C8.4 Probabilistic Combinatorics: Sheet #1

Due on January 29, 2026 at 12:00pm

Professor P. Balister

Ray Tsai

Problem 1

(a) Show that for $1 \leq k \leq n$

$$\exp\left(-\frac{k(k-1)}{2(n-k+1)}\right) \leq \prod_{i=0}^{k-1}\left(1-\frac{i}{n}\right) \leq \exp\left(-\frac{k(k-1)}{2n}\right).$$

Proof. Since $e^{-x/(1-x)} \leq 1-x \leq e^{-x}$ for $x < 1$,

$$-\frac{i}{n-i} \leq \ln\left(1-\frac{i}{n}\right) \leq -\frac{i}{n}.$$

It now follows that

$$-\frac{k(k-1)}{2(n-k+1)} \leq \sum_{i=0}^{k-1} -\frac{i}{n-i} \leq \ln \prod_{i=0}^{k-1}\left(1-\frac{i}{n}\right) = \sum_{i=0}^{k-1} \ln\left(1-\frac{i}{n}\right) \leq \sum_{i=0}^{k-1} -\frac{i}{n} = -\frac{k(k-1)}{2n}.$$

$\square$

(b) Deduce, using Stirling's formula, that for $n, k \to \infty$ and $k = o(n^{2/3})$,

$$\binom{n}{k} \sim \frac{1}{\sqrt{2\pi k}}\left(\frac{en}{k}\right)^k e^{-k^2/2n}.$$

Proof. Since $\prod_{i=0}^{k-1}\left(1-\frac{i}{n}\right) = \frac{n!}{k!n^k}$,

$$\frac{n^k}{k!}\exp\left(-\frac{k(k-1)}{2(n-k+1)}\right) \leq \binom{n}{k} = \frac{n^k}{k!}\prod_{i=0}^{k-1}\left(1-\frac{i}{n}\right) \leq \frac{n^k}{k!}\exp\left(-\frac{k(k-1)}{2n}\right).$$

By Stirling's Formula,

$$\frac{n^k}{k!} \sim \frac{1}{\sqrt{2\pi k}}\left(\frac{en}{k}\right)^k.$$

Since $k = o(n^{2/3})$, we have $k(k-1)/2n \sim k^2/2n$ and $k(k-1)/2(n-k+1) = k^2/2(1+o(1))n + o(1) \sim k^2/2n$. The result now follows from combining all of the above. $\square$

Problem 2

Show that for $r \geq 2$, any graph G contains an r-partite subgraph H with $e(H) \geq \frac{r-1}{r}e(G)$. [*Hint: randomly assign each vertex to a part.*]

Proof. Randomly color the vertices of G with r colors uniformly and independently. Let H be the r-partite subgraph with no edges between vertices with the same color. Then

$$\mathbb{E}[e(H)] = \sum_{\{u,v\} \in E(G)} \mathbb{P}(u \text{ and } v \text{ have different colors}) = \frac{r-1}{r}e(G).$$

Thus there is a coloring such that $e(H) \geq \frac{r-1}{r}e(G)$. $\square$

Problem 3

Prove the Paley–Zygmund inequality: for any non-negative random variable X with finite variance,

$$\mathbb{P}(X > 0) \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

[*Hint: Cauchy–Schwarz.*] How does this compare with the bound from Chebyshev?

Proof. Write $X = X \cdot \mathbb{1}_{\{X>0\}}$. Then by the Cauchy–Schwarz inequality,

$$\mathbb{E}[X]^2 \leq \mathbb{E}[X^2] \cdot \mathbb{E}[\mathbb{1}_{\{X>0\}}^2] = \mathbb{E}[X^2] \cdot \mathbb{P}(X > 0).$$

The result now follows from rearranging the above. $\square$

Problem 4

A *dominating set* in a graph $G = (V, E)$ is a set $U \subseteq V$ such that every vertex $v \in V$ is either in U or has a neighbour in U.

Suppose that $|V| = n$ and that G has minimum degree $\delta \geq 2$. Choose a subset X of V by including each vertex independently with probability p. Let Y be the set of all vertices which are not in X and which have no neighbour in X.

