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)

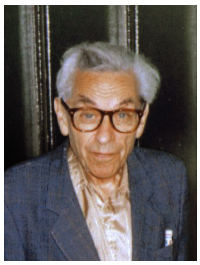


Figure 1: Paul Erdős (1913–1996) is considered the father of the probabilistic method. He published around 1,500 papers during his lifetime, and had more than 500 collaborators. To learn more about Erdős, see his biography *The man who loved only numbers* by Hoffman and the documentary *N is a number*.

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 1.1.1 (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 1.1.2 (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 1.1.3. 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 1.1.4 (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 1.1.5. 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 1.1.6. 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 1.1.7 (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 1.1.8. 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 1.1.9 (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 1.1.10. 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 1.2.5. 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 1.2.6. 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 1.2.7. 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 1.2.8. 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 1.2.9 (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 1.2.10. 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

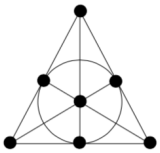
An **k-uniform hypergraph** (or **k-graph**) is a pair $H = (V, E)$, where V (vertices) is a finite set and E (edges) is a set of k-element subsets of E, i.e., $E \subseteq \binom{V}{k}$ (so hypergraphs are really the same concept as set families).

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 1.3.2. 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 1.3.3 (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 1.3.3 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 2.2.2. 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 2.3.3. Equality occurs if G is a disjoint union of cliques.

Remark 2.3.4 (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 2.3.5. 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 2.3.6. (Turán's 1941) Every n-vertex K_{r+1}-free graph has $\leq \left(1 - \frac{1}{r}\right) \frac{n^2}{2}$ edges.

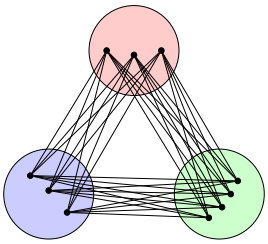


Figure 3: The Turán graph $T_{10,3}$.

Proof. Since G is K_{r+1}-free, by Corollary 2.3.5, 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 2.3.7. 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.,

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

This is a "cheap bound" that we will boost using the probabilistic method.

For graphs with $|E| = \Theta(n^2)$, this gives $\operatorname{cr}(G) \gtrsim n^2$. 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 $\mathbb{E}\,\mathrm{cr}(G') \le p^4\,\mathrm{cr}(G)$, $\mathbb{E}|E'| = p^2|E|$ and $\mathbb{E}|V'| = p\mathbb{E}|V|$. So

$$p^4\,\mathrm{cr}(G) \ge p^2|E| - 3p|V|.$$

Thus

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

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: $I(\mathcal{P},\mathcal{L}) \le |\mathcal{P}||\mathcal{L}|$

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 $p \ne p'$ and $p,p' \in \ell$, this is $\le |\mathcal{P}|^2$ and

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

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 $y_1, \ldots, y_n$ 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 $V_1, \ldots, V_k$ 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 $V_1, \ldots, V_k$ 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

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

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

Theorem 3.3.3. 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 3.5.2. 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 4.0.2. 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 4.0.3. If $\mathrm{Var}[X] = o(\mathbb{E}X)^2$ then $X \sim \mathbb{E}X$ whp.

Remark 4.0.4. 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 4.0.5 (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.

Example 4.0.6 (The number of triangles in a random graph)**.** Let

$$X = \text{the number of triangles in the random graph } G(n, p).$$

For vertices $i, j, k \in [n]$, denote the edge indicator variables by $X_{ij} = 1_{ij \text{ is an edge}}$. Let the triangle indicator variables be $X_{ijk} = 1_{ijk \text{ is a triangle}} = X_{ij} X_{ik} X_{jk}$. Then

$$X = \sum_{i<j<k} X_{ijk} = \sum_{i<j<k} X_{ij} X_{ik} X_{jk}.$$

Its expectation is easy to compute, since $\mathbb{E}[X_{ij}X_{ik}X_{jk}] = \mathbb{E}[X_{ij}]\mathbb{E}[X_{ik}]\mathbb{E}[X_{jk}] = p^3$ by independence. So

$$\mathbb{E}X = \binom{n}{3}p^3$$

Now we compute $\operatorname{Var} X$. Unlike in the earlier example, the summands of X are not all independent. Nonetheless, it is easy to compute the variance.

Given two triples T_1, T_2 of vertices

$$\begin{aligned}\operatorname{Cov}[X_{T_1}, X_{T_2}] = \mathbb{E}[X_{T_1}X_{T_2}] - \mathbb{E}[X_{T_1}]\mathbb{E}[X_{T_2}] &= p^{e(T_1\cup T_2)} - p^{e(T_1)+e(T_2)} \\ &= \begin{cases} 0 & \text{if } |T_1 \cap T_2| \le 1 \\ p^5 - p^6 & \text{if } |T_1 \cap T_2| = 2 \\ p^3 - p^6 & \text{if } T_1 = T_2 \end{cases}\end{aligned}$$

Thus

$$\operatorname{Var} X = \sum_{T_1,T_2} \operatorname{Cov}[X_{T_1}, X_{T_2}] = \binom{n}{3}(p^3 - p^6) + \binom{n}{2}n(n-1)(p^5 - p^6) \lesssim n^3p^3 + n^4p^5$$

When do we have $\sigma \ll \mu$? It is equivalent to satisfying both $n^{3/2}p^{3/2} \ll n^3p^3$ (which gives $p \gg 1/n$) and $n^2p^{5/2} \ll n^3p^3$ (which gives $p \gg n^{-2}$). So $\sigma \ll \mu$ if and only if $p \gg 1/n$, and as we saw earlier, in this case $X \sim \mathbb{E}X$ with high probability.

Remark 4.0.7. Later on we will use more powerful tools (including martingale methods/Azuma-Hoeffding inequalities, and also Janson inequalities) to prove better tail bounds on triangle (and other subgraph) counts.

Remark 4.0.8. Actually the number X of triangles in $G(n,p)$ satisfies an asymptotic central limit theorem, i.e., $(X-\mu)/\sigma \to N(0,1)$ in distribution (Rucinski 1988), initially proved via moment of moments (by showing that higher moments of $(X-\mu)/\sigma$ match those of the normal distribution). Later a different proof was found using the "method of projections."

On the other hand, for much sparser random graphs, when $p \lesssim 1/n$, X is asymptotically Poisson.

4.1 Threshold functions for small subgraphs in random graphs

Question 4.1.1. For which $p = p_n$ is $K_4 \subset G(n,p)$ true with high probability (i.e., with probability $1 - o(1)$)?

There are two statements that one wants to show:

- (0-statement) if $p = p_n$ is small, then $\mathbb{P}(K_4 \subset G(n,p)) \to 0$ as $n \to \infty$.
- (1-statement) if $p = p_n$ is large, then $\mathbb{P}(K_4 \subset G(n,p)) \to 1$ as $n \to \infty$.

Let X be the number of copies of K_4 in $G(n,p)$.

- To show the 0-statement, it suffices to have $\mathbb{E}X \to 0$, in which case Markov's inequality implies that $\mathbb{P}(X \geq 1) \leq \mathbb{E}X \to 0$ (here we are only using the first moment method).
- To show the 1-statement, it suffices to show $\operatorname{Var} X = o((\mathbb{E}X)^2)$, by the lemma below (second moment method).

For simple applications, e.g., $K_4 \subset G(n,p)$, these two methods turn out to be sufficient. Other applications may require stronger techniques (though sometimes "only" second moment, but much more difficult applications).

Lemma 4.1.2. For any random variable X,

$$\mathbb{P}(X = 0) \leq \frac{\operatorname{Var} X}{(\mathbb{E}X)^2}$$

Proof. By Chebyshev inequality, writing $\mu = \mathbb{E}X$,

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mu| \geq |\mu|) \leq \frac{\operatorname{Var} X}{\mu^2}. \qquad \square$$

Corollary 4.1.3. If $\operatorname{Var} X = o((\mathbb{E}X)^2)$, then $X > 0$ with probability $1 - o(1)$.

