Lecture notes (MIT 18.226, Fall 2020)

Probabilistic Methods in Combinatorics

Yufei Zhao yufeiz@mit.edu

http://yufeizhao.com/pm/

These notes were created primarily for my own lecture preparation. The writing style is far below that of formal writing and publications (in terms of complete sentences, abbreviations, citations, etc.). The notes are not meant to be a replacement of the lectures.

Please let me know if you spot any errors.

Asymptotic notation convention

Each line below has the same meaning for positive functions f and g (as some parameter, usually n, tends to infinity)

- $f \lesssim g$, $f = O(g)$, $g = \Omega(f)$, $f \leq Cg$ (for some constant $C > 0$)
- $f/g \to 0$, $f \ll g$, $f = o(g)$ (and sometimes $g = \omega(f)$)
- $f = \Theta(g)$, $f \asymp g$, $g \lesssim f \lesssim g$
- $f \sim g$, $f = (1 + o(1))g$
- *whp* (= *with high probability*) means with probability $1 - o(1)$

Warning: analytic number theorists use $\ll$ to mean $O(\cdot)$ (Vinogradov notation)

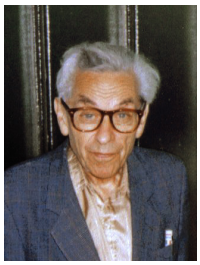


Figure 1: Paul Erdős (1913–1996) is considered the father of the probabilistic method. He published around 1,500 papers during his lifetime, and had more than 500 collaborators. To learn more about Erdős, see his biography *The man who loved only numbers* by Hoffman and the documentary *N is a number*.

1 Introduction

Probabilistic method: to prove that an object exists, show that a random construction works with positive probability

Tackle combinatorics problems by introducing randomness

Theorem 1.0.1. Every graph $G = (V, E)$ contains a bipartite subgraph with at least $|E|/2$ edges.

Proof. Randomly color every vertex of G with black or white, iid uniform

Let $E' =$ edges with one end black and one end white

Then (V, E') is a bipartite subgraph of G

Every edge belongs to E' with probability $\frac{1}{2}$, so by linearity of expectation, $\mathbb{E}[|E'|] = \frac{1}{2}|E|$.

Thus there is some coloring with $|E'| \geq \frac{1}{2}|E|$, giving the desired bipartite subgraph. □

1.1 Lower bounds to Ramsey numbers

Ramsey number $R(k, \ell) =$ smallest n such that in every red-blue edge coloring of K_n, there exists a red K_k or a blue K_ℓ.

e.g., $R(3, 3) = 6$

Ramsey (1929) proved that $R(k, \ell)$ exists and is finite

Figure 2: Frank Ramsey (1903–1930) wrote seminal papers in philosophy, economics, and mathematical logic, before his untimely death at the age of 26 from liver problems. See a recent profile of him in the New Yorker.

1.1.1 Erdős' original proof

The probabilistic method started with:
P. Erdős, Some remarks on the theory of graphs, BAMS, 1947

Remark (Hungarian names). Typing "Erdős" in LaTeX: `Erd\H{o}s` and *not* `Erd\"os`

Hungarian pronunciations: s = /sh/ and sz = /s/, e.g., Erdős, Szekeres, Lovász

Theorem 1.1.1 (Erdős 1947)**.** If $\binom{n}{k}2^{1-\binom{k}{2}} < 1$, then $R(k,k) > n$. In other words, there exist a red-blue edge-coloring of K_n without a monochromatic K_k.

Proof. Color edges uniformly at random

For every fixed subset R of k vertices, let A_R denote the event that R induces a monochromatic K_k. Then $\mathbb{P}(A_R) = 2^{1-\binom{k}{2}}$.

$$\mathbb{P}(\text{there exists a monochromatic } K_k) = \mathbb{P}\left(\bigcup_{R\in\binom{[n]}{k}} A_R\right) \le \sum_{R\in\binom{[n]}{k}} \mathbb{P}(A_R) = \binom{n}{k}2^{1-\binom{k}{2}} < 1.$$

Thus, with positive probability, the random coloring gives no monochromatic K_k. □

Remark. By optimizing n (using Stirling's formula) above, we obtain

$$R(k,k) > \left(\frac{1}{e\sqrt{2}} + o(1)\right) k2^{k/2}$$

Can be alternatively phrased as counting: of all $2^{\binom{n}{2}}$ possible colorings, not all are bad (this was how the argument was phrased in the original Erdős 1947 paper.

In this course, we almost always only consider finite probability spaces. While in principle the finite probability arguments can be rephrased as counting, but some of the later more

involved arguments are impractical without a probabilistic perspective.

Constructive lower bounds? Algorithmic? Open! "Finding hay in a haystack"

Remark (Ramsey number upper bounds). Erdős–Szekeres (1935):

$$R(k+1,\ell+1) \le \binom{k+\ell}{k}.$$

Recent improvements by Conlon (2009), and most recently Sah (2020+):

$$R(k+1,k+1) \le e^{-c(\log k)^2}\binom{2k}{k}.$$

All these bounds have the form $R(k,k) \le (4+o(1))^k$. It is a major open problem whether $R(k,k) \le (4-c)^k$ is true for some constant $c > 0$ and all sufficiently large k.

1.1.2 Alteration method

Two steps: (1) randomly color (2) get rid of bad parts

Theorem 1.1.2. For any k, n, we have $R(k,k) > n - \binom{n}{k}2^{1-\binom{k}{2}}$.

Proof. Construct in two steps:
(1) Randomly 2-color the edges of K_n
(2) Delete a vertex from every monochromatic K_k
Final graph has no monochromatic K_k

After step (1), every fixed K_k is monochromatic with probability $2^{1-\binom{k}{2}}$, let X be the number of monochromatic K_k's. $\mathbb{E}X = \binom{n}{k}2^{1-\binom{k}{2}}$.

We delete at most $|X|$ vertices in step (2). Thus final graph has size $\ge n - |X|$, which has expectation $n - \binom{n}{k}2^{1-\binom{k}{2}}$.

Thus with positive probability, the remaining graph has size at least $n - \binom{n}{k}2^{1-\binom{k}{2}}$ (and no monochromatic K_k by construction) □

Remark. By optimizing the choice of n in the theorem, we obtain

$$R(k,k) > \left(\frac{1}{e} + o(1)\right) k2^{k/2},$$

which improves the previous bound by a constant factor of $\sqrt{2}$.

1.1.3 Lovász local lemma

We give one more improvement to the lower bound, using the Lovász local lemma, which we will prove later in the course

Consider "bad events" $E_1, \ldots, E_n$. We want to avoid all.

If all $\mathbb{P}(E_i)$ small, say $\sum_i \mathbb{P}(E_i) < 1$, then can avoid all bad events.

Or, if they are all independent, then the probability that none of E_i occurs is $\prod_{i=1}^{n}(1 - \mathbb{P}(E_i)) > 0$ (provided that all $\mathbb{P}(E_i) < 1$).

What if there are some weak dependencies?

Theorem 1.1.3 (Lovász local lemma)**.** Let $E_1, \ldots, E_n$ be events, with $\mathbb{P}[E_i] \leq p$ for all i. Suppose that each E_i is independent of all other E_j except for at most d of them. If

$$ep(d+1) < 1,$$

then with some positive probability, none of the events E_i occur.

Remark. The meaning of "independent of ..." is actually somewhat subtle (and easily mistaken). We will come back to this issue later on when we discuss the local lemma in more detail.

Theorem 1.1.4 (Spencer 1977)**.** If $e\left(\binom{k}{2}\binom{n}{k-2} + 1\right) 2^{1-\binom{k}{2}} < 1$, then $R(k,k) > n$.

Proof. Random 2-color edges of K_n

For each k-vertex subset R, let E_R be the event that R induces a monochromatic K_k. $\mathbb{P}[E_R] = 2^{1-\binom{k}{2}}$.

E_R is independent of all E_S other than those such that $|R \cap S| \geq 2$

For each R, there are at most $\binom{k}{2}\binom{n}{k-2}$ choices S with $|S| = k$ and $|R \cap S| \geq 2$.

