Math 262A/ CSE 291A
Final solutions
Due: 4:00pm, Wednesday, Dec. 6

Please solve four of the following five problems:

1. For any n and k, with $k > 10$, suppose $A_1, \ldots, A_n \subset \{1, \ldots, n\}$, where each A_i has exactly k points and each i, $1 \le i \le n$, lies in exactly k sets. Show that there exists a two-coloring of $\{1, \ldots, n\}$ so that no A_i is monochromatic. (Hint: Use the Lovasz Local Lemma.)

   Proof:
   Color the set $\{1, \ldots, n\}$ red with probability p and blue with probability $1-p$, where each element is colored with independent, identical distribution. Let M_i represent the event that A_i is monochromatic, for $i \in \{1, \ldots, n\}$. We wish to show that there is a coloring for which no A_i is monochromatic, i.e., $\Pr\left[\wedge \overline{M}_i\right] > 0$ by showing that the problem satisfies the preconditions of the *Lovász Local Lemma* . The following is a statement of the lemma.

   Lovász Local Lemma (Symmetric case). Let $M_1, \ldots, M_n$ be events with dependency graph G (on vertices $\{1, \ldots, n\}$) such that

   $$\Pr[M_i] < p \quad \text{for all } i, \qquad \deg(i) \le d \quad \text{for all } i$$

   and

   $$4dp < 1.$$

   Then

   $$\Pr\left[\wedge \overline{M}_i\right] > 0.$$

   We now verify the hypothesis of the lemma.

   - $\Pr[M_i] = p^{k-1}$, since we are free to choose the color of the first element of A_i, but the rest of the elements must be the same color as the first.
   - $\deg(i) \le k(k-1) < k^2$, since each of k elements which appear in A_i may appear in at most $k-1$ other sets A_j. Thus the coloring of A_i is dependent upon the coloring of at most $k(k-1)$ other sets A_j.
   - $4k^2p^{k-1} < 1$ since $p = 1/2$ and $k = 10$.

All hypotheses of the lemma hold, and so there is a coloring of the $\{1, \ldots, n\}$ such that no A_i is monochromatic; i.e.,

$$\Pr\left[\wedge \overline{M}i\ \right] > 0.$$

2. Let $X = X_1 + \ldots + X_n$ where the X_i are indicator random variables with $E[X_i] = 1/2$. Assume that for each i, X_i is independent of X_j for all $j \neq i$ with at most δn exceptions. Bound $Var[X]$ from above and use this to give an upper bound for

$$Pr[|X - \frac{n}{2}| > \epsilon n]$$

Proof:
First we note that

$$E[X] = \sum E[X_i] = \frac{n}{2}.$$

In order to bound Var$[X]$, consider the formula

$$\text{Var}[X] \;=\; \sum_i \text{Var}[X_i] + \sum^{*} \text{cov}[X_i, X_j] + \sum^{**} \text{cov}[X_i, X_j]$$

where $\sum^{**}$ is over those i, j with X_i, X_j independent and $\sum^{*}$ is over all other $i \neq j$. When X_i, X_j are independent $Cov[X_i, X_j] = 0$ so $\sum^{**} = 0$. Otherwise,

$$\begin{aligned} Cov[X_i, X_j] &= E[X_i X_j] - E[X_i]E[X_j] \\ &\leq E[X_i] - E[X_i]E[X_j] \\ &\leq \frac{1}{4}. \end{aligned}$$

So, there are at most δn^2 nonzero terms in the sume of $\sum_i$ and $\sum^{*}$. We have

$$Var[X] \leq \frac{\delta n^2}{4}$$

Next, we need the statement of Chebyshev's inequality.