Remark 4.1.4. Here is a slightly stronger inequality in the case of nonnegative random variables. It is a special case of the Paley–Zygmund inequality. I am showing it here because it is neat. It makes no difference for our applications whether we use the next lemma or the previous one.

Lemma 4.1.5. For any random variable $X \geq 0$,

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

Proof. We have $\mathbb{P}(X > 0) = \mathbb{E}[1_{X>0}]$. By the Cauchy–Schwarz inequality

$$\mathbb{E}[1_{X>0}]\ \mathbb{E}[X^2] \geq (\mathbb{E}[1_{X>0}X])^2 = (\mathbb{E}X)^2. \qquad \square$$

Definition 4.1.6 (Graph properties)**.** A **graph property** $\mathcal{P}$ is a subset of all graphs. We say that $\mathcal{P}$ is **monotone (increasing)** if whenever $G \in \mathcal{P}$, then any graph obtained by adding edges to G also satisfies $\mathcal{P}$. We say that $\mathcal{P}$ is **non-trivial** if for all sufficiently large n, there exists an n-vertex graph in $\mathcal{P}$ and an n-vertex graph not in $\mathcal{P}$.

Example 4.1.7. Examples of graph properties

- Contains K_4; i.e., $\mathcal{P} = \{G : K_4 \subset G\}$
- Connected
- Hamiltonian
- 3-colorable (a monotone decreasing property)
- Planar (monotone decreasing)
- Contains a vertex of degree 1 (not monotone increasing or decreasing)

Definition 4.1.8 (Threshold function)**.** We say that r_n is a **threshold function** for some graph property $\mathcal{P}$ if

$$\mathbb{P}(G(n, p_n) \text{ satisfies } \mathcal{P}) \to \begin{cases} 0 & \text{if } p_n/r_n \to 0, \\ 1 & \text{if } p_n/r_n \to \infty. \end{cases}$$

Remark 4.1.9. The above definition is most suitable for monotone increasing properties. For other types of properties one may need to adjust the definition appropriately.

Remark 4.1.10. From the definition, we see that if r_n and r'_n are both threshold functions, then they must be within a constant factor of each other. So it is fine to say "*the* threshold" of some property, with the understanding that we do not care about constant factors. Later on we will see that every monotone property *has* a threshold function.

Theorem 4.1.11. A threshold function for containing a K_3 is $1/n$, i.e.,

$$\lim_{n\to\infty} \mathbb{P}(K_3 \subset G(n, p_n)) = \begin{cases} 0 & \text{if } p_n n \to 0 \\ 1 & \text{if } p_n n \to \infty \end{cases}$$

Proof. Let X be the number of triangles in $G(n, p)$. Then $\mu := \mathbb{E}X = \binom{n}{3} p^3 \sim n^3 p^3/6$. Let $\sigma^2 = \operatorname{Var} X$.

If $p \ll 1/n$, then $\mu = o(1)$, so $\mathbb{P}(X \geq 1) = o(1)$ by Markov, and hence $X = 0$ w.h.p.

If $p \gg 1/n$, then $\mu \to \infty$, and we saw earlier that $\sigma \ll \mu$, so whp $X \sim \mu$ and thus $X > 0$ whp. □

Question 4.1.12. What is the threshold for containing a fixed H as a subgraph?

The next calculation is similar in spirit to what we did earlier for triangles, but we would like to be more organized as there may be more interacting terms in the variance calculation.

General setup. Suppose $X = X_1 + \cdots + X_m$ where X_i is the indicator random variable for event A_i. Write $i \sim j$ if $i \neq j$ and the pair of events (A_i, A_j) are not independent. (For variance calculation, we are only considering pairwise dependence. Warning: later on when we study the Lovász Local Lemma, we will need a strong notion of a dependency graph.)

If $i \neq j$ and $i \not\sim j$ then $\mathrm{Cov}[X_i, X_j] = 0$. Otherwise,

$$\mathrm{Cov}[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \leq \mathbb{E}[X_i X_j] = \mathbb{P}[A_i \wedge A_j].$$

Thus

$$\mathrm{Var}\, X = \sum_{i,j} \mathrm{Cov}[X_i, X_j] \leq \mathbb{E}X + \Delta$$

where

$$\Delta = \sum_{(i,j):i\sim j} \mathbb{P}(A_i \wedge A_j)$$

The earlier second moment results (Corollary 4.0.3) imply that

$$\text{If } \mathbb{E}X \to \infty \text{ and } \Delta = o(\mathbb{E}X)^2 \text{ then and } X \sim \mathbb{E}X \text{ and } X > 0 \text{ whp.}$$

We have

$$\sum_{(i,j):i\sim j} \mathbb{P}(A_i \wedge A_j) = \sum_i \mathbb{P}(A_i) \sum_{j:j\sim i} \mathbb{P}\left(A_j \mid A_i\right)$$

In many symmetric situations (e.g. our examples), the following quantity does not depend on i:

$$\Delta^* = \sum_{j:j\sim i} \mathbb{P}\left(A_j \mid A_i\right)$$

(or take Δ^* to be the maximum such value ranging over all i). Then

$$\Delta = \sum_i \mathbb{P}[A_i]\Delta^* = \Delta^* \mathbb{E}X$$

Thus we have

Lemma 4.1.13. If $\mathbb{E}X \to \infty$ and $\Delta^* = o(\mathbb{E}X)$, then $X \sim \mathbb{E}X$ and $X > 0$ whp.

Theorem 4.1.14. A threshold function for containing K_4 is $n^{-2/3}$.

Proof. Let X denote the number of copies of K_4 in $G(n,p)$. Then $\mathbb{E}X = \binom{n}{4}p^6 \sim n^4p^6/24$.

If $p \ll n^{-2/3}$ then $\mathbb{E}X = o(1)$ so $X = 0$ whp

Now suppose $p \gg n^{-2/3}$, so $\mathbb{E}X \to \infty$. For each 4-vertex subset S, let A_S be the event that S is a clique in $G(n,p)$.

For each fixed S, one has $A_S \sim A_{S'}$ if and only if $|S \cap S'| \geq 2$.

- The number of S' that share exactly 2 vertices with S is $6\binom{n}{2} = O(n^2)$, and for each such S' one has $\mathbb{P}(A_{S'}|A_S) = p^5$ (as there are 5 additional edges, no in the S-clique, that needs to appear clique to form the S'-clique).
- The number of S' that share exactly 3 vertices with S is $4(n-4) = O(n)$, and for each such S' one has $\mathbb{P}(A_{S'}|A_S) = p^3$.

Summing over all above S', we find Then

$$\Delta^* = \sum_{S':|S'\cap S|\in\{2,3\}} \mathbb{P}(A_{S'}|A_S) \lesssim n^2p^5 + np^3 \ll n^4p^6 \asymp \mathbb{E}X.$$

Thus $X > 0$ whp by Lemma 4.1.13. □

For both K_3 and K_4, we saw that any choice of $p = p_n$ with $\mathbb{E}X \to \infty$ one has $X > 0$ whp. Is this generally true?

Example 4.1.15 (First moment is not enough)**.** Let $H =$  . We have $\mathbb{E}X_H \asymp n^5p^7$. If $\mathbb{E}X = o(1)$ then $X = 0$ whp. But what if $\mathbb{E}X \to \infty$, i.e., $p \gg n^{-5/7}$?

We know that if $n^{-5/7} \ll p \ll n^{-2/3}$, then $X_{K_4} = 0$ whp, so $X_H = 0$ whp since $K_4 \subset H$.

On the other hand, if $p \gg n^{-2/3}$, then whp can find K_4, and pick an arbitrary edge to extend to H (we'll prove this).

Thus the threshold for $H =$  is actually $n^{-2/3}$, and not $n^{-5/7}$ as one might have naively predicted from the first moment alone.