Apply Lovász local lemma to the events $\left\{E_R : R \in \binom{V}{k}\right\}$ and $p = 2^{1-\binom{k}{2}}$ and $d = \binom{k}{2}\binom{n}{k-2}$, we get that with positive probability none of the events E_R occur, which gives a coloring with no monochromatic K_k's. □

Remark. By optimizing the choice of n, we obtain

$$R(k,k) > \left(\frac{\sqrt{2}}{e} + o(1)\right) k 2^{k/2}$$

once again improving the previous bound by a constant factor of $\sqrt{2}$. This is the best known lower bound to $R(k,k)$ to date.

1.2 Set systems

1.2.1 Sperner's theorem

Let $\mathcal{F}$ a collection of subsets of $\{1,2,\ldots,n\}$. We say that $\mathcal{F}$ is an **antichain** if no set in $\mathcal{F}$ is contained in another set in $\mathcal{F}$.

Question 1.2.1. What is the maximum number of sets in an antichain?

Example: $\mathcal{F} = \binom{[n]}{k}$ has size $\binom{n}{k}$. Maximized when $k = \lfloor \frac{n}{2} \rfloor$ or $\lceil \frac{n}{2} \rceil$. The next result shows that we cannot do better.

Theorem 1.2.2 (Sperner 1928)**.** If $\mathcal{F}$ is an antichain of subsets of $\{1,2,\ldots,n\}$, then $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.

In fact, we will show an even stronger result:

Theorem 1.2.3 (LYM inequality; Bollobás 1965, Lubell 1966, Meshalkin 1963, and Yamamoto 1954)**.** If $\mathcal{F}$ is an antichain of subsets of $[n]$, then

$$\sum_{A \in \mathcal{F}} \frac{1}{\binom{n}{|A|}} \leq 1.$$

Sperner's theorem follows since $\binom{n}{|A|} \geq \binom{n}{\lfloor n/2 \rfloor}$.

Proof. Consider a random permutation σ of $\{1,2,\ldots,n\}$, and its associated chain of subsets

$$\emptyset, \{\sigma(1)\}, \{\sigma(1),\sigma(2)\}, \{\sigma(1),\sigma(2),\sigma(3)\}, \ldots, \{\sigma(1),\ldots,\sigma(n)\}$$

where the last set is always equal to $\{1,2,\ldots,n\}$. For each $A \subset \{1,2,\ldots,n\}$, let E_A denote the event that A is found in this chain. Then

$$\mathbb{P}(E_A) = \frac{|A|!(n-|A|)!}{n!} = \frac{1}{\binom{n}{|A|}}.$$

Since $\mathcal{F}$ is an antichain, if $A, B \in \mathcal{F}$ are distinct, then E_A and E_B cannot both occur. So $\{E_A : A \in \mathcal{F}\}$ is a set of disjoint event, and thus their probabilities sum to at most 1. $\square$

1.2.2 Bollobás two families theorem

Sperner's theorem is generalized by the following celebrated result of Bollobás, which has many more generalizations that we will not discuss here.

Theorem 1.2.4 (Bollobás (1965) "two families theorem"). Let $A_1, \dots, A_m$ be r-element sets and $B_1, \dots, B_m$ be s-element sets such that $A_i \cap B_i = \emptyset$ for all i and $A_i \cap B_j \neq \emptyset$ for all $i \neq j$. Then $m \leq \binom{r+s}{r}$.

Remark. The bound is sharp: let A_i range over all r-element subsets of $[r+s]$ and set $B_i = [r+s] \setminus A_i$.

Let us give an application/motivation for Bollobás' two families theorem in terms of transversals.

Given a set family $\mathcal{F}$, say that T is a **transversal** for $\mathcal{F}$ if $T \cap S \neq \emptyset$ for all $S \in \mathcal{F}$ (i.e., T hits every element of $\mathcal{F}$).

Let $\tau(\mathcal{F})$, the **transversal number** of $\mathcal{F}$, be the size of the smallest transversal of $\mathcal{F}$.

Say that $\mathcal{F}$ is **τ-critical** if $\tau(\mathcal{F} \setminus \{S\}) < \tau(\mathcal{F})$ for all $S \in \mathcal{F}$.

Question 1.2.5. What is the maximum size of a τ-critical r-uniform $\mathcal{F}$ with $\tau(\mathcal{F}) = s+1$?

We claim that the answer is $\binom{r+s}{r}$. Indeed, let $\mathcal{F} = \{A_1, \dots, A_m\}$, and B_i an s-element transversal of $\mathcal{F} \setminus \{A_i\}$ for each i. Then the condition is satisfied. Thus $m \leq \binom{r+s}{r}$.

Conversely, $\mathcal{F} = \binom{[r+s]}{r}$ is τ-critcal r-uniform with $\tau(\mathcal{F}) = s+1$. (why?)

Here is a more general statement of the Bollobás' two-family theorem.

Theorem 1.2.6. Let $A_1, \dots, A_m$ and $B_1, \dots, B_m$ be finite sets such that $A_i \cap B_i = \emptyset$ for all i and $A_i \cap B_j \neq \emptyset$ for all $i \neq j$. Then

$$\sum_{i=1}^{m} \binom{|A_i| + |B_i|}{|A_i|}^{-1} \leq 1.$$

Note that Sperner's theorem and LYM inequality are also special cases, since if $\{A_1, \dots, A_m\}$ is an antichain, then setting $B_i = [n] \setminus A_i$ for all i satisfies the hypothesis.

Proof. Consider a uniform random ordering of all elements.

Let X_i be the event that all elements of A_i come before B_i.

Then $\mathbb{P}[X_i] = \binom{|A_i|+|B_i|}{|A_i|}^{-1}$ (all permutations of $A_i \cup B_i$ are equally likely to occur).

Note that the events X_i are disjoint (X_i and X_j both occuring would contradict the hypothesis for A_i, B_i, A_j, B_j). Thus $\sum_i \mathbb{P}[X_i] \leq 1$. □

1.2.3 Erdős–Ko–Rado theorem on intersecting families

A family $\mathcal{F}$ of sets is **intersecting** if $A \cap B \neq \emptyset$ for all $A, B \in \mathcal{F}$.

Question 1.2.7. What is the largest intersecting family of k-element subsets of $[n]$?

Example: $\mathcal{F}$ = all subsets containing the element 1. Then $\mathcal{F}$ is intersecting and $|\mathcal{F}| = \binom{n-1}{k-1}$

Theorem 1.2.8 (Erdős–Ko–Rado 1961; proved in 1938)**.** If $n \geq 2k$, then every intersecting family of k-element subsets of $[n]$ has size at most $\binom{n-1}{k-1}$.

Remark. The assumption $n \geq 2k$ is necessary since if $n < 2k$, then the family of all k-element subsets of $[n]$ is automatically intersecting by pigeonhole.

Proof. Consider a uniform random circular permutation of $1, 2, \ldots, n$ (arrange them randomly around a circle)

For each k-element subset A of $[n]$, we say that A is **contiguous** if all the elements of A lie in a contiguous block on the circle.

The probability that A forms a contiguous set on the circle is exactly $n/\binom{n}{k}$.

So the expected number of contiguous sets in $\mathcal{F}$ is exactly $n\,|\mathcal{F}|\,/\binom{n}{k}$.

Since $\mathcal{F}$ is intersecting, there are at most k contiguous sets in $\mathcal{F}$ (under every circular ordering of $[n]$). Indeed, suppose that $A \in \mathcal{F}$ is contiguous. Then there are $2(k-1)$ other contingous sets (not necessarily in $\mathcal{F}$) that intersect A, but they can be paired off into disjoint pairs. Since $\mathcal{F}$ is intersecting, it follows that it contains at most k contiguous sets.

Combining with result from the previous paragraph, we see that $n\,|\mathcal{F}|\,/\binom{n}{k} \leq k$, and hence $|\mathcal{F}| \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}$. □

1.3 2-colorable hypergraphs

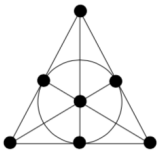
An ***k*-uniform hypergraph** (or ***k*-graph**) is a pair $H = (V, E)$, where V (vertices) is a finite set and E (edges) is a set of k-element subsets of E, i.e., $E \subseteq \binom{V}{k}$ (so hypergraphs are really the same concept as set families).