Chebyshev's inequality. Let X have mean m and variance σ^2, and let $\lambda \geq 0$. Then

$$\Pr\left[|X - m| \geq \lambda\sigma\right] \leq \lambda^{-2}.$$

Now let us apply Chebyshev to our problem. Since $\epsilon n = \sqrt{\delta n^2/4} \cdot \frac{2\epsilon}{\sqrt{\delta}}$, we have

$$Pr[|\ X - \frac{n}{2}\ |> \epsilon n] < \left(\frac{2\epsilon}{\sqrt{\delta}}\right)^2 = \frac{\delta}{4\epsilon^2}$$

3. In the random graph $G(n,p)$ with $p = c/n$ for a fixed constant c, let X denote the number of isolated edges (that are edges disconnected from the rest of the graph). Find asymptotic formulas for $E[X]$ and $Var[X]$.

Proof:
We want to show that

$$\begin{aligned} E[X] &\sim (cn/2)e^{-2c} \\ \text{Var}[X] &\sim n\left(\frac{ce^{-2c}}{2} - c^2e^{-4c} + c^3e^{-4c}\right) \end{aligned}$$

Let X_e be the indicator variable that e is an isolated edge. Let $X = \sum_e X_e$. Then we have

$$E[X_e] = \Pr[X_e = 1] = p(1-p)^{2(n-2)},$$

since an isolated edge i requires edge e to be present and all edges connecting from the other $n-2$ vertices to either endpoint of e to be absent. Thus we have

$$\begin{aligned} E[X] &= \sum_e E[X_e] \\ &\leq \frac{n(n-1)}{2}\frac{c}{n}\ e^{-(c/n)2(n-2)} \cong (cn/2)e^{-2c}. \end{aligned}$$

In a very rough sense, we might say that $E[X]$ is $O(n)$.

Now, let us determine a bound for Var$[X]$. From the expression

$$\text{Var}[X] = \sum_i \text{Var}[X_i] + \sum_{|e\cap f|=1} \text{cov}[X_e, X_f] + \sum_{e\cap f=\emptyset} \text{cov}[X_e, X_f]$$

All three parts are significant!

Since $E[X_e] = o(1)$, $Var[X_e] \sim E[X_e]$ and

$$\sum_e Var[X_e] \sim E[X] \sim n\frac{ce^{-2c}}{2}$$

Suppose $|e \cap f| = 1$. Then

$$\sum_{|e\cap f|=1} \text{Cov}[X_e, X_f] = E[X_e, X_f] - E[X_e]E[X_f] = -E[X_e, X_f]$$

since e, f can't both be isolated idges and sharing a vertex. There are $\binom{n}{2}2(n-2) \sim n^3$ such terms so

$$\sum_{|e\cap f|=1} \text{cov}[X_e, X_f] \sim -n^3(p(1-p)^{2(n-2)})^2 \sim -n^3(\frac{c}{n}e^{-2c})^2 = -nc^2e^{-4c}$$

Suppose $e \cap f = \emptyset$. Then

$$\begin{aligned} \mathrm{Cov}[X_e, X_f] &= E[X_e, X_f] - E[X_e]E[X_f] = p^2(1-p)^{4n-12} - (p(1-p)^{2(n-2)})^2 \\ &= p^2(1-p)^{4n-12}(1-(1-p)^4) \end{aligned}$$

Since $p = o(1), (1-p)^4 = 1 - 4p + o(p)$, there are $\binom{n}{2}\binom{n-2}{2} \sim n^4/4$ terms so

$$\sum_{e \cap f = \emptyset} \mathrm{Cov}[X_e, X_f] \sim \frac{n^4}{4} 4(\frac{c}{n})^3 e^{-4c} = nc^3 e^{-4c}$$

Altogether, we have

$$\mathrm{Var}[X] \sim n\left(\frac{ce^{-2c}}{2} - c^2e^{-4c} + c^3e^{-4c}\right)$$

4. Let $R \subseteq \{1, \ldots, n\}$ be a random subset given by

$$Pr[x \in R] = p = 100n^{-2/3}(\ln n)^{1/3}$$

Show that the probability that there do not exist distinct $x, y, z \in R$ with $x + y + z = n$ is $o(n^{-5})$.
(Hint: Use Janson's inequality.)