Why didn't $\mathbb{E}X_H \to \infty$ give $X_H > 0$ whp? In the calculation of Δ^*, one of the terms is $\asymp np$ (from two copies of H with a K_4-overlap), and $np \not\ll n^5p^7 \asymp \mathbb{E}X_H$ if $p \ll n^{-2/3}$.

Definition 4.1.16. Define the **edge-vertex ratio** of a graph H by $\rho(H) = e_H/v_H$. Define the **maximum edge-vertex ratio of a subgraph** of H:

$$m(H) := \max_{H' \subset H} \rho(H').$$

Example 4.1.17. Let $H =$  . We have $\rho(H) = 7/5$ whereas $\rho(K_4) = 3/2 > 7/5$. It is not hard to check that $m(H) = \rho(K_4) = 3/2$ as K_4 is the subgraph of H with the maximum edge-vertex ratio.

Theorem 4.1.18 (Bollobás 1981)**.** Fix a graph H with v_H vertices and e_H edges. Then $p = n^{-1/m(H)}$ is a threshold function for containing H has a subgraph. Furthermore, if $p \gg n^{-1/m(H)}$, then the number X_H of copies of H in $G(n,p)$ satisfies, with probability $1 - o(1)$,

$$X_H \sim \mathbb{E}X_H = \binom{n}{v_H} \frac{v_H!}{\mathrm{aut}(H)} p^{e_H} \sim \frac{n^{v_H} p^{e_H}}{\mathrm{aut}(H)}.$$

Proof. Let H' be a subgraph of H achieving the maximum edge-vertex ratio, i.e., $\rho(H') = m(H)$.

If $p \ll n^{-1/m(H)}$, then $\mathbb{E}X_{H'} \asymp n^{v_{H'}} p^{e_{H'}} = o(1)$, so $X_{H'} = 0$ whp, hence $X_H = 0$ whp.

Now suppose $p \gg n^{-1/m(H)}$. Let us count *labeled* copies of the subgraph H in $G(n,p)$. Let J be a labeled copy of H in K_n, and let A_J denote the event that J appears in $G(n,p)$. We have, for fixed J,

$$\Delta^* = \sum_{J' \sim J} \mathbb{P}\left(A_{J'} \mid A_J\right) = \sum_{J' \sim J} p^{|E(J') \setminus E(J)|}$$

For any $J' \sim J$, we have

$$n^{|V(J') \setminus V(J)|} p^{|E(J') \setminus E(J)|} \ll n^{|V(J)|} p^{|E(J)|}$$

since

$$p \gg n^{-1/m(H)} \geq n^{-1/\rho(J \cap J')} = n^{-|V(J) \cap V(J')|/|E(J) \cap E(J')|}.$$

It then follows, after consider all the possible ways that J' can overlap with J, that $\Delta^* \ll n^{|V(J)|} p^{|E(J)|} \asymp \mathbb{E}X_H$. So Lemma 4.1.13 yields the result. □

4.2 Existence of thresholds

Question 4.2.1. Does every monotone graph property $\mathcal{P}$ have a threshold function?

E.g., could it be the case that $\mathbb{P}(G(n, n^{-1/3}) \in \mathcal{P}), \mathbb{P}(G(n, n^{-1/4}) \in \mathcal{P}) \in [0.1, 0.9]$ for all sufficiently large n?

First, an even simpler question, why is it that if $\mathcal{P}$ is a nontrivial monotone property, then $\mathbb{P}(G(n,p) \in \mathcal{P})$ is an increasing function of p? This is intuitively obvious, but how to prove it?

Let us give two (related) proofs of this basic fact. Both are quite instructive.

More abstractly, this is not really about graphs, but rather about random subsets (for random graphs, we are taking random subgraphs of edges).

Given a collection $\mathcal{F}$ of subsets of $[n]$, we say that $\mathcal{F}$ is an **upward closed set** (or **up-set**) if whenever $A \subset B$ and $A \in \mathcal{F}$ then $B \in \mathcal{F}$. We say that an up-set $\mathcal{F}$ is nontrivial if $\emptyset \notin \mathcal{F}$ and $[n] \in \mathcal{F}$.

Let $[n]_p$ denote the random subset of $[n]$ obtained by including every element independently with probability p.

Theorem 4.2.2. Let $\mathcal{F}$ a nontrivial up-set of $[n]$. Then $p \mapsto \mathbb{P}([n]_p \in \mathcal{F})$ is a strictly increasing function.

The first proof is by **coupling**. Coupling is powerful probabilistic idea. Given two random variables X and Y with individually prescibed distributions, we "couple" them together by considering a single probabilistic process that generates both X and Y in a way that clarifies their relationship. More formally, we construct a joint distribution (X, Y) whose marginals agree with those of X and Y.

Proof 1. (By coupling) Let $0 \leq p < q \leq 1$. Consider the following process to generate two random subsets of $[n]$: pick a uniform random vector $(x_1, \ldots, x_n) \in [0,1]^n$. Let $A = \{i : x_i \leq p\}$ and $B = \{i : x_i \leq q\}$. Then A has the same distribution as $[n]_p$ and B has the same distribution as $[n]_q$. Furthermore, we see that $A \in \mathcal{F}$ implies $B \in \mathcal{F}$. Thus

$$\mathbb{P}([n]_p \in \mathcal{F}) = \mathbb{P}(A \in \mathcal{F}) \leq \mathbb{P}(B \in \mathcal{F}) = \mathbb{P}([n]_q \in \mathcal{F}).$$

To see that the inequality strict, we simply have to observe that with positive probability, one has $A \notin \mathcal{F}$ and $B \in \mathcal{F}$ (e.g., $A = \emptyset$ and $B = [n]$). □

The second proof is also uses coupling, but viewed somewhat differently. The idea is that we can obtain $[n]_p$ as the union of several independent $[n]_{p'}$ for some smaller values of p'.

In other words, we are exposing the random subset in several rounds.

Proof 2. (By two-round exposure) Let $0 \le p < q \le 1$. Note that $B = [n]_q$ has the same distribution as the union of two independent $A = [n]_p$ and $A' = [n]_{p'}$, where p' is chosen to satisfy $1 - q = (1 - p)(1 - p')$. Thus

$$\mathbb{P}(A \in \mathcal{F}) \le \mathbb{P}(A \cup A' \in \mathcal{F}) = \mathbb{P}(B \in \mathcal{F}).$$

Like earlier, to observe that the inequality is strict, one observes that with positive probability, one has $A \notin \mathcal{F}$ and $A \cup A' \in \mathcal{F}$. □

The above technique (generalized from two round exposure to multiple round exposures) gives a nice proof of the following theorem (originally proved using the Kruskal–Katona theorem).

Theorem 4.2.3 (Bollobás and Thomason 1987)**.** Every nontrivial monotone graph property has a threshold function.

Proof. Note that $G(n, 1-(1-p)^k)$ has the same distribution has the union of k independent copies $G^1, \dots, G^k$ of $G(n, p)$. Furthermore, by the monotonicity of the property, if $G^1 \cup \dots \cup G^k \notin \mathcal{P}$, then $G^1, \dots, G^k \notin \mathcal{P}$. By independence,

$$\mathbb{P}(G(n, 1 - (1 - p)^k) \notin \mathcal{P}) = \mathbb{P}(G^1 \cup \dots \cup G^k \notin \mathcal{P}) \le \mathbb{P}(G^1 \notin \mathcal{P}) \cdots \mathbb{P}(G^k \notin \mathcal{P})$$

To simplify notation, let us write

$$f_p = f_p(n) = \mathbb{P}(G(n, p) \in \mathcal{P}).$$

Since $1 - (1 - p)^k \le kp$ (check by convexity), we have that for any monotone graph property $\mathcal{P}$, any positive integer $k \le 1/p$,

$$1 - f_{kp} \le 1 - f_{1-(1-p)^k} \le (1 - f_p)^k. \tag{4.1}$$

Fix any large enough n (so that set of n-vertex graphs satisfying the property $\mathcal{P}$ is a nontrival up-set). Since $p \mapsto f_p(n)$ is a continuous strictly increasing function from 0 to 1 as p goes from 0 to 1 (in fact it is a polynomial in p for each fixed n), there is some "critical" $p_c = p_c(n)$ with $f_{p_c}(n) = 1/2$.