We say that H is **r-colorable** if the vertices can be colored using r colors so that no edge is monochromatic.

Let $m(k)$ denote the minimum number of edges in a k-uniform hypergraph that is not 2-colorable (elsewhere in the literature, "2-colorable" = "property B", named after Bernstein who introduced the concept in 1908)

$m(2) = 3$

$m(3) = 7$. Example: Fano plane (below) is not 2-colorable (the other direction is by exhaustive search)



$m(4) = 23$, proved via exhaustive computer search (Östergård 2014)

Exact value of $m(k)$ is unknown for all $k \geq 5$

The probabilistic method gives a short proof of a lower bound (random coloring):

Theorem 1.3.1 (Erdős 1964)**.** For any $k \geq 2$, $m(k) \geq 2^{k-1}$, i.e., every k-uniform hypergraph with fewer than 2^{k-1} edges is 2-colorable.

Proof. Let there be $m < 2^{k-1}$ edges. In a random 2-coloring, the probability that there is a monochromatic edge is $\leq 2^{-k+1}m < 1$. □

Remark. Later on we will prove an better lower bound $m(k) \gtrsim 2^k\sqrt{k/\log k}$, which is the best known to date.

Perhaps somewhat surprisingly, the state of the art upper bound is also proved using probabilistic method (random construction).

Theorem 1.3.2 (Erdős 1964)**.** $m(k) = O(k^2 2^k)$, i.e., there exists a k-uniform hypergraph with $O(k^2 2^k)$ edges that is not 2-colorable.

Proof. Fix $|V| = n$ to be decided. Let H be the k-uniform hypergraph obtained by choosing m random edges (with replacement) $S_1, \ldots, S_m$.

Given a coloring $\chi\colon V \to [2]$, let A_χ denote the event that χ is a proper coloring (i.e., no monochromatic edges). It suffices to check that $\sum_\chi \mathbb{P}[A_\chi] < 1$.

If χ colors a vertices with one color and b vertices with the other color, then the probability that (random) S_1 is monochromatic under (fixed) χ is

$$\frac{\binom{a}{k}+\binom{b}{k}}{\binom{n}{k}} \geq \frac{2\binom{n/2}{k}}{\binom{n}{k}} = \frac{2(n/2)(n/2-1)\cdots(n/2-k+1)}{n(n-1)\cdots(n-k+1)}$$
$$\geq 2\left(\frac{n/2-k+1}{n-k+1}\right)^k = 2^{-k+1}\left(1-\frac{k-1}{n-k+1}\right)^k$$

Setting $n=k^2$, we see that the above quantity is at least $c2^{-k}$ for some constant $c>0$.

Thus, the probability that χ is a proper coloring (i.e., no monochromatic edges) is at most $(1-c2^{-k})^m \leq e^{-c2^{-k}m}$ (using $1+x \leq e^x$ for all real x).

Thus, $\sum_\chi \mathbb{P}[A_\chi] \leq 2^n e^{-c2^{-k}m} < 1$ for some $m = O(k^2 2^k)$ (recall $n=k^2$). □

1.4 List chromatic number of $K_{n,n}$

Given a graph G, its **chromatic number** $\chi(G)$ is the minimum number of colors required to proper color its vertices.

In **list coloring**, each vertex of G is assigned a list of allowable colors. We say that G is **k-choosable** (also called **k-list colorable**) if it has a proper coloring no matter how one assigns a list of k colors to each vertex.

We write $\mathrm{ch}(G)$, called the **choosability** (also called: **choice number**, **list colorability**, **list chromatic number**) of G, to be the smallest k so that G is k-choosable.

It should be clear that $\chi(G) \leq \mathrm{ch}(G)$, but the inequality may be strict.

For example, while every bipartite graph is 2-colorable, $K_{3,3}$ is not 2-choosable. Indeed, no list coloring of $K_{3,3}$ is possible with color lists (check!):

$$\begin{matrix}\{2,3\} & \{2,3\}\\ \{1,3\} & \{1,3\}\\ \{1,2\} & \{1,2\}\end{matrix}$$

Easy to check then that $\mathrm{ch}(K_{3,3}) = 3$.

Question 1.4.1. What is the asymptotic behavior of $\mathrm{ch}(K_{n,n})$?

First we prove an upper bound on $\mathrm{ch}(K_{n,n})$.

Theorem 1.4.2. If $n < 2^{k-1}$, then $K_{n,n}$ is k-choosable.

In other words, $\mathrm{ch}(K_{n,n}) \le \lfloor \log_2(2n) \rfloor + 1$.

Proof. For each color, mark it either "L" or "R" iid uniformly.

For any vertex of $K_{n,n}$ on the left part, remove all its colors marked R.

For any vertex of $K_{n,n}$ on the right part, remove all its colors marked L.

The probability that some vertex has no colors remaining is at most $2n2^{-k} < 1$. So with positive probability, every vertex has some color remaining. Assign the colors arbitrarily for a valid coloring. □

The lower bound on $\mathrm{ch}(K_{n,n})$ turns out to follow from the existence of non-2-colorable k-uniform hypergraph with many edges.

Theorem 1.4.3. If there exists a non-2-colorable k-uniform hypergraph with n edges, then $K_{n,n}$ is not k-choosable.

Proof. Let $H = (V, E)$ be a k-uniform hypergraph $|E| = n$ edges. Label the vertex of $K_{n,n}$ by v_e and w_e as e ranges over E. View V as colors and assign to both v_e and w_e a list of colors given by the k-element set e.

If this $K_{n,n}$ has a proper list coloring with the assigned colors. Let C be the colors used among the n vertices. Then we get a proper 2-coloring of H by setting C black and $V \setminus C$ white. So if H is not 2-colorable, then this $K_{n,n}$ is not k-choosable. □

Recall from Theorem 1.3.2 that there exists a non-2-colorable k-uniform hypergraph with $O(k^2 2^k)$ edges. Thus $\mathrm{ch}(K_{n,n}) > (1 - o(1)) \log_2 n$.

Putting these bounds together:

Corollary 1.4.4. $\mathrm{ch}(K_{n,n}) = (1 + o(1)) \log_2 n$

It turns out that, unlike the chromatic number, the list chromatic number always grows with the average degree. The following result was proved using the method of **hypergraph containers** (a very important modern development in combinatorics) provides the optimal asymptotic dependence (the example of $K_{n,n}$ shows optimality).

Theorem 1.4.5 (Saxton and Thomason 2015)**.** If a graph G has average degree d, then $\mathrm{ch}(G) > (1 + o(1)) \log_2 d$.

They also proved similar results for the list chromatic number of hypergraphs. For graphs, a slightly weaker result, off by a factor of 2, was proved earlier by Alon (2000).

2 Linearity of expectations

Let $X = c_1X_1 + \cdots + c_nX_n$ where $X_1, \ldots, X_n$ are random variables, and $c_1, \ldots, c_n$ constants. Then

$$\mathbb{E}[X] = c_1\mathbb{E}[X_1] + \cdots + c_n\mathbb{E}[X_n]$$

Note: this identity does not require any assumption of independence. On the other hand, generally $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$ unless X and Y are uncorrelated (Independent random variables are always uncorrelated)

Here is a simple question with a simple solution (there are also much more involved solutions via enumerations, but linearity of expectations nearly trivializes the problem).

Question 2.0.1. What is the average number of fixed points of a random permutation of $[n]$ chosen uniformly at random?

Let X_i be the event that i is fixed. Then $\mathbb{E}[X_i] = 1/n$. So the expected number of fixed points is $\mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = 1$

2.1 Hamiltonian paths in tournaments

Important observation for proving existence: With positive probability, $X \geq \mathbb{E}[X]$ (likewise for $X \leq \mathbb{E}[X]$)

A **tournament** is a directed complete graph.

Theorem 2.1.1 (Szele 1943)**.** There is a tournament on n vertices with at least $n!2^{-(n-1)}$ Hamiltonian paths

Proof. Let X be the number of Hamiltonian paths in a random tournament.

