = 0, \sum_i x_i = 1 \right\}$$

The dual program is:

$$\min_{\lambda,\mu} \{\mu : \lambda a_i + \mu \geq \exp\{2a_i\},\ |i| \leq 2005\}\,.$$

Observe that the function $\lambda a + \mu - \exp\{2a\}$ of a is concave on $[-1, 1]$, so that its minimum on this segment is achieved either at $a = 1$, or at $a = -1$. Thus, the constraints in the dual are satisfied if and only if they are satisfied when $i = -2005$ ($a_{-2005} = -1$) and when $i = 2005$ ($a_{2005} = 1$). Therefore the dual is equivalent to the problem

$$\min_{\lambda,\mu} \{\mu : \lambda + \mu \geq \exp\{2\}, -\lambda + \mu \geq \exp\{-2\}\}\,;$$

the solution to this problem is $\lambda_* = \frac{1}{2}(\exp\{2\} - \exp\{-2\})$, $\mu_* = \frac{1}{2}(\exp\{2\} + \exp\{-2\})$, and the optimal value is $\cosh\{2\}$.

5. Combinatorial Optimization

In a round-robin competition with n players, where each player has a match with each other player, let w_i denote the number wins for player i (we assume there are no ties in the matches). Describe a network optimization model that gives a polynomial-time algorithm to test if a given vector $w = (w_1, w_2, \ldots, w_n)$ arises from such a competition.

Solution: Consider a bipartite directed graph G with a vertex p_i for each player, a vertex g_{ij} for each game between players $i < j$, and arcs (g_{ij}, p_i) and (g_{ij}, p_j). Give each arc capacity 1, each vertex p_i demand w_i, and each vertex g_{ij} demand -1. The network has an integral circulation if and only if w is a win-loss vector.

6. Algebra

If G is a group of order 140 and H is a subgroup of G of order 35, prove that H is normal in G.

Solution: Let G act by left multiplication on the set S of left cosets of H in G. Since $|S| = 4$, this action defines a homomorphism $\varphi : G \to S_4$. The kernel K of φ is the largest normal subgroup of G contained in H. If $K \neq H$, then $|K| \in \{1, 5, 7\}$, and therefore $|G/K| \in \{140, 28, 20\}$. As the natural map $\overline{\varphi} : G/K \to S_4$ induced by φ is injective, G/K is isomorphic to a subgroup of S_4, and therefore $|G/K|$ divides 24 by Lagrange's theorem. This leaves $K = H$ as the only possibility, and thus H is normal in G.

7. Graph Algorithms

(1) Let $G = (V, E)$ be an undirected graph with positive integral weights on edges, and let $s \in V$ be a special vertex. Let T be the shortest path tree obtained by running Dijkstra's algorithm from s in G. Is T a minimum spanning tree in G? If yes, prove it; otherwise, give a counterexample.
(2) Define a bottleneck spanning tree in an undirected graph to be one that minimizes the maximum weight edge used. Prove that any minimum spanning tree is a bottleneck spanning tree. Is the converse not true? Give a proof or counterexample.

Solution. **(1)** Here is a counterexample. Let G be the complete graph on three vertices, let one of the edges incident with s have weight 3, and let the the other two edges have weight 2.
(2) Suppose that a minimum spanning tree, say T_1, uses an edge e that is heavier than any edge in a bottleneck spanning tree, say T_2. Remove e from T_1 to obtain two components. The tree T_2 must have edges connecting these two components. Any such edge is lighter than e and can be used to obtain a spanning tree lighter than T_1, contradicting the fact that it is a minimal spanning tree.

The converse is not true. Take a clique in which an edge e has weight 1 and the rest of the edges have weight 2. A minimum spanning tree must pick e, but a bottleneck tree need not.

7. Randomized Algorithms

There is a driver who visits n cities $c_1, \ldots, c_n$ in order, and then repeats. Thus, the driver does the following cycle:

$$c_1 \to c_2 \to c_3 \to \cdots \to c_{n-1} \to c_n \to c_1 \to c_2 \cdots \to c_n \to c_1 \to c_2 \ldots$$

The driver has n packages initially. At each visit to a city, he drops a single package with probability p (independently between visits), and with probability $1-p$ does not drop anything. The driver stops once all n packages are dropped.

(a) Suppose $p > 1/2$. Give a high probability upper bound on the maximum number of packages at any city.

(b) Suppose $p < 1/n^6$. Once again, give a high probability upper bound on the maximum number of packages at any city.

To be precise, say high probability means $\geq 1 - n^{-3}$ and state your bound for n sufficiently large. For answers that depend on n (e.g., $\sqrt{n}$), it's OK to get it within a constant factor (so $O(\sqrt{n})$ is sufficient).

Solution

Part (a): The expected total number of cities visited is n/p. By Chernoff's bound, with high probability at most $2n$ cities will be visited. Hence, each city is visited at most twice with high probability, i.e., the max number of packages at any city is at most 2.

Details of the Chernoff bound: Let X_i be an indicator random variable which is 1 if a package is dropped at the i-th city visited. Thus, $E[X_i] = p$. Let $X = \sum_{i=1}^{2n} X_i$. We have $E[X] = 2np$. Using Chernoff's inequality (e.g., Remark 2.5 in *Random Graphs* by [Janson,Luczak,Ruczinski])

$$\begin{aligned} Pr(\text{ number of visited cities} > 2n) &= Pr(X < n) \\ &= Pr(X < E[X] - n(2p-1)) \\ &\leq \exp(-n^2(2p-1)^2). \end{aligned}$$

Since $2p - 1 > 0$, then clearly for n sufficiently large, the tail probability is $\leq n^{-3}$.

Part (b): The probability more than 2 packages are dropped in any cycle is at most n^2p^2. The expected number of cycles the driver makes is $\lceil 1/p \rceil$. By Chernoff's bound, with probability $\geq 1 - n^{-20}$ for n sufficiently large, the driver makes at most $N = O(1/p)$ cycles. Thus, by a union bound, except with probability at most $n^2p^2N = O(n^{-4})$, the driver makes at most N cycles and in no cycle drops more than one package.

The probability that a package is dropped at city i given exactly one package is dropped during the cycle, is the same probability for all i. Hence, the package is dropped at a random city. So n packages are each dropped independently at random cities. This is the standard n balls in n bins, for which the maximum load is $O(\log n / \log \log n)$ with high probability. More precisely,

$$Pr(\text{bin } i \text{ receives} \geq \ell \text{ balls}) = \binom{n}{\ell}(1/n)^\ell \leq (e/\ell)^\ell \leq 1/n^4 \text{ for } \ell = 10 \log n / \log \log n$$