We claim that p_c is a threshold function. Indeed, (4.1) implies, if $p = p(n) \gg p_c(n)$, then, letting $k = k(n) = \lfloor p/p_c \rfloor \to \infty$,

$$1 - f_p \le (1 - f_{p_c})^k = 2^{-k} \to 0$$

so $f_p \to 1$. Likewise, if $p \ll p_c$, then, letting $k = \lfloor p_c/p \rfloor \to \infty$, we have

$$\frac{1}{2} = 1 - f_{p_c} \leq (1 - f_p)^k,$$

and thus $f_p \to 0$ as $n \to \infty$. Thus $p_c(n)$ is a threshold function for $\mathcal{P}$. □

Remark 4.2.4. Note that, by definition, if $p_1(n)$ and $p_2(n)$ are both threshold functions for the same property, then $cp_1(n) \leq p_2(n) \leq Cp_2(n)$ for some constants $0 < c < C$.

Last section we identified the threshold for the property of containing a fixed subgraph. Let us state the result (at least in the case of triangles, but similar results are known for every subgraph) a bit more precisely, where we use the fact that for a constant $c > 0$, the number of triangles in $G(n, c/n)$ converges to a Poisson distribution with mean $c^3/6$ (this can be proved using the "method of moments" but we will not do it here). So

$$\mathbb{P}\left(G\left(n, \frac{c_n}{n}\right) \text{ contains a triangle}\right) \to \begin{cases} 0 & \text{if } c_n \to 0 \\ 1 - e^{-c^3/6} & \text{if } c_n \to c \text{ constant} \\ 1 & \text{if } c_n \to \infty \end{cases}$$

What about other graph properties? It turns out that we can sometimes identity the transition very precisely.

Example 4.2.5. Here are some more examples of threshold functions. The first two statements are in the original Erdős–Rényi (1959) paper on random graphs. The first is an easy (and instructive) exercise in the second moment method.

- With $p = \dfrac{\log n + c_n}{n}$

$$\mathbb{P}\left(G\left(n,p\right) \text{ has no isolated vertices}\right) \to \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty \end{cases}$$

- With $p = \dfrac{\log n + c_n}{n}$

$$\mathbb{P}\left(G\left(n,p\right) \text{ is connected}\right) \to \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty \end{cases}$$

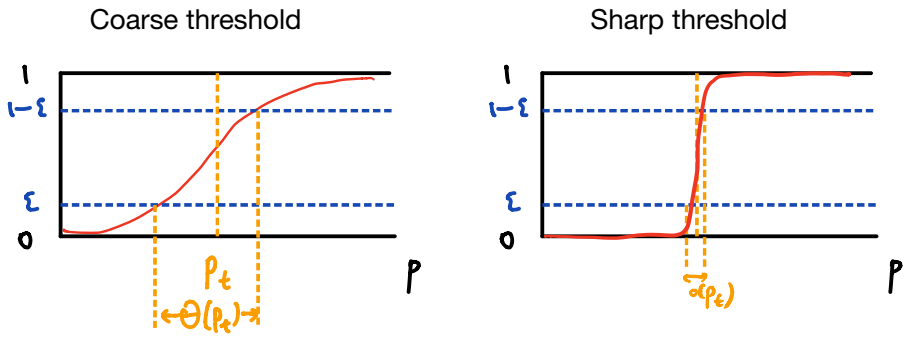


Figure 4: Examples of coarse and sharp thresholds. The vertical axis is the probability that $G(n,p)$ satisfies the property.

In fact, a much stronger statement is true, connecting the above two examples: consider a process where one adds an random edges one at a time, then with probability $1-o(1)$, the graph becomes connected as soon as there are no more isolated vertices.

- With $p = \dfrac{\log n + \log\log n + c_n}{n}$

$$\mathbb{P}(G(n,p) \text{ has a Hamiltonian cycle}) \to \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty \end{cases}$$

Like earlier, it is true that with high probability, a random graph becomes Hamiltonian as soon as its minimum degree reaches 2.

In the above examples, the probability that $G(n,p)$ satisfies the property changes quickly and dramatically as p crosses the threshold (physical analogy: similar to how the structure of water changes dramatically as the temperature drops below freezing). For example, while for connectivity, while $p = \log n/n$ is a threshold function, we see that $G(n, 0.99\log n/n)$ is whp not connected and $G(n, 1.01\log n/n)$ is whp connected, unlike the situation for containing a triangle earlier. We call this the **sharp threshold phenomenon**.

Definition 4.2.6 (Sharp thresholds)**.** We say that r_n is a **sharp threshold** for some graph property $\mathcal{P}$ if, for every $\delta > 0$,

$$\mathbb{P}(G(n,p_n) \text{ satisfies } \mathcal{P}) \to \begin{cases} 0 & \text{if } p_n \le (1-\delta)r_n, \\ 1 & \text{if } p_n \ge (1-\delta)r_n. \end{cases}$$

Equivalently, a graph property $\mathcal{P}$ exhibits a sharp threshold at r_n if, for every $\epsilon > 0$,

for a given large n, as p increases from 0 to 1, the probability $\mathbb{P}(G(n,p) \in \mathcal{P})$ increases from ϵ to $1-\epsilon$ over a short window of width $o(r_n)$ around r_n. On the other hand, if this transition window has width $\Omega(r_n)$ for some $\epsilon > 0$, then we say that it is a **coarse threshold**. See Figure 4.

We saw coarse thresholds for the "local" property of containing some given subgraph, whereas we saw sharp thresholds for "global" properties such as connectivity. It turns out that this is a general phenomenon.

Friedgut's sharp threshold theorem (1999), a deep and important result, roughly says that:

> All monotone graph properties with a coarse threshold may be approximated by a local property.

In other words, informally, if a monotone graph property $\mathcal{P}$ has a coarse threshold, then there is finite list of graph $G_1, \ldots, G_m$ such that $\mathcal{P}$ is "close to" the property of containing one of $G_1, \ldots, G_m$ as a subgraph.

We need "close to" since the property could be "contains a triangle and has at least $\log n$ edges", which is not exactly local but it is basically the same as "contains a triangle."

There is some subtlety here since we can allow very different properties depending on the value of n. E.g., $\mathcal{P}$ could be the set of all n-vertex graphs that contain a K_3 if n is odd and K_4 if n is even. Friedgut's theorem tells us that if there is a threshold, then there is a partition $\mathbb{N} = \mathbb{N}_1 \cup \cdots \cup \mathbb{N}_k$ such that on each $\mathbb{N}_i$, $\mathcal{P}$ is approximately the form described in the previous paragraph.

In the last section, we derived that the property of containing some fixed H has threshold $n^{-1/m(H)}$ for some rational number $m(H)$. It follows as a corollary of Friedgut's theorem that every coarse threshold must have this form.

Corollary 4.2.7 (of Friedgut's sharp threshold theorem)**.** Suppose $r(n)$ is a coarse threshold function of some graph property. Then there is a partition of $\mathbb{N} = \mathbb{N}_1 \cup \cdots \cup \mathbb{N}_k$ and rationals $\alpha_1, \ldots, \alpha_k > 0$ such that $r(n) \asymp n^{-\alpha_j}$ for every $n \in \mathbb{N}_j$.

In particular, if $(\log n)/n$ is a threshold function of some monotone graph property (e.g., this is the case for connectivity), then we automatically know that it must be a sharp threshold, even without knowing anything else about the property. Likewise if the threshold has the form $n^{-\alpha}$ for some irrational α.