For every permutation σ of $[n]$, one has the directed path $\sigma(1) \to \sigma(2) \to \cdots \to \sigma(n)$ with probability 2^{-n+1}.

Let X be the number of σ satisfying the above. $\mathbb{E}X = n!2^{-n+1}$. □

This was considered the first use of the probabilistic method. Szele conjectured that the maximum number of Hamiltonian paths in a tournament on n players is $n!/(2 - o(1))^n$. This was proved by Alon (1990) using the Minc–Brégman theorem on permanents (we will see this later in the course when discussing the entropy method).

2.2 Sum-free set

A subset A in an abelian group is **sum-free** if there do not exist $a, b, c \in A$ with $a+b=c$.

Does every n-element set contain a large sum-free set?

Theorem 2.2.1 (Erdős 1965)**.** Every set of n nonzero integers contains a sum-free subset of size $\geq n/3$.

Proof. Let $A \subset \mathbb{Z} \setminus \{0\}$ with $|A| = n$. For $\theta \in [0,1]$, let

$$A_\theta := \{a \in A : \{a\theta\} \in (1/3, 2/3)\}$$

where $\{\cdot\}$ denotes fractional part. Then A_θ is sum-free since $(1/3, 2/3)$ is sum-free in $\mathbb{R}/\mathbb{Z}$.

For θ uniformly chosen at random, $\{a\theta\}$ is also uniformly random in $[0,1]$, so $\mathbb{P}(a \in A_\theta) = 1/3$. By linearity of expectations, $\mathbb{E}|A_\theta| = n/3$. □

Remark. Alon and Kleitman (1990) noted that one can improve the bound to $\geq (n+1)/3$ by noting that $|A_\theta| = 0$ for $\theta \approx 0$.

Bourgain (1997) improved it to $\geq (n+2)/3$ via a difficult Fourier analytic argument. This is currently the best bound known.

Eberhard, Green, and Manners (2014) showed that there exist n-element sets of integers whose largest sum-free subset has size $(1/3 + o(1))n$.

It remains an open problem to prove $\geq (n + \omega(n))/3$ for some function $\omega(n) \to \infty$

2.3 Turán's theorem and independent sets

Question 2.3.1. What is the maximum number of edges in an n-vertex K_k-free graph?

Taking the complement of a graph changes its independent sets to cliques and vice versa. So the problem is equivalent to one about graphs without large independent sets.

The following result, due to Caro (1979) and Wei (1981), shows that a graph with small degrees much contain large independent sets. The probabilistic method proof shown here is due to Alon and Spencer.

Theorem 2.3.2 (Caro 1979, Wei 1981)**.** Every graph G contains an independent set of size at least

$$\sum_{v \in V(G)} \frac{1}{d_v + 1},$$

where d_v is the degree of vertex v.

Proof. Consider a random ordering (permutation) of the vertices. Let I be the set of vertices that appear before all of its neighbors. Then I is an independent set.

For each $v \in V$, $\mathbb{P}(v \in I) = \frac{1}{1+d_v}$ (this is the probability that v appears first among $\{v\} \cup N(v)$). Thus $\mathbb{E}|I| = \sum_{v \in V(G)} \frac{1}{d_v+1}$. Thus with positive probability, $|I|$ is at least this expectation. □

Remark. Equality occurs if G is a disjoint union of cliques.

Remark (Derandomization). Here is an alternative "greedy algorithm" proof of the Caro–Wei inequality.

Permute the vertices in non-increasing order of their degree.

And then greedily construct an independent set: at each step, take the first available vertex (in this order) and then discarding all its neighbors.

If each vertex v is assigned weight $1/(d_v + 1)$, then the total weight removed at each step is at most 1. Thus there must be at least $\sum_v 1/(d_v + 1)$ steps.

Taking the complement

Corollary 2.3.3. Every n-vertex graph G contains a clique of size at least $\sum_{v \in V(G)} \frac{1}{n-d_v}$.

Note that equality is attained when G is multipartite.

Now let us answer the earlier question about maximizing the number of edges in a K_{r+1}-free graph.

The **Turán graph** $T_{n,r}$ is the complete multipartite graph formed by partitioning n vertices into r parts with sizes as equal as possible (differing by at most 1).

Easy to see that $T_{n,r}$ is K_{r+1}-free.

Turán's theorem (1941) tells us that $T_{n,r}$ indeed maximizes the number of edges among n-vertex K_{r+1}-free graphs.

We will prove a slightly weaker statement, below, which is tight when n is divisible by r.

Theorem 2.3.4. (Turán's 1941) Every n-vertex K_{r+1}-free graph has $\leq \left(1 - \frac{1}{r}\right) \frac{n^2}{2}$ edges.

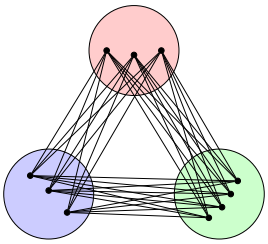


Figure 3: The Turán graph $T_{10,3}$.

Proof. Since G is K_{r+1}-free, by Corollary 2.3.3, letting $\overline{d}$ be average degree and $m = n\overline{d}/2$ be the number of edges, we see that the size $\omega(G)$ of the largest clique of G satisfies

$$r \geq \omega(G) \geq \sum_{v \in V} \frac{1}{n - d_v} \geq \frac{n}{n - \overline{d}} = \frac{n}{n - 2m/n}.$$

Rearranging gives $m \leq \left(1 - \frac{1}{r}\right) \frac{n^2}{2}$. □

Remark. By a careful refinement of the above argument, we can deduce Turán's theorem that $T_{n,r}$ maximizes the number of edges in a n-vertex K_{r+1}-free graph, by noting that $\sum_{v \in V} \frac{1}{n - d_v}$ is minimized over fixed $\sum_v d_v$ when the degrees are nearly equal.

2.4 Crossing number inequality

Consider drawings of graphs on a plane using continuous curves as edges.

The **crossing number** $\mathrm{cr}(G)$ is the minimum number of crossings in a drawing of G.

A graph is **planar** if $\mathrm{cr}(G) = 0$.

$K_{3,3}$ and K_5 are non-planar; furthermore, the following famous theorem characterizes these two graphs as the only obstructions to planarity

Kuratowski's theorem (1930): every non-planar graph contains a subgraph that is topologically homeomorphic to $K_{3,3}$ or K_5

(Also related: Wagner's theorem (1937) says that a graph is planar if and only if it does not have $K_{3,3}$ or K_5 as a minor. It is not too hard to show that Wagner's theorem and Kuratowski's theorem are equivalent)

Question 2.4.1. What is the minimum possible number of crossings that a drawing of:

- K_n? (Hill's conjecture)
- $K_{n,n}$? (Zarankiewicz conjecture; Turán's brick factory problem)
- a graph on n vertices and $n^2/100$ edges?

The following result, due to Ajtai–Chvátal–Newborn–Szemerédi (1982) and Leighton (1984), lower bounds the number of crossings for graphs with many edges.

Theorem 2.4.2 (Crossing number inequality)**.** In a graph $G = (V, E)$, if $|E| \geq 4|V|$, then

$$\operatorname{cr}(G) \gtrsim \frac{|E|^3}{|V|^2}$$

Corollary 2.4.3. In a graph $G = (V, E)$, if $|E| \gtrsim |V|^2$, then $\operatorname{cr}(G) \gtrsim |V|^4$.

Proof. Recall **Euler's formula:** $v - e + f = 2$ for every connected planar graph

For every connected planar graph with at least one cycle, $3|F| \leq 2|E|$ since every face is adjacent to ≥ 3 edges, whereas every edge is adjacent to exactly 2 faces. Plugging into Euler, $|E| \leq 3|V| - 6$.

Thus $|E| \leq 3|V|$ for all planar graphs. Hence $\operatorname{cr}(G) > 0$ whenever $|E| > 3|V|$.

By deleting one edge for each crossing, we get a planar graph, so $|E| - \operatorname{cr}(G) > 3|V|$, i.e.,

$$\operatorname{cr}(G) \geq |E| - 3|V|$$

For graphs with $|E| = \Theta(n^2)$, this games $\operatorname{cr}(G) \gtrsim n^2$. This not a great bound. We will use the probabilistic method to boost this bound.