Show that $\mathbb{E}[|X \cup Y|] \leq np + ne^{-p(\delta+1)}$. What can you say about the set $X \cup Y$?

By optimizing over p, show that the graph G has a dominating set which contains at most $n\frac{1+\log(\delta+1)}{\delta+1}$ vertices.

Proof. Note that

$$\begin{aligned}
\mathbb{E}[|X \cup Y|] &= \sum_{v \in V} \mathbb{P}(v \in X \cup Y) \\
&\leq \sum_{v \in V} \mathbb{P}(v \in X) + \sum_{v \in V} \mathbb{P}(v \in Y) \\
&= np + \sum_{v \in V} \mathbb{P}(v \notin X) \cdot \mathbb{P}(N(v) \cap X = \emptyset) \\
&\leq np + \sum_{v \in V} (1-p)^{\delta+1} \\
&\leq np + ne^{-p(\delta+1)}.
\end{aligned}$$

Notice that for $v \in V$, if $v \notin X$, then either $v \in Y$ or v has a neighbour in X. Thus, $X \cup Y$ is a dominating set. Set $p = \log(\delta+1)/(\delta+1)$. Then

$$\mathbb{E}[|X \cup Y|] \leq n\left(\frac{\log(\delta+1)}{\delta+1} + \frac{1}{\delta+1}\right) = n \cdot \frac{1+\log(\delta+1)}{\delta+1}.$$

□

Problem 5

For $n, r \in \mathbb{N}$, $1 < r < n$, let $z(r,n)$ be the largest possible number of 0 entries in an $n \times n$ matrix which has no $r \times r$ submatrix whose entries are all 0. (Here a submatrix is obtained by selecting any r rows and any r columns; the rows or columns need not be consecutive.)

Consider a random matrix in which each entry is 0 with probability p and 1 with probability $1-p$, independently. What is the expected number of entries which are 0? What is the expected number of $r \times r$ submatrices which are "all 0"?

Deduce that $z(r,n) \geq pn^2 - p^{r^2} n^{2r}/r!^2$.

Optimize over p to find the best lower bound on $z(r,n)$ that you can, for fixed r and large n.

Proof. Let M be the random matrix. Let X be the number of 0 entries in M. Then $\mathbb{E}[X] = pn^2$. Let R be the number of $r \times r$ submatrices which are "all 0". Then

$$\mathbb{E}[R] = \sum_{A \text{ is a } r \times r \text{ submatrix of } M} \mathbb{P}(A \text{ is all } 0) = \binom{n}{r}^2 p^{r^2} \leq \frac{n^{2r} p^{r^2}}{r!^2}.$$

Obtain matrix M' by flipping an entry of each $r \times r$ submatrix of M which is "all 0". Note that M' has no $r \times r$ submatrix which is "all 0" and the number of 0 entries in M' is $Z = X - R$. Thus,

$$\mathbb{E}[Z] = \mathbb{E}[X] - \mathbb{E}[R] = pn^2 - \frac{n^{2r} p^{r^2}}{r!^2}.$$

But then there is an $n \times n$ matrix M' without any $r \times r$ submatrix which is "all 0" and the number of 0 entries in M' is at least $pn^2 - n^{2r} p^{r^2}/r!^2$. Thus, $z(r,n) \geq pn^2 - n^{2r} p^{r^2}/r!^2$. By calculus, the bound is optimized when $p = c_r n^{-2/(r+1)}$, where $c_r = (r-1)!^{2/(r^2-1)}$. $\square$

Problem 6

Let G be a graph with n vertices, and let d_v denote the degree of vertex v.

(a) Consider a random ordering of $V = V(G)$ (chosen uniformly from all $n!$ possibilities). What is the probability that v precedes all its neighbours in the ordering?

Proof. The probability that v precedes all its neighbours in the ordering is equals the probability that it is the first vertex in the ordering among its neighbours, which is $\frac{1}{d_v+1}$. □

(b) Show that G has an independent set of size at least $\sum_{v \in V} \frac{1}{d_v+1}$.