The exact statement of Friedgut's theorem is more cumbersome. We refer those who are interested to Friedgut's original 1999 paper and his later survey for details and applications. This topic is connected more generally to an area known as the **analysis of**

boolean functions.

Also, it is known that the transition window of every monotone graph property is $(\log n)^{-2+o(1)}$ (Friedgut—Kalai (1996), Bourgain–Kalai (1997)).

Curiously, tools such as Friedgut's theorem sometimes allow us to prove the existence of a sharp threshold without being able to identify its exact location. For example, it is an important open problem to understand where exactly is the transition for a random graph to be k-colorable.

Conjecture 4.2.8 (k-colorability threshold)**.** For every $k \geq 3$ there is some real constant $d_k > 0$ such that for any constant $d > 0$,

$$\mathbb{P}(G(n, d/n) \text{ is } k\text{-colorable}) \to \begin{cases} 1 & \text{if } d < d_k, \\ 0 & \text{if } d > d_k. \end{cases}$$

We do know that there *exists* a sharp threshold for k-colorability.

Theorem 4.2.9 (Achlioptas and Friedgut 2000)**.** For every $k \geq 3$, there exists a function $d_k(n)$ such that for every $\epsilon > 0$, and sequence $d(n) > 0$,

$$\mathbb{P}\left(G\left(n, \tfrac{d(n)}{n}\right) \text{ is } k\text{-colorable}\right) \to \begin{cases} 1 & \text{if } d(n) < d_k(n) - \epsilon, \\ 0 & \text{if } d(n) > d_k(n) + \epsilon. \end{cases}$$

On the other hand, it is not known whether $\lim_{n\to\infty} d_k(n)$ exists, which would imply Conjecture 4.2.8. Further bounds on $d_k(n)$ are known, e.g. the landmark paper of Achlioptas and Naor (2006) showing that for each fixed $d > 0$, whp $\chi(G(n, d/n) \in \{k_d, k_d + 1\}$ where $k_d = \min\{k \in \mathbb{N} : 2k \log k > d\}$. Also see the later work of Coja-Oghlan and Vilenchik (2013).

4.3 Clique number

The **clique number** $\omega(G)$ of a graph is the maximum number of vertices in a clique of G.

Question 4.3.1. What is the clique number of $G(n, 1/2)$?

Let X be the number of k-cliques of $G(n, 1/2)$. We have

$$f(k) := \mathbb{E}X = \binom{n}{k} 2^{-\binom{k}{2}}.$$

Theorem 4.3.2. Let $k = k(n)$ satisfy $f(k) \to \infty$. Then $\omega(G(n, 1/2)) \geq k$ whp.

Proof. For each k-element subset S of vertices, let A_S be the event that S is a clique. Let X_S be the indicator random variable for A_S. Let $X = \sum_{S \in \binom{[n]}{k}} X_S$ denote the number of k-cliques.

For fixed k-set S, consider all k-set T with $|S \cap T| \geq 2$:

$$\Delta^* = \sum_{\substack{T \in \binom{[n]}{k} \\ 2 \leq |S \cap T| \leq k-1}} \mathbb{P}(A_T | A_S) = \sum_{i=2}^{k-1} \binom{k}{i} \binom{n-k}{k-i} 2^{\binom{i}{2} - \binom{k}{2}} \overset{\text{omitted}}{\ll} \mathbb{E}X = \binom{n}{k} 2^{-\binom{k}{2}}.$$

It then follows from Lemma 4.1.13 that $X > 0$ (i.e., $\omega(G) \geq k$) whp. □

Theorem 4.3.3 (Bollobás–Erdős 1976 and Matula 1976)**.** There exists a $k = k(n) \sim 2 \log_2 n$ such that $\omega(G(n, 1/2)) \in \{k, k+1\}$ whp.

Proof. (Sketch) For $k \sim 2 \log_2 n$,

$$\frac{f(k+1)}{f(k)} = \frac{n-k}{k+1} 2^{-k} = n^{-1+o(1)} = o(1).$$

So the value of $f(k)$ drops rapidly for $k \sim 2 \log_2 n$. Let $k_0 = k_0(n)$ be the value with $f(k_0) \geq 1 > f(k_0 + 1)$. If n is such that $f(k_0) \to \infty$ while $f(k_0 + 1) \to 0$ (it turns out that this is true for most integers n), and thus $\omega(G) = k_0$ whp. When $f(k_0) = O(1)$, we have $f(k_0 - 1) \to \infty$ and $f(k_0 + 1) \to 0$ so one has $\omega(G(n, 1/2)) \in \{k_0 - 1, k_0\}$ whp. □

Remark 4.3.4. The result also implies the same about size of largest independent set in $G(n, 1/2)$ (take complement). Also extends to constant p: $\omega(G(n, p)) \sim 2 \log_{1/(1-p)} n$ whp.

Since the chromatic number satisfies $\chi(G) \geq n/\alpha(G)$, we have

$$\chi(G(n, 1/2)) \geq (1 + o(1)) \frac{n}{2 \log_2 n} \qquad \text{whp.}$$

Later on, using more advanced methods, we will prove $\chi(G(n, 1/2)) \sim n/(2 \log_2 n)$ whp (Bollobás 1987).

Also, later, using martingale concentration, we know show that $\chi(G(n, p))$ is tightly concentrated around its mean without a priori needing to know where the mean is located.

4.4 Hardy–Ramanujan theorem on the number of prime divisors

Let $\nu(n)$ denote the number of primes p dividing n (do not count multiplicities).

The next theorem says that "almost all" n have $(1+o(1))\log\log n$ prime factors

Theorem 4.4.1 (Hardy and Ramanujan 1917)**.** For every $\epsilon > 0$, there exists C such that all but ϵ-fraction of $x \in [n]$ satisfy

$$|\nu(x) - \log\log n| \leq C\sqrt{\log\log n}$$

The original proof of Hardy and Ramanujan was quite involved. Here we show a "probabilistic" proof due to Turán (1934), which played a key role in the development of probabilistic methods in number theory.

Proof. Choose $x \in [n]$ uniformly at random. For prime p, let

$$X_p = \begin{cases} 1 & \text{if } p|x, \\ 0 & \text{otherwise.} \end{cases}$$

Set $M = n^{1/10}$, and (the sum is taken over primes p).

$$X = \sum_{p \leq M} X_p$$

We have $\nu(x) - 10 \leq X(x) \leq \nu(x)$ since x cannot have more than 10 prime factors $> n^{1/10}$. So it suffices to analyze X. Since exactly $\lfloor n/p \rfloor$ positive integers $\leq n$ are divisible by p, we have

$$\mathbb{E}X_p = \frac{\lfloor n/p \rfloor}{n} = \frac{1}{p} + O\left(\frac{1}{n}\right)$$

So

$$\mathbb{E}X = \sum_{p \leq M} \left(\frac{1}{p} + O\left(\frac{1}{n}\right)\right) = \log\log n + O(1)$$

Here we are applying Merten's theorem from analytic number theory: $\sum_{p \leq n} 1/p = \log\log n + O(1)$ (the $O(1)$ error term converges to the Meissel–Mertens constant).

Next we compute the variance. The intuition is that distinct primes should be have independently. Indeed, if pq divides n, then X_p and X_q are independent. Then pq does not divide n, but n is large enough, then there is some small covariance contribution. (Contrast to the earlier calculations in random graphs, where there are very few nonzero

covariance terms, but each can be more significant.)