Let $p \in [0, 1]$ to be decided. Let $G' = (V', E')$ be obtained from G by randomly keeping each vertex with probability p. Then

$$\operatorname{cr}(G') \geq |E'| - 3|V'|$$

So

$$\mathbb{E} \operatorname{cr}(G') \geq \mathbb{E}|E'| - 3\mathbb{E}|V'|$$

We have $\mathbb{E} \operatorname{cr}(G') \leq p^4 \operatorname{cr}(G)$, $\mathbb{E}|E'| = p^2|E|$ and $\mathbb{E}|V'| = p\mathbb{E}|V|$. So

$$p^4 \operatorname{cr}(G) \geq p^2|E| - 3p|V|.$$

Thus

$$\mathrm{cr}(G) \geq p^{-2}|E| - 3p^{-3}|V|.$$

Setting $p \in [0,1]$ so that $4p^{-3}|V| = p^{-2}|E|$, we obtain $\mathrm{cr}(G) \gtrsim |E|^3 / |V|^2$. □

2.4.1 Application to incidence geometry

Question 2.4.4. What is the maximum number of incidences between n distinct points and n distinct lines on a plane?

Let $\mathcal{P}$ be a set of points and $\mathcal{L}$ a set of lines. Denote the number of incidences by

$$I(\mathcal{P}, \mathcal{L}) := |\{(p, \ell) \in \mathcal{P} \times \mathcal{L} : p \in \ell\}|$$

Example: n points and n lines:

$$\mathcal{P} = [k] \times [2k^2] \quad \text{and} \quad \mathcal{L} = \{y = mx + b : m \in [k], b \in [k^2]\}$$

Every line contains k points from $\mathcal{P}$. Taking $3k^3 \approx n$ gives $k^4 = \Theta(n^{4/3})$ incidences.

Can we do better?

No. The following foundational theorem in incidence geometry implies that one has $O(n^{4/3})$ incidences between n points and n lines.

Theorem 2.4.5 (Szemerédi–Trotter 1983)**.** Given a set $\mathcal{P}$ of points and $\mathcal{L}$ of lines in $\mathbb{R}^2$,

$$I(\mathcal{P}, \mathcal{L}) \lesssim |\mathcal{P}|^{2/3}|\mathcal{L}|^{2/3} + |\mathcal{P}| + |\mathcal{L}|.$$

We will show how to prove the Szemerédi–Trotter theorem using the crossing number inequality. This proof is due to Székely (1997).

Trivial bound: $I(\mathcal{P}, \mathcal{L}) \leq |\mathcal{P}||\mathcal{L}|$

Using that every pair of points determine at most one line, and counting triples $(p, p', \ell) \in \mathcal{P} \times \mathcal{P} \times \mathcal{L}$ with $p \neq p'$ and $p, p' \in \ell$, this is $\leq |\mathcal{P}|^2$ and

$$\geq \sum_{\ell \in \mathcal{L}} |\mathcal{P} \cap \ell|(|\mathcal{P} \cap \ell| - 1) \geq |I(\mathcal{P}, \mathcal{L})|^2/|\mathcal{L}| - |I(\mathcal{P}, \mathcal{L})|$$

Combining we get

$$I(\mathcal{P}, \mathcal{L}) \lesssim |\mathcal{P}||\mathcal{L}|^{1/2} + |\mathcal{L}|$$

By point-line duality, also

$$I(\mathcal{P},\mathcal{L}) \lesssim |\mathcal{L}||\mathcal{P}|^{1/2} + |\mathcal{P}|$$

This gives $n^{3/2}$ for n points and n lines. Can we do better? Note that this is tight for planes over finite fields. Need to use topology of Euclidean space.

Proof of Szemerédi–Trotter theorem. Assume that there are no lines with < 2 incidences (otherwise remove such lines repeatedly until this is the same; we remove $\leq |\mathcal{L}|$ incidences this way).

Draw a graph based on incidences. Vertices are point in $\mathcal{P}$ and edges join consecutive points of $\mathcal{P}$ on a given line of $\mathcal{L}$.

A line with k incidences gives $k-1 \geq k/2$ edges, so the total number of edges is $\leq |I(\mathcal{P},\mathcal{L})|/2$.

There are at most $|\mathcal{L}|^2$ crossings. So by crossing number inequality

$$|\mathcal{L}|^2 \geq \mathrm{cr}(G) \gtrsim \frac{|E|^3}{|V|^2} \gtrsim \frac{|I(\mathcal{P},\mathcal{L})|^3}{|\mathcal{P}|^2} \quad \text{if } |I(\mathcal{P},\mathcal{L})| \geq 8|\mathcal{P}|.$$

So $I(\mathcal{P},\mathcal{L}) \lesssim |\mathcal{P}|^{2/3}|\mathcal{L}|^{2/3} + |\mathcal{P}|$. Remember to add $|\mathcal{L}|$ to the bound from the first step of the proof (removing lines with < 2 incidences). □

2.5 Dense packing of spheres in high dimensions

Question 2.5.1. What is the maximum density of a packing of non-overlapping unit balls in $\mathbb{R}^n$ for large n?

Here the **density** is fraction of volume occupied (fraction of the box $[-n,n]^d$ as $n \to \infty$)

Let Δ_n denote the supremum of unit ball packing densities in $\mathbb{R}^n$

Exact maximum only solved in dimension $1, 2, 3, 8, 24$. Dimensions 8 and 24 were only solved recently (see this Quanta magazine story). Dimensions 8 and 24 are special because of the existences of highly symmetric lattices (E_8 lattice in dimension 8 and Leech lattice in dimension 24).

What are examples of dense packings?

We can add balls greedily. Any *maximal* packing has density $\geq 2^{-n}$. Doubling the ball radius would cover space

What about lattices? $\mathbb{Z}^n$ has sphere packing density $\mathrm{vol}(B(1/2)) = \frac{\pi^{n/2}}{(n/2)!2^n} < n^{-cn}$.

Best upper bound: Kabatiansky–Levenshtein (1978): $\Delta_n \leq 2^{-(0.599\cdots+o(1))n}$

Existence of a dense lattice? (Optimal lattices known in dimensions 1–8 and 24)

We will use the probabilistic method to show that a random lattice has high density.

How does one pick a random lattice?

A **lattice** the $\mathbb{Z}$-span of of its basis vectors $v_1, \ldots, v_n$. It's covolume (volume of its fundamental domain) is given by $|\det(v_1|v_2|\cdots|v_n)|$.

So every matrix in $\mathrm{SL}_n(\mathbb{R})$ corresponds to a unimodular lattice (i.e., covolume 1).

Every lattice can be represented in different ways by picking a different basis (e.g., $\{v_1 + v_2, v_2\}$). The matrices $A, A' \in \mathrm{SL}_n(\mathbb{R})$ represent the same lattice iff $A' = AU$ for some $U \in \mathrm{SL}_n(\mathbb{Z})$.

So the space of unimodular lattices is $\mathrm{SL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{Z})$, which has a finite Haar measure (even though this space not compact), so can normalize to a probability measure.

We can pick a **random unimodular lattice** in $\mathbb{R}^n$ by picking a random point in $\mathrm{SL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{Z})$ according to its Haar probability measure.

The following classic result of Siegel acts as like a linearity of expectations statement for random lattices.

Theorem 2.5.2 (Siegel mean value theorem)**.** Let L be the random lattice in $\mathbb{R}^n$ as above and $S \subset \mathbb{R}^n$. Then

$$\mathbb{E}|S \cap L \setminus \{0\}| = \lambda_{\mathrm{Leb}}(S)$$

Proof sketch.

1. $\mu(S) = \mathbb{E}|S \cap L \setminus \{0\}|$ defines a measure on $\mathbb{R}^n$ (it is additive by linearity of expectations)
2. This measure is invariant under $\mathrm{SL}_n(\mathbb{R})$ action (since the random lattice is choosen with respect to Haar measure)
3. Every $\mathrm{SL}_n(\mathbb{R})$-invariant measure on $\mathbb{R}^n$ is a constant multiple of the Lebesgue measure.
4. By considering a large ball S, deduce that $c = 1$. □

Theorem 2.5.3 (Minkowski 1905)**.** For every n, there exist a lattice sphere packing in $\mathbb{R}^n$ with density $\geq 2^{-n}$.