Proof. Consider a random ordering of V, and let I be the set of all vertices that precede all their neighbours in the ordering. Notice that for $u, v \in I$, u and v are not neighbours, otherwise one would precede the other in the ordering. Thus, I is an independent set. Since

$$\mathbb{E}[|I|] = \sum_{v \in V} \mathbb{P}(v \in I) = \sum_{v \in V} \frac{1}{d_v + 1},$$

there exists an independent set of size at least $\sum_{v \in V} \frac{1}{d_v+1}$. □

(c) Deduce that any graph with n vertices and m edges has an independent set of size at least $\frac{n^2}{2m+n}$.

Proof. Note that the function $f(x) = 1/(x+1)$ is convex on $[0, \infty)$. Thus by Jensen's inequality,

$$\sum_{v \in V} \frac{1}{d_v + 1} \geq \frac{n}{1 + \frac{1}{n} \sum_{v \in V} d_v} = \frac{n^2}{2m + n}.$$

The result now follows from (b). □

Problem 7

Let G be a bipartite graph with n vertices. Suppose each vertex v has a list $S(v)$ of more than $\log_2 n$ colours associated to it. Show that there is a proper colouring of G in which each vertex v receives a colour from its list $S(v)$.

Proof. Let A and B be the two parts of G. Let $C = \bigcup_{v \in V(G)} S(v)$ and let C_A be a random subset of C such that each element of C is included in C_A with probability $1/2$. Then for $v \in V(G)$,

$$\mathbb{P}(S(v) \cap C_A = \emptyset) = \mathbb{P}(S(v) \cap C_B = \emptyset) < 2^{-\log_2 n} = \frac{1}{n}.$$

Let $C_B = C \setminus C_A$. For $v \in V(G)$, assign it the first available color in C_A if $v \in A$ and the first available color in C_B if $v \in B$. Then

$$\mathbb{P}(\text{coloring is improper}) \leq \sum_{v \in A} \mathbb{P}(S(v) \cap C_A = \emptyset) + \sum_{v \in B} \mathbb{P}(S(v) \cap C_B = \emptyset) < n \cdot \frac{1}{n} = 1.$$

Thus there is a proper coloring of G in which each vertex v receives a colour from its list $S(v)$. □

Problem 8

Let $p = p(n) = \frac{\log n + f(n)}{n}$. Show that if $f(n) \to \infty$ then the probability that $G(n,p)$ contains an isolated vertex tends to 0, and that if $f(n) \to -\infty$ then this probability tends to 1. [*Hint. Apply the first and second moment methods to the number of isolated vertices. We may assume (why?) that $f(n)$ is not too large, say $|f(n)| \leq \log n$.*]

(This shows in particular that $p^*(n) = \log n/n$ is a threshold function for $G(n,p)$ to have minimum degree at least 1.)

Proof. Let X be the number of isolated vertices in $G(n,p)$. Then

$$\mathbb{E}[X] = \sum_{v \in V(G)} \mathbb{P}(v \text{ is isolated}) = n(1-p)^{n-1} \leq ne^{-np} = e^{-f(n)}.$$

Thus if $f(n) \to \infty$, then $\mathbb{E}[X] \to 0$ and by Markov's inequality, $\mathbb{P}(X > 0) \to 0$. Now suppose $f(n) \to -\infty$. Then

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{v \in V(G)} \mathbb{E}[\mathbb{1}^2_{d(v)=0}] + \sum_{\{u,v\} \subseteq V(G)} \mathbb{E}[\mathbb{1}_{d(u)=0} \cdot \mathbb{1}_{d(v)=0}] \\
&= \mathbb{E}[X] + \sum_{u \neq v} \mathbb{P}(N(u) \cup N(v) = \emptyset) \\
&= \mathbb{E}[X] + n(n-1)(1-p)^{2n-3} \\
&= \mathbb{E}[X] + n(n-1)(\mathbb{E}[X]^2 \cdot (1-p)^{-1}).
\end{aligned}$$

But then $p \to 0$, so

$$\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2} = \frac{1}{\mathbb{E}[X]} + \left(1 - \frac{1}{n}\right)(1-p)^{-1} \to 1.$$

By Chebyshev's inequality,

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2} - 1 \to 0.$$

□

Problem 9

Let $S_{n,p}$ be a random subset of $\{1, 2, \ldots, n\}$ chosen by including each element independently with probability p.