If $p \neq q$, then $X_p X_q = 1$ if and only if $pq|x$. Thus

$$\begin{aligned}|\mathrm{Cov}[X_p, X_q]| &= |\mathbb{E}[X_p X_q] - \mathbb{E}[X_p]\mathbb{E}[X_q]| \\ &= \left| \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n} \frac{\lfloor n/q \rfloor}{n} \right| \\ &= O\left(\frac{1}{n}\right)\end{aligned}$$

Thus

$$\sum_{p \neq q} |\mathrm{Cov}[X_p, X_q]| \lesssim \frac{M^2}{n} \lesssim n^{-4/5}$$

Also, $\mathrm{Var}\, X_p = \mathbb{E}[X_p] - (\mathbb{E}X_p)^2 = (1/p)(1 - 1/p) + O(1/n)$. Combining, we have

$$\begin{aligned}\mathrm{Var}\, X &= \sum_{p \leq M} \mathrm{Var}\, X_p + \sum_{p \neq q} \mathrm{Cov}[X_p, X_q] \\ &= \sum_{p \leq M} \frac{1}{p} + O(1) = \log\log n + O(1) \sim \mathbb{E}X\end{aligned}$$

Thus by Chebyshev, for every constant $\lambda > 0$

$$\mathbb{P}\left(|X - \log\log n| \geq \lambda \sqrt{\log\log n}\right) \leq \frac{(\mathrm{Var}\, X)^2}{\lambda^2 (\log\log n)} = \frac{1}{\lambda^2} + o(1).$$

Finally, recall that $|X - \nu| \leq 10$, so same asymptotic bound holds with X replaced by ν. □

Theorem 4.4.2 (Erdős and Kac 1940)**.** With $x \in [n]$ uniformly chosen at random, $\nu(x)$ is asymptotically normal, i.e., for every $\lambda \in \mathbb{R}$,

$$\lim_{n \to \infty} \mathbb{P}_{x \in [n]} \left(\frac{\nu(x) - \log\log n}{\sqrt{\log\log n}} \geq \lambda \right) = \frac{1}{\sqrt{2\pi}} \int_{\lambda}^{\infty} e^{-t^2/2}\, dt$$

The intuition is that the number of prime divisors $X = \sum_p X_p$ (from the previous proof) behaves like a sum of independent random variables, the central limit theorem should imply an asymptotic normal distribution.

The original proof of Erdős and Kac verifies the above intuition using some more involved results in analytic number theory. Simpler proofs have been subsequently given, and we outline one below, which is based on computing the moments of the distribution. The idea of computing moments for this problem was first used by Delange (1953), who was

apparently not aware of the Erdős–Kacs paper. Also see a more modern account by Granville and Soundararajan (2007).

The following tool from probability theory allows us to verify asymptotic normality from convergence of moments.

Theorem 4.4.3 (Method of moments)**.** Let X_n be a sequence of real valued random variables such that for every positive integer k, $\lim_{n\to\infty} \mathbb{E}[X_n^k]$ equals to the k-th moment of the standard normal distribution. Then X_n converges in distribution to the standard normal, i.e., $\lim_{n\to\infty} \mathbb{P}(X_n \le a) = \mathbb{P}(Z \le a)$ for every $a \in \mathbb{R}$, where Z is a standard normal.

Remark 4.4.4. The same conclusion holds for any probability distribution (other than normal) that is "determined by its moments," i.e., there are no other distributions sharing the same moments. Many common distributions that arise in practice, e.g., the Poisson distribution, satisfy this property. There are various sufficient conditions for guaranteeing this moments property, e.g., Carleman's condition tells us that any probability distribution whose moments do not increase too quickly is determined by its moments.

Proof sketch of Erdős–Kacs Theorem 4.4.2. We compare higher moments of $X = \nu(x)$ with that of an idealized Y treating the prime divisors as truly random variables.

Set $M = n^{1/s(n)}$ where $s(n) \to \infty$ sufficiently slowly. As earlier, $\nu(x) - s(n) \le \nu(x) \le v(x)$.

We construct a "model random variable" mimicking X. Let $Y = \sum_{p \le M} Y_p$, where $Y_p \sim$ Bernoulli$(1/p)$ independently for all primes $p \le M$. We can compute:

$$\mu := \mathbb{E}Y \sim \mathbb{E}X \sim \log\log n$$

and

$$\sigma^2 := \operatorname{Var} Y \sim \operatorname{Var} X \sim \log\log n.$$

Let $\widetilde{X} = (X - \mu)/\sigma$ and $\widetilde{Y} = (Y - \mu)/\sigma$.

By the central limit theorem (e.g., the Lindeberg CLT), $\widetilde{Y} \to N(0,1)$ in distribution. In particular, $\mathbb{E}[\widetilde{Y}^k] \sim \mathbb{E}[Z^k]$ (asymptotics as $n \to \infty$) where Z is a standard normal.

Let us compare $\widetilde{X}$ and $\widetilde{Y}$. It suffices to show that for every fixed k, $\mathbb{E}[\widetilde{X}^k] \sim \mathbb{E}[\widetilde{Y}^k]$.

For every set of distinct primes $p_1, \dots p_r \le M$,

$$\mathbb{E}[X_{p_1} \cdots X_{p_r} - Y_{p_1} \cdots Y_{p_r}] = \frac{1}{n}\left\lfloor \frac{n}{p_1 \cdots p_r} \right\rfloor - \frac{1}{p_1 \cdots p_r} = O\left(\frac{1}{n}\right)$$

Comparing expansions of $\widetilde{X}^k$ in terms of the X_p's ($n^{o(1)}$ terms), we get

$$\mathbb{E}[\widetilde{X}^k - \widetilde{Y}^k] = n^{-1+o(1)} = o(1).$$

So the moments of $\widetilde{X}$ approach those of $N(0,1)$. The method of moments theorem from probability then implies that $\widetilde{X}$ is asymptotically normally distributed. □

4.5 Distinct sums

Question 4.5.1. Let S be a k-element subset of positive integers such that all 2^k subset sums of S. What is the minimum possible $\max S$?

E.g., $S = \{1, 2, 2^2, \ldots, 2^{k-1}\}$ (the greedy choice).

We begin with an easy pigeonhole argument. On the other hand, since all 2^k sums are distinct and are at most $k \max S$, we have $2^k \leq k \max S$, so $\max S \geq 2^k/k$.

Erdős offered \$300 for a proof or disproof that $\max S \gtrsim 2^k$. This remains an interesting open problem.

Let us use the second moment to give a modest improvement on the earlier pigeonhole argument. The main idea here is that, by second moment, most of the subset sums lie within an $O(\sigma)$-interval, so that we can improve on the pigeonhole estimate ignoring outlier subset sums.

Theorem 4.5.2. Let S be a k-element subset of positive integers such that all 2^k subset sums of S. Then $\max S \gtrsim 2^k/\sqrt{k}$.

Proof. Let $S = \{x_1, \ldots, x_k\}$ and $n = \max S$. Set

$$X = \epsilon_1 x_1 + \cdots + \epsilon_k x_k$$

where $\epsilon_i \in \{0, 1\}$ are chosen uniformly at random independently. We have

$$\mu := \mathbb{E}X = \frac{x_1 + \cdots + x_k}{2}$$

and

$$\sigma^2 := \operatorname{Var} X = \frac{x_1^2 + \cdots + x_k^2}{4} \leq \frac{n^2 k}{4}.$$

By Chebyshev,

$$\mathbb{P}(|X - \mu| < n\sqrt{k}) \geq \frac{3}{4}.$$

Since X takes distinct values for every $(\epsilon_1, \dots, \epsilon_k) \in \{0,1\}^k$, we have $\mathbb{P}(X = x) \le 2^{-k}$ for all x, so we have the lower bound

$$\mathbb{P}(|X - \mu| < n\sqrt{k}) \le 2^{-k}(2n\sqrt{k} + 1).$$

Putting them together, we get

$$2^{-k}(2n\sqrt{k} + 1) \le \frac{3}{4}.$$

So $n \gtrsim 2^k/\sqrt{k}$. □

Recently, this July, Dubroff–Fox–Xu gave another short proof of this result (with an improved error term $O(1)$) by applying Harper's vertex-isoperimetric inequality on the cube (this is an example of "concentration of measure", which we will explore more later this course).