Proof. Let S be a ball of volume 1 (think $1 - \epsilon$ for arbitrarily small $\epsilon > 0$ if you like) centered at the origin. By the Siegel mean value theorem, the random lattice is has expected 1 nonzero lattice point in S, so with positive probability it has no nonzero

lattice point in S. Putting a copy of $\frac{1}{2}S$ (volume 2^{-n}) at each lattice point then gives a lattice packing of density $\geq 2^{-n}$ □

Here is a factor 2 improvement. Take S to be a ball of volume 2. Note that the number of nonzero lattice points in S must be even (if $x \in S$ then $-x \in S$). So same argument gives lattice packing of density $\geq 2^{-n+1}$.

The above improvement uses 2-fold symmetry of $\mathbb{R}^n$. Can we do better by introducing more symmetry?

Historically, a bunch of improvements of the form $\geq cn2^{-n}$ for a sequence of improving constants $c > 0$

Venkatesh (2012) showed that one can get a lattice with a k-fold symmetry by building it using two copies of the cyclotomic lattice $\mathbb{Z}[\omega]$ where $\omega = e^{2\pi/k}$. Every lattice of this form has k-fold symmetry by multiplication by ω.

Skipping details, one can extend the earlier idea to choose a random unimodular lattice in in dimension $n = 2\phi(k)$ with k-fold length-preserving symmetry (without fixed points). An extension of Siegel mean value theorem also holds in this case.

By apply same argument with S being a ball of volume k, we get a a lattice packing of density $\geq k2^{-n}$ in $\mathbb{R}^n$. This bound can be optimized (in term of asymptotics along a subsequence of n) by taking primorial $k = p_1 p_2 \cdots p_m$ where $p_1 < p_2 < \cdots$ are the prime numbers. This gives the current best known bound:

Theorem 2.5.4 (Venkatesh 2012)**.** For infinitely many n, there exists a lattice sphere packing in $\mathbb{R}^n$ of density

$$\geq (e^{-\gamma} - o(1))n \log\log n 2^{-n}.$$

Here $\gamma = 0.577\ldots$ is Euler's constant.

Open problem 2.5.5. Do there exist lattices (or sphere packings) in $\mathbb{R}^n$ with density $\geq (c + o(1))^n$ for some constant $c > 1/2$?

2.6 Unbalancing lights

Theorem 2.6.1. Let $a_{ij} = \pm 1$ for all $i, j \in [n]$. There exists $x_i, y_j \in \{-1, 1\}$ for all $i, j \in [n]$ such that

$$\sum_{i,j=1}^{n} a_{ij} x_i y_j \geq \left(\sqrt{\frac{2}{\pi}} + o(1)\right) n^{3/2}$$

Interpretation: $n \times n$ array of lights. Can flip rows and columns. Want to turn on as many lights as possible.

Proof. Choose $y_1, \dots, y_n$ randomly. And then choose x_i to make the i-th row sum nonnnegative. Let

$$R_i = \sum_{j=1}^{n} a_{ij} y_j \qquad \text{and} \qquad R = \sum_{i=1}^{n} |R_i|.$$

How is R_i distributed? Same distribution as $S_n = \epsilon_1 + \cdots + \epsilon_n$, a sum of n i.i.d. uniform $\{-1, 1\}$. And so for every i

$$\mathbb{E}[|R_i|] = \mathbb{E}[|S_n|] = \left(\sqrt{\frac{2}{\pi}} + o(1)\right)\sqrt{n},$$

e.g., by central limit theorem

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}\left[\frac{|S_n|}{\sqrt{n}}\right] &= \mathbb{E}[|X|] \qquad \text{where } X \sim \text{Normal}(0,1) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |x| e^{-x^2/2}\, dx = \sqrt{\frac{2}{\pi}}\end{aligned}$$

(one can also use binomial sum identities to compute exactly: $\mathbb{E}[|S_n|] = n2^{1-n}\binom{n-1}{\lfloor (n-1)/2 \rfloor}$, though it is rather unnecessary to do so.) Thus

$$\mathbb{E}[R] = \left(\sqrt{\frac{2}{\pi}} + o(1)\right) n^{3/2}.$$

Thus with positive probability, $R \geq \left(\sqrt{\frac{2}{\pi}} + o(1)\right) n^{3/2}$. □

The next example is tricky. The proof will set up a probabilistic process where the parameters are not given explicitly. A compactness argument will show that a good choice of parameters exists.

Theorem 2.6.2. Let $V = V_1 \cup \cdots \cup V_k$, where $V_1, \dots, V_k$ are disjoint sets of size n. The edges of the complete k-uniform hypergraph on V are colored with red/blue. Suppose that every edge formed by taking one vertex from each $V_1, \dots, V_k$ is colored blue. Then there exists $S \subset V$ such that the number of red edges and blue edges in S differ by more than $c_k n^k$, where $c_k > 0$ is a constant.

Proof. Let's do this proof for $k = 3$. Proof easily generalizes to other k.

Let p_1, p_2, p_3 be real numbers to be decided. We are going to pick S randomly by including each vertex in V_i with probability p_i, independently. Let

$$a_{i,j,k} = \#\{\text{blue edges in } V_i \times V_j \times V_k\} - \#\{\text{red edges in } V_i \times V_j \times V_k\}.$$

Then

$$\mathbb{E}[\#\{\text{red edges in } S\} - \#\{\text{blue edges in } S\}]$$

equals to some polynomial

$$f(p_1, p_2, p_3) = \sum_{i \le j \le k} a_{i,j,k} p_i p_j p_k = n^3 p_1 p_2 p_3 + a_{1,1,1} p_1^3 + a_{1,1,2} p_1^2 p_2 + \cdots .$$

(note that $a_{1,2,3} = n^3$ by hypothesis). We would be done if we can find $p_1, p_2, p_3 \in [0,1]$ such that $|f(p_1, p_2, p_3)| > c$ for some constant $c > 0$ (not depending on the $a_{i,j,k}$'s). Note that $|a_{i,j,k}| \le n^3$. We are done after the following lemma

Lemma 2.6.3. Let P_k denote the set of polynomials $g(p_1, \ldots, p_k)$ of degree k, whose coefficients have absolute value ≤ 1, and the coefficient of $p_1 p_2 \cdots p_k$ is 1. Then there is a constant $c_k > 0$ such that for all $g \in P_k$, there is some $p_1, \ldots, p_k \in [0,1]$ with $|g(p_1, \ldots, p_k)| \ge c$.

Proof of Lemma. Set $M(g) = \sup_{p_1, \ldots, p_k \in [0,1]} |g(p_1, \ldots, p_k)|$ (note that sup is achieved as max due to compactness). For $g \in P_k$, since g is nonzero (its coefficient of $p_1 p_2 \cdots p_k$ is 1), we have $M(g) > 0$. As P_k is compact and $M \colon P_k \to \mathbb{R}$ is continuous, M attains a minimum value $c = M(g) > 0$ for some $g \in P_k$. ■ □

3 Alterations

3.1 Ramsey numbers

Recall from Section 1.1:

$R(s,t)$ = smallest n such that every red/blue edge coloring of K_n contains a red K_s or a blue K_t

Using the basic method (union bounds), we deduce

Theorem 3.1.1. If there exists $p \in [0,1]$ with

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

then $R(s,t) > n$.

Proof sketch. Color edge red with prob p and blue with prob $1-p$. LHS upper bounds the probability of a red K_s or a blue K_t. □

Using the alteration method, we deduce

Theorem 3.1.2. For all $p \in [0,1]$ and n,

$$R(s,t) > n - \binom{n}{s} p^{\binom{s}{2}} - \binom{n}{t} (1-p)^{\binom{t}{2}}$$

Proof sketch. Color edge red with prob p and blue with prob $1-p$ remove one vertex from each red K_s or blue K_t. RHS lower bounds the expected number remaining vertices. □

3.2 Dominating set in graphs

In a graph $G = (V, E)$, we say that $U \subset V$ is **dominating** if every vertex in $V \setminus U$ has a neighbor in U.