(a) Show that $p = n^{-2/3}$ is a threshold function for the property "$S_{n,p}$ contains an arithmetic progression of length 3".

Proof. Suppose $n^{2/3}p \to 0$. For $a \in S_{n,p}$, let X be the number of arithmetic progressions of length 3. Then

$$\mathbb{E}[X] = \sum_{a \in [n]} \sum_{d \leq (n-a)/2} p^3 = \frac{p^3}{2} \sum_{a \in [n]} n - a = \frac{p^3 n(n-1)}{4} \to 0.$$

Thus by Markov's inequality, $\mathbb{P}(X > 0) \to 0$. Now suppose $n^{2/3}p \to \infty$. For 3 term arithmetic progression $A \subseteq [n]$, let $\mathbb{1}_A$ be the indicator random variable that $A \subseteq S_{n,p}$. Since the events $A \subseteq S_{n,p}$ and $B \subseteq S_{n,p}$ are positively correlated, $\mathbb{E}[\mathbb{1}_A \mathbb{1}_B] \geq \mathbb{E}[\mathbb{1}_A]\mathbb{E}[\mathbb{1}_B]$. Thus,

$$\mathbb{E}[X^2] = \sum_{A \text{ is 3-AP}} \mathbb{E}[\mathbb{1}_A^2] + \sum_{A \neq B \text{ are 3-APs}} \mathbb{E}[\mathbb{1}_A \mathbb{1}_B] \geq \mathbb{E}[X] + \sum_{A \neq B \text{ are 3-APs}} p^6.$$

Note that there are $\binom{n(n-1)/4}{2} \sim n^4/32$ distinct pairs of 3-APs in $[n]$. Thus,

$$\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2} = \frac{1}{\mathbb{E}[X]} + \frac{\binom{n(n-1)/4}{2} p^6}{(p^3 n(n-1)/4)^2} \to 1.$$

By Chebyshev's inequality,

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2} - 1 \to 0.$$

□

(b) Show that for $k \geq 3$ fixed, $p = n^{-2/k}$ is a threshold function for $S_{n,p}$ to contain an arithmetic progression of length k.

Proof. Suppose $n^{2/k}p \to 0$. For $a \in S_{n,p}$, let X be the number of arithmetic progressions of length k. Each k-AP can be determined by the first element $a \in [n]$ and the common difference $d \leq (n-a)/(k-1)$. Thus, the number of k-APs in $[n]$ is $N = n(n-1)/2(k-1) \sim n^2/2(k-1)$. But then

$$\mathbb{E}[X] = \sum_{A \text{is } k\text{-AP}} p^k = \frac{n(n-1)p^k}{k-1} \to 0.$$

Thus by Markov's inequality, $\mathbb{P}(X > 0) \to 0$. Now suppose $n^{2/k}p \to \infty$. For k-term arithmetic progression $A \subseteq [n]$, let $\mathbb{1}_A$ be the indicator random variable that $A \subseteq S_{n,p}$. Since the events $A \subseteq S_{n,p}$ and $B \subseteq S_{n,p}$ are positively correlated, $\mathbb{E}[\mathbb{1}_A \mathbb{1}_B] \geq \mathbb{E}[\mathbb{1}_A]\mathbb{E}[\mathbb{1}_B]$. Thus,

$$\mathbb{E}[X^2] = \sum_{A \text{ is } k\text{-AP}} \mathbb{E}[\mathbb{1}_A^2] + \sum_{A \neq B \text{ are } k\text{-APs}} \mathbb{E}[\mathbb{1}_A \mathbb{1}_B] \geq \mathbb{E}[X] + \sum_{A \neq B \text{ are } k\text{-APs}} p^{2k}.$$

Note that there are $\binom{N}{2} \sim n^4/4(k-1)^2$ distinct pairs of k-APs in $[n]$. Thus,

$$\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2} \sim \frac{1}{\mathbb{E}[X]} + \frac{n^4 p^{2k}/4(k-1)^2}{(p^k n(n-1)/2(k-1))^2} \to 1.$$

By Chebyshev's inequality,

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2} - 1 \to 0.$$

□