Here for the "n-dimensional boolean cube" we consider the graph on the vertex set $\{0,1\}^n$ with an edge between every pair of n-tuples that differ in exactly one coordinate. Given $A \subseteq \{0,1\}^n$, let δA be the set of all vertices outside A that is adjacent to some vertex of A.

Theorem 4.5.3 (Harper 1966)**.** Every $A \subset \{0,1\}^k$ with $|A| = 2^{k-1}$ has $|\partial A| \ge \binom{k}{\lfloor k/2 \rfloor}$.

Remark 4.5.4. Harper's theorem, more generally, gives the precise value of $\min_{A \subset \{0,1\}^n : |A| = m} |\partial A|$ for every (n, m). Basically, the minimum is achieved when A is a a Hamming ball (or, if m is not exactly the size of some Hamming ball, then take the first m elements of $\{0,1\}^n$ when ordered lexicographically).

Theorem 4.5.5 (Dubroff–Fox–Xu)**.** If S is a set of k positive integers with distinct subset sums, then

$$\max S \ge \binom{k}{\lfloor k/2 \rfloor} = \left(\sqrt{\frac{2}{\pi}} + o(1)\right) \frac{2^k}{\sqrt{k}}.$$

Remark 4.5.6. The above bound has the currently best known leading constant factor.

Proof. Let $S = \{x_1, \dots, x_k\}$. Let

$$A = \left\{ (\epsilon_1, \dots, \epsilon_k) \in \{0,1\}^k : \epsilon_1 x_1 + \cdots + \epsilon_1 x_k < \frac{x_1 + \cdots + x_k}{2} \right\}.$$

Note that due to the distinct sum hypothesis, one can never have $x_1 s_1 + \cdots + x_n s_n = (s_1 + \cdots + s_n)/2$. It thus follows by symmetry that $|A| = 2^{k-1}$.

Note that every element of ∂A corresponds to some subset sum in the open interval

$$\left(\frac{x_1 + \cdots + x_k}{2}, \frac{x_1 + \cdots + x_k}{2} + \max S\right)$$

Since all subset sums are distinct, we must have $\max S \geq |\partial A| \geq \binom{k}{\lfloor k/2 \rfloor}$ by Harper's theorem (Theorem 4.5.3). □

4.6 Weierstrass approximation theorem

We finish off the chapter with an application to analysis.

Weierstrass approximation theorem every continuous real function on an interval can be uniformly approximated by a polynomial.

Theorem 4.6.1 (Weierstrass approximation theorem 1885)**.** Let $f\colon [0,1] \to \mathbb{R}$ be a continuous function. Let $\epsilon > 0$. Then there is a polynomial $p(x)$ such that $|p(x) - f(x)| \leq \epsilon$ for all $x \in [0,1]$.

Proof. (Bernstein 1912) The idea is to approximate f by a sum of polynomials look like "bumps":

$$P_n(x) = \sum_{i=0}^{n} E_i(x) f(i/n)$$

where $E_j(x)$ chosen as some polynomials peaks at $x = i/n$ and then decaysaway from $x = i/n$. To this end, set set

$$E_i(x) = \mathbb{P}(\mathrm{Bin}(n,x) = i) = \binom{n}{i} x^i (1-x)^{n-i} \quad \text{for } 0 \leq i \leq n.$$

For each $x \in [0,1]$, the binomial distribution $\mathrm{Bin}(n,x)$ has mean nx and variance $nx(1-x) \leq n$. By Chebyshev's inequality,

$$\sum_{i:|i-nx|>n^{2/3}} E_i(x) = \mathbb{P}(|\mathrm{Bin}(n,x) - nx| > n^{2/3}) \leq n^{-1/3}.$$

Since $[0,1]$ is compact, f is uniformly continuous and bounded. By rescaling, assume that $|f(x)| \leq 1$ for all $x \in [0,1]$. Also there exists $\delta > 0$ such that $|f(x) - f(y)| \leq \epsilon/2$ for all $x, y \in [0,1]$ with $|x - y| \leq \delta$.

Take $n > \max\{64\epsilon^{-3}, \delta^{-3}\}$. Then for every $x \in [0,1]$ (note that $\sum_{j=0}^{n} E_j(x) = 1$),

$$\begin{aligned}
|P_n(x) - f(x)| &\leq \sum_{i=0}^{n} E_i(x)|f(i/n) - f(x)| \\
&\leq \sum_{i:|i/n-x|<n^{-1/3}<\delta} E_i(x)|f(i/n) - f(x)| + \sum_{i:|i-nx|>n^{2/3}} 2E_i(x) \\
&\leq \frac{\epsilon}{2} + 2n^{-1/3} \leq \epsilon. \square
\end{aligned}$$

5 Chernoff bound

Chernoff bounds give us much better tail bounds than the second moment method when applied to sums of independent random variables. This is one of the most useful bounds in probabilistic combinatorics.

The proof technique of bounding the exponential moments is perhaps just as important as the resulting bounds themselves. We will see this proof method come up again later on when we prove martingale concentration inequalities. The method allows us to adapt the proof of the Chernoff bound to other distributions. Let us give the proof in the most basic case for simplicity and clarity.

Theorem 5.0.1. Let $S_n = X_1 + \cdots + X_n$ where $X_i \in \{-1, 1\}$ uniformly iid. Let $\lambda > 0$. Then

$$\mathbb{P}(S_n \geq \lambda\sqrt{n}) \leq e^{-\lambda^2/2}$$

Note that in contrast, $\operatorname{Var} S_n = n$, so Chebyshev's inequality would only give a tail bound $\leq 1/\lambda^2$

Proof. Let $t \geq 0$. Consider the **moment generating function**

$$\mathbb{E}\left[e^{tS_n}\right] = \mathbb{E}\left[e^{t\sum_i X_i}\right] = \mathbb{E}\left[\prod_i e^{tX_i}\right] = \prod_i \mathbb{E}\left[e^{tX_i}\right] = \left(\frac{e^{-t} + e^t}{2}\right)^n.$$

We have (by comparing Taylor series coefficients $\frac{1}{(2n)!} \leq \frac{1}{n!2^n}$), for all $t \geq 0$,

$$\frac{e^{-t} + e^t}{2} \leq e^{t^2/2}.$$

By Markov's inequality,

$$\mathbb{P}(S_n \geq \lambda\sqrt{n}) \leq \frac{\mathbb{E}\left[e^{tS}\right]}{e^{t\lambda\sqrt{n}}} \leq e^{-t\lambda\sqrt{n} + t^2 n/2}$$

Set $t = \lambda/\sqrt{n}$ gives the bound. □

Remark 5.0.2. The technique of considering the moment generating function can be thought morally as taking an appropriately high moment. Indeed, $\mathbb{E}[e^{tS}] = \sum_{n\geq 0} \mathbb{E}[S^n]t^n/n!$ contains all the moments data of the random variable.

The second moment method (Chebyshev + Markov) can be thought of as the first iteration of this idea. By taking fourth moments (now requiring 4-wise independence of the summands), we can obtain tail bounds of the form $\lesssim \lambda^{-4}$. And similarly with higher

moments.

In some applications, where one cannot assume independence, but can estimate high moments, the above philosophy can allow us to prove good tail bounds as well.

Also by symmetry, $\mathbb{P}(S_n \leq -\lambda\sqrt{n}) \leq e^{-\lambda^2/2}$. Thus we have the following two-sided tail bound.

Corollary 5.0.3. $\mathbb{P}(|S_n| \geq \lambda\sqrt{n}) \leq 2e^{-\lambda^2/2}$

Remark 5.0.4. It is easy to adapt the above proof so that each X_i is a mean-zero random variable taking $[-1,1]$-values, and independent (but not necessarily identical) across all i. Indeed, by convexity, we have $e^{tx} \leq \frac{1-x}{2}e^{-t} + \frac{1+x}{2}e^{t}$ for all $x \in [-1,1]$ by convexity, so that $\mathbb{E}[e^{tX}] \leq \frac{e^t+e^{-t}}{2}$. In particular, we obtain the following tail bounds on the binomial distribution.