Proof:
Let B_i denote a triple $\{x, y, z\}$ with $x + y + z = n$ and x, y, z distinct. Let A_i denote the event that $B_i \subset R$. Then $Pr[A_i] = p^3$. Now, we need an upper bound on the possible number of B_i's . Certainly choosing x and y from the first $n/3$ integers, and setting $z = n - x - y$ will accomplish this. So $\#B_i$'s $\geq \binom{n/3+1}{2} \geq n^2/18$. Therefore,

$$\begin{aligned} M &= \prod_{i \in I} Pr[\bar{A}_i] \\ &\leq (1-p^3)^{n^2/18} \\ &\leq e^{-p^3 n^2/18} \\ &= e^{-100^{-2n} n^2 \ln n/18} \\ &= o(n^{-5}) \end{aligned}$$

Note that $Pr[A_i] = p^3 < \epsilon = 1/2$. To calculate Δ, we must see how a B_i and B_j can be dependent. In this case, it means they share precisely one element. There are n choices for the shared element and at most n choices from each B_i and B_j. The third element in the two sets wll then be determined. Since all five elements must exist (not siz, since one is shared),, the probability associated with each case is p^5. So,

$$\begin{aligned}\Delta &= \sum_{i\sim j} Pr[A_i \cap A_j] \\ &\leq n \cdot n^2 \cdot p^5 \\ &= n^3 \cdot 100^{-10n/3}(\ln n)^{5/3} \\ &= o(1)\end{aligned}$$

Therefore, by Janson's inequality, we have

$$\begin{aligned}Pr[\cap_{i\in I}\bar{A}_i] &\leq Me^{\Delta/(2(1-\epsilon))} \\ &= Me^{\Delta} \\ &= o(n^{-5})(1+o(1)) \\ &= o(n^{-5})\end{aligned}$$

5. Prove that for every k, r with $k \geq 2$, there exists a family $\mathcal{F}$ of k-sets so that $|A \cap B| \leq 1$ for all $A, B \in \mathcal{F}$ and, for any r-coloring of the underlying points, some $A \in \mathcal{F}$ is monochromatic.
(Hint: Use the deletion method.)

Proof:
Choose the family of k-sets randomly and independently from an n-set so that the probability that a given set lies in the family is

$$p = \frac{c}{n^{k-1}}$$

where c is a constant to be determined. Let B be the event that a set of $n/2r$ points exists not containing any of the k-sets in the family. Then,

$$\begin{aligned}Pr[B] &\leq \binom{n}{n/2r}(1-p)^{\binom{n/2r}{k}} \\ &\leq 2^n e^{-p(n/2r)^k/k!} \\ &\leq 2^n e^{-cn/[(2r)^k/k!]} \\ &\leq o(1)\end{aligned}$$

where $c > (2r)^k k!/\ln 2$.

Now that we are assured that there is a k-set in every $n/2r$-set, we want to show that there are at least $n/2kr$ k-sets in every n/r-set. This can be seen by looking at a n/r-set and choosing a $n/2r$-set within it. This subset contains a k-set, so the n/r-set contains this k-set, as well. Now, ignoring the points in this k-set, select an $n/2r$-set from the remaining $(n/r - k)$-set. Again this $n/2r$-set contains a k-set. Repeat this process until there are fewer than $n/2r$ points to choose from. By this time, we have found $n/2kr$ k-sets.

Now all we have left to show is that the expected total number of k-sets which overlap in more than one point is constant. If this is true, we will delete only a constant number of k-sets, leaving still many k-sets in every $n/2r$-set. Let X be the number of "bad" k-sets (that overlap in more than one point). Then,

$$\begin{aligned} E[X] &\leq p^2 \binom{n}{k} \binom{k}{2} \binom{n-2}{k-2} \\ &= O(n^{2k-2} p^2) \\ &= O(1) \end{aligned}$$

and we are done.