Theorem 3.2.1. Every graph on n vertices with minimum degree $\delta > 1$ has a dominating set of size at most $\left(\frac{\log(\delta+1)+1}{\delta+1}\right) n$.

Naive attempt: take out vertices greedily. The first vertex eliminates $1+\delta$ vertices, but subsequent vertices eliminate possibly fewer vertices.

Proof. Two-step process (alteration method):
1. Choose a random subset
2. Add enough vertices to make it dominating

Let $p \in [0,1]$ to be decided later. Let X be a random subset of V where every vertex is included with probability p independently.

Let $Y = V \setminus (X \cup N(X))$. Each $v \in V$ lies in Y with probability $\leq (1-p)^{1+\delta}$.

Then $X \cup Y$ is dominating, and

$$\mathbb{E}[|X \cup Y|] = \mathbb{E}[|X|] + \mathbb{E}[|Y|] \leq pn + (1-p)^{1+\delta} n \leq (p + e^{-p(1+\delta)})n$$

using $1 + x \leq e^x$ for all $x \in \mathbb{R}$. Finally, setting $p = \frac{\log(\delta+1)}{\delta+1}$ to minimize $p + e^{-p(1+\delta)}$, we bound the above expression by

$$\leq \left(\frac{1 + \log(\delta+1)}{\delta+1} \right). \qquad \square$$

3.3 Heilbronn triangle problem

Question 3.3.1. How can one place n points in the unit square so that no three points forms a triangle with small area?

Let

$$\Delta(n) = \sup_{\substack{S \subset [0,1]^2 \\ |S|=n}} \min_{\substack{p,q,r \in S \\ \text{distinct}}} \text{area}(pqr)$$

Naive constructions fair poorly. E.g., n points around a circle has a triangle of area $\Theta(1/n^3)$ (the triangle formed by three consectutive points has side lengths $\asymp 1/n$ and angle $\theta = (1 - 1/n)2\pi$). Even worse is arranging points on a grid, as you would get triangles of zero area.

Heilbronn conjectured that $\Delta(n) = O(n^{-2})$.

Komlós, Pintz, and Szemerédi (1982) disproved the conjecture, showing $\Delta(n) \gtrsim n^{-2} \log n$. They used an elaborate probabilistic construction. Here we show a much simpler version probabilistic construction that gives a weaker bound $\Delta(n) \gtrsim n^{-2}$.

Remark. The currently best upper bound known is $\Delta(n) \leq n^{-8/7+o(1)}$ (Komlós, Pintz, and Szemerédi 1981)

Theorem 3.3.2. For every positive integer n, there exists a set of n points in $[0,1]^2$ such that every triple spans a triangle of area $\geq cn^{-2}$, for some absolute constant $c > 0$.

Proof. Choose $2n$ points at random. For every three random points p, q, r, let us estimate

$$\mathbb{P}_{p,q,r}(\text{area}(p,q,r) \leq \epsilon).$$

By considering the area of a circular annulus around p, with inner and outer radii x and $x + \Delta x$, we find

$$\mathbb{P}_{p,q}(|pq| \in [x, x + \Delta x]) \leq \pi((x + \Delta x)^2 - x^2)$$

So the probability density function satisfies

$$\mathbb{P}_{p,q}(|pq| \in [x, x + dx]) \leq 2\pi x dx$$

For fixed p, q

$$\mathbb{P}_r(\text{area}(pqr) \leq \epsilon) = \mathbb{P}_r\left(\text{dist}(pq, r) \leq \frac{2\epsilon}{|pq|}\right) \lesssim \frac{\epsilon}{|pq|}$$

Thus, with p, q, r at random

$$\mathbb{P}_{p,q,r}(\text{area}(pqr) \leq \epsilon) \lesssim \int_0^{\sqrt{2}} 2\pi x \frac{\epsilon}{x} dx \asymp \epsilon.$$

Given these $2n$ random points, let X be the number of triangles with area $\leq \epsilon$. Then $\mathbb{E}X = O(\epsilon n^3)$.

Choose $\epsilon = c/n^2$ with $c > 0$ small enough so that $\mathbb{E}X \leq n$.

Delete a point from each triangle with area $\leq \epsilon$.

The expected number of remaining points is $\mathbb{E}[2n - X] \geq n$, and no triangles with area $\leq \epsilon = c/n^2$.

Thus with positive probability, we end up with $\geq n$ points and no triangle with area $\leq c/n^2$. □

Algebraic construction. Here is another construction due to Erdős (in appendix of Roth (1951)) also giving $\Delta(n) \gtrsim n^{-2}$:

Let p be a prime. The set $\{(x, x^2) \in \mathbb{F}_p^2 : x \in \mathbb{F}_p\}$ has no 3 points collinear (a parabola meets every line in ≤ 2 points). Take the corresponding set of p points in $[p]^2 \subset \mathbb{Z}^2$. Then every triangle has area $\geq 1/2$ due to Pick's theorem. Scale back down to a unit square. (If n is not a prime, then use that there is a prime between n and $2n$.)

3.4 Markov's inequality

We note an important tool that will be used next.

Markov's inequality. Let $X \geq 0$ be random variable. Then for every $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Proof. $\mathbb{E}[X] \geq \mathbb{E}[X 1_{X \geq a}] \geq \mathbb{E}[a 1_{X \geq a}] = a\mathbb{P}(X \geq a)$ □

Take-home message: for r.v. $X \geq 0$, if $\mathbb{E}X$ is *very* small, then *typically* X is small.

3.5 High girth and high chromatic number

If a graph has a k-clique, then you know that its chromatic number is at least k.

Conversely, if a graph has high chromatic number, is it always possible to certify this fact from some "local information"?

Surprisingly, the answer is no. The following ingenious construction shows that a graph can be "locally tree-like" while still having high chromatic number.

The **girth** of a graph is the length of its shortest cycle.

Theorem 3.5.1 (Erdős 1959)**.** For all k, ℓ, there exists a graph with girth $> \ell$ and chromatic number $> k$.

Proof. Let $G \sim G(n,p)$ with $p = (\log n)^2/n$ (the proof works whenever $\log n/n \ll p \ll n^{-1+1/\ell}$). Here $G(n,p)$ is Erdős–Rényi random graph (n vertices, every edge appearing with probability p independently).

Let X be the number of cycles of length at most ℓ in G. By linearity of expectations, as there are exactly $\binom{n}{i}(i-1)!/2$ cycles of length i in K_n for each $3 \leq i \leq n$, we have (recall that ℓ is a constant)

$$\mathbb{E}X = \sum_{i=3}^{\ell} \binom{n}{i} \frac{(i-1)!}{2} p^i \leq \sum_{i=3}^{\ell} n^i p^i = o(n).$$

By Markov's inequality

$$\mathbb{P}(X \geq n/2) \leq \frac{\mathbb{E}X}{n/2} = o(1).$$

(This allows us to get rid of all short cycles.)

How can we lower bound the chromatic number $\chi(\cdot)$? Note that $\chi(G) \geq |V(G)|/\alpha(G)$, where $\alpha(G)$ is the independence number (the size of the largest independent set).

With $x = (3/p)\log n$,

$$\mathbb{P}(\alpha(G) \geq x) \leq \binom{n}{x}(1-p)^{\binom{x}{2}} < n^x e^{-px(x-1)/2} = (ne^{-p(x-1)/2})^x = o(1).$$

Let n be large enough so that $\mathbb{P}(X \geq n/2) < 1/2$ and $\mathbb{P}(\alpha(G) \geq x) < 1/2$. Then there is some G with fewer than $n/2$ cycles of length $\leq \ell$ and with $\alpha(G) \leq (3/p)\log n$.

Remove a vertex from each cycle to get G'. Then $|V(G')| \geq n/2$, girth $> \ell$, and $\alpha(G') \leq \alpha(G) \leq (3/p)\log n$, so

$$\chi(G') \geq \frac{|V(G')|}{\alpha(G')} \geq \frac{np}{6\log n} = \frac{\log n}{6} > k$$

if n is sufficiently large. □

Remark. Erdős (1962) also showed that in fact one needs to see at least a linear number of vertices to deduce high chromatic number: for all k, there exists $\epsilon = \epsilon_k$ such that for all sufficiently large n there exists an n-vertex graph with chromatic number $> k$ but every subgraph on $\lfloor \epsilon n \rfloor$ vertices is 3-colorable. (In fact, one can take $G \sim G(n, C/n)$; see "Probabilistic Lens: Local coloring" in Alon–Spencer)

3.6 Greedy random coloring

Recall $m(k)$ is the minimum number of edges in a k-uniform hypergraph that is not 2-colorable.