Theorem 5.0.5. Let each X_i be an independent random variable taking values in $[-1,1]$ and $\mathbb{E}X_i = 0$. Then $S_n = X_1 + \cdots + X_n$ satisfies

$$\mathbb{P}(S_n \geq \lambda\sqrt{n}) \leq e^{-\lambda^2}.$$

Corollary 5.0.6. Let X be a sum of n independent Bernoulli's (not necessarily the same probability). Let $\mu = \mathbb{E}X$ and $\lambda > 0$. Then Then

$$\mathbb{P}(X \geq \mu + \lambda\sqrt{n}) \leq e^{-\lambda^2/2} \quad \text{and} \quad \mathbb{P}(X \leq \mu - \lambda\sqrt{n}) \leq e^{-\lambda^2/2}$$

The quality the Chernoff compares well to that of the normal distribution. For the standard normal $Z \sim N(0,1)$, one has $\mathbb{E}[e^{tZ}] = e^{t^2/2}$ and so

$$\mathbb{P}(Z \geq \lambda) = \mathbb{P}(e^{tZ} \geq e^{t\lambda}) \leq e^{-t\lambda}\mathbb{E}[e^{tX}] = e^{-t\lambda+t^2/2}$$

Set $t = \lambda$ and get

$$\mathbb{P}(Z \geq \lambda) \leq e^{-\lambda^2/2}$$

And this is actually pretty tight, as, for $\lambda \to \infty$,

$$\mathbb{P}(Z \geq \lambda) = \frac{1}{\sqrt{2\pi}} \int_\lambda^\infty e^{-t^2/2}\, dt \sim \frac{e^{-\lambda^2/2}}{\sqrt{2\pi\lambda}}$$

The same proof method allows you to prove bounds for other sums of random variables, suitable for whatever application you have in mind. See Alon–Spencer Appendix A for some calculations.

For example, for a sum of independent Bernoulli's with small means, we can improve on the above estimates as follows

Theorem 5.0.7. Let X be the sum of independent Bernoulli random variables (not necessarily same probability). Let $\mu = \mathbb{E}X$. For all $\epsilon > 0$,

$$\mathbb{P}(X \geq (1+\epsilon)\mu) \leq e^{-((1+\epsilon)\log(1+\epsilon)-\epsilon)\mu} \leq e^{-\frac{\epsilon^2}{1+\epsilon}\mu}$$

and

$$\mathbb{P}(X \leq (1-\epsilon)\mu) \leq e^{-\epsilon^2\mu/2}.$$

Remark 5.0.8. The bounds for upper and lower tails are necessarily asymmetric, when the probabilities are small. Why? Think about what happens when $X \sim \text{Bin}(n, c/n)$, which, for a constant $c > 0$, converges as $n \to \infty$ to a Poisson distribution with mean c, whose value at k is $c^k e^k/k! = e^{-\Theta(k \log k)}$ and not $e^{-\Omega(k^2)}$ as one might naively predict by an incorrect application of the Chernoff bound formula.

Nonetheless, both formulas tell us that both tails exponentially decay like ϵ^2 for small values of ϵ, say, $\epsilon \in [0, 1]$.

5.1 Discrepancy

Theorem 5.1.1. Let $\mathcal{F}$ be a collection of m subsets of $[n]$. Then there exists some assignment $[n] \to \{-1, 1\}$ so that the sum on every set in $\mathcal{F}$ is at most $2\sqrt{n \log m}$ in absolute value.

Proof. Put ± 1 iid uniformly at random on each vertex. On each edge, the probability that the sum exceeds $2\sqrt{n \log m}$ in absolute value is, by Chernoff bound, less than $2e^{-2\log m} = 2/m^2$. By union bound over all m edges, with probability greater than $1 - 2/m \geq 0$, no edge has sum exceeding $2\sqrt{n \log m}$. □

Remark 5.1.2. In a beautiful landmark paper titled *Six standard deviations suffice*, Spencer (1985) showed that one can remove the logarithmic term by a more sophisticated semi-random assignment algorithm.

Theorem 5.1.3 (Spencer (1985))**.** Let $\mathcal{F}$ be a collection of n subsets of $[n]$. Then there exists some assignment $[n] \to \{-1, 1\}$ so that the sum on every set in $\mathcal{F}$ is at most $6\sqrt{n}$ in absolute value.
More generally, if $\mathcal{F}$ be a collection of $m \geq n$ subsets of $[n]$, then we can replace $6\sqrt{n}$ by $11\sqrt{n \log(2m/n)}$.

Remark 5.1.4. More generally, Spencer proves that the same holds if vertices have $[0,1]$-valued weights.

The idea, very roughly speaking, is to first generalize from $\{-1,1\}$-valued assignments to $[-1,1]$-valued assignments. Then the all-zero vector is a trivially satisfying assignment. We then randomly, in iterations, alter the values from 0 to other values in $[-1,1]$, while avoiding potential violations (e.g., edges with sum close to $6\sqrt{n}$ in absolute value), and finalizing a color of a color when its value moves to either -1 and 1.

Spencer's original proof was not algorithmic, and he suspected that it could not be made efficiently algorithmic. In a breakthrough result, Bansal (2010) gave an efficient algorithm for producing a coloring with small discrepancy. Another very nice algorithm with another beautiful proof of the algorithmic result was given by Lovett and Meka (2015).

Here is a famous conjecture on discrepancy.

Conjecture 5.1.5 (Komlós)**.** There exists some absolute constant K so that for every set of vectors $v_1,\ldots,v_m$ in the unit ball in R^n, there exists signs $\epsilon_1,\ldots,\epsilon_m \in \{-1,1\}$ such that

$$\epsilon_1 v_1 + \cdots + \epsilon_m v_m \in [-K,K]^n.$$

Banaszczyk (1998) proved the bound $K = O(\sqrt{\log n})$ in a beautiful paper using deep ideas from convex geometry.

Spencer's theorem's implies the Komlós conjecture if all vectors v_i have the form $n^{-1/2}(\pm 1,\ldots,\pm 1)$ (or more generally when all coordinates are $O(n^{-1/2})$). The deduction is easy when $m \le n$. When $m > n$, we use the following observation.

Lemma 5.1.6. Let $v_1,\ldots,v_m \in \mathbb{R}^n$. Then there exists $a_1,\ldots,a_m \in [-1,1]^m$ with $|\{i : a_i \notin \{-1,1\}\}| \le n$ such that

$$a_1 v_1 + \cdots + a_m v_m = 0$$

Proof. Find $(a_1,\ldots,a_m) \in [-1,1]^m$ satisfying and as many $a_i \in \{-1,1\}$ as possible. Let $I = \{i : a_i \notin \{-1,1\}\}$. If $|I| > n$, then we can find some nontrivial linear combination of the vectors $v_i, i \in I$, allowing us to to move $(a_i)_{i\in I}$'s to new values, while preserving $a_1 v_1 + \cdots + a_m v_m = 0$, and end up with at one additional a_i taking $\{-1,1\}$-value. □

Letting $a_1,\ldots,a_m$ and $I = \{i : a_i \notin \{-1,1\}\}$ as in the Lemma, we then take $\epsilon_i = a_i$ for all $i \notin I$, and apply a corollary of Spencer's theorem to find $\epsilon_i \in \{-1,1\}^n$, $i \in I$ with

$$\sum_{i\in I} (\epsilon_i - a_i) v_i \in [-K,K]^n,$$

which would yield the desired result. The above step can be deduced from Spencer's theorem by first assuming that each $a_i \in [-1, 1]$ has finite binary length (a compactness argument), and then rounding it off one digit at a time during Spencer's theorem, starting from the least significant bit (see Corollary 8 in Spencer's paper for details).

5.2 Hajós conjecture counterexample

We begin by reviewing some classic result from graph theory. Recall some definitions:

- H is an **induced subgraph** of G if H can be obtained from G