Earlier we proved that $m(k) \geq 2^{k-1}$. Indeed, given a k-graph with $< 2^{k-1}$ edges, by randomly coloring the vertices, the expected number of monochromatic numbers is < 1.

We also proved an upper bound $m(k) = O(k^2 2^k)$ by taking a random k-uniform hypergraph on k^2 vertices.

Here is the currently best known lower bound.

Theorem 3.6.1 (Radhakrishnan and Srinivasan (2000)). $m(k) \gtrsim \sqrt{\frac{k}{\log k}} 2^k$

Here we present a simpler proof, based on a **random greedy coloring**, due to Cherkashin and Kozik (2015), following an approach of Pluhaár (2009).

Proof. Suppose H is a k-graph with m edges.

Map $V(H) \to [0,1]$ uniformly at random.

Color vertices greedily from left to right: color a vertex blue unless it would create a monochromatic edge, in which case color it red (i.e., every red vertex is the final vertex in an edge with all earlier $k-1$ vertices have been colored blue).

The resulting coloring has no all-blue edges. What is the probability of seeing a red edge?

If there is a red edge, then there must be two edges e, f so that the last vertex of e is the first vertex of f. Call such pair (e,f) **conflicting**.

Want to bound probability of seeing a conflicting pair in a random $V(H) \to [0,1]$.

Here is an attempt (an earlier weaker result due to Pluhaár (2009)). Each pair of edges with exactly one vertex in common conflicts with probability $\frac{(k-1)!^2}{(2k-1)!} = \frac{1}{2k-1}\binom{2k-2}{k-1}^{-1} \asymp k^{-1/2}2^{-2k}$; union bounding over $< m^2$ pairs of edges, the probability of of getting a conflicting edge is $\lesssim m^2 k^{-1/2}2^{-2k}$, which is < 1 for some $m \asymp k^{1/4}2^k$.

We'd like to do better by more carefully analyzing conflicting edges. Continuing ...

Write $[0,1] = L \cup M \cup R$ where (p to be decided)

$$L := \left[0, \frac{1-p}{2}\right) \qquad M := \left[\frac{1-p}{2}, \frac{1+p}{2}\right] \qquad R := \left(\frac{1+p}{2}, 1\right].$$

The probability that a given edge lands entirely in L is $(\frac{1-p}{2})^k$, and likewise with R

So probability that some edge of H is entirely contained in L or contained in R is $\leq 2m(\frac{1-p}{2})^k$.

Suppose that no edge of H lies entirely in L or entirely in R. If (e,f) conflicts, then their unique common vertex $x_v \in e \cap f$ must lie in M. So the probability that (e,f) conflicts is (here we use $x(1-x) \leq 1/4$)

$$\int_{(1-p)/2}^{(1+p)/2} x^{k-1}(1-x)^{k-1}\,dx \leq p4^{-k+1}.$$

Thus the probability of seeing any conflicting pair is

$$\leq 2m\left(\frac{1-p}{2}\right)^k + m^2 p 4^{-k+1} < 2^{-k+1}me^{-pk} + (2^{-k+1}m)^2 p.$$

Set $p = \log(2^{-k+2}k/m)/k$, we find that the above probability is < 1 for $m = c2^k\sqrt{k/\log k}$, with $c > 0$ being a sufficiently small constant. □

4 Second moment method

Previously, we used $\mathbb{E}X \geq a$ to deduce $\mathbb{P}(X \geq a) > 0$. We also saw from Markov's inequality that for $X \geq 0$, if $\mathbb{E}X$ is very small, then X is small with high probability.

Does $\mathbb{E}X$ being (very) large imply that X is large with high probability?

No! X could be almost always small but $\mathbb{E}X$ could still be large due to outliers (rare large values of X).

Often we want to show that some random variable is **concentrated** around its mean. This would then imply that outliers are unlikely.

We will see many methods in this course on proving concentrations of random variables. We begin with the simplest method. It is the easiest to execute, requires the least hypotheses, but only produces weak (though often useful) concentration bounds.

Second moment method: show that a random variable is concentrated near its mean by bounding its variance.

Variance: $\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Notation convention: mean μ, variance σ^2, standard deviation σ.

Theorem 4.0.1 (Chebyshev's inequality)**.** Let X be a random variable with mean μ and standard deviation σ. For any $\lambda > 0$

$$\mathbb{P}(|X - \mu| \geq \lambda\sigma) \leq \lambda^{-2}.$$

Proof. By Markov's inequality,

$$LHS = \mathbb{P}(|X - \mu|^2 \geq \lambda^2\sigma^2) \leq \frac{\mathbb{E}[(X-\mu)^2]}{\lambda^2\sigma^2} = \frac{1}{\lambda^2}. \qquad \square$$

Remark. Concentration bounds that show small probability of deviating from the mean are called **tail bounds** (also: upper tail bounds for bounding $\mathbb{P}(X \geq \mu + a)$ and lower tail bounds for bounding $\mathbb{P}(X \leq \mu - a)$). Chebyshev's inequality gives tail bounds with polynomial decay. Later on we will see tools that give much better decay (usually exponential) provided additional assumptions on the random variable (e.g., independence).

We can rewrite Chebyshev's inequality as

$$\mathbb{P}(|X - \mathbb{E}X| \geq \epsilon\mathbb{E}X) \leq \frac{\mathrm{Var}\, X}{\epsilon^2(\mathbb{E}X)^2}.$$

Corollary 4.0.2. If $\mathrm{Var}[X] = o(\mathbb{E}X)^2$ then $X \sim \mathbb{E}X$ whp.

Remark. We are invoking asymptotics here (so we are actually considering a sequence X_n of random variables instead of a single one). The conclusion is equivalent to that for every $\epsilon > 0$, one has $|X - \mathbb{E}X| \leq \epsilon \mathbb{E}X$ with probability $1 - o(1)$ as $n \to \infty$.

Variance can be calculated from pairwise covariances. Recall the **covariance**

$$\mathrm{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

So $\mathrm{Var}[X] = \mathrm{Cov}[X, X]$. Covariance is bilinear in X and Y, i.e., for constants $a_1, \ldots$ and $b_1, \ldots$, one has

$$\mathrm{Cov}\left[\sum_i a_i X_i, \sum_j b_j Y_j\right] = \sum_{i,j} a_i b_j \,\mathrm{Cov}[X_i, Y_j].$$

Thus, given $X = X_1 + \cdots + X_n$ (no assumptions on dependencies between the X_i's), we have

$$\mathrm{Var}[X] = \mathrm{Cov}[X, X] = \sum_{i,j \in [n]} \mathrm{Cov}[X_i, X_j] = \sum_{i \in [n]} \mathrm{Var}[X_i] + 2 \sum_{i<j} \mathrm{Cov}[X_i, X_j]$$

We have $\mathrm{Cov}[X, Y] = 0$ if X and Y are independent. Thus in the sum we only need to consider dependent pairs (i, j).

Example 4.0.3 (Sum of independent Bernoulli)**.** Suppose $X = X_1 + \cdots + X_n$ with X_i iid $X_i \sim \mathrm{Bernoulli}(p)$, i.e., $X = 1$ with prob p and $X = 0$ with prob $1 - p$.

Then $\mu = np$ and $\sigma^2 = np(1 - p)$. If $np \gg 1$ then $\sigma \ll \mu$ and thus $X = \mu + o(\mu)$ whp.

Note that the above computation remains identical even if we only knew that the X_i's are *pairwise uncorrelated* (much weaker than assuming full independence).

Here the "tail probability" (the bound hidden in "whp") decays polynomially in the deviation. Later on we will derive much sharper rates of decay (exponential) using more powerful tools such as the Chernoff bound when the r.v.'s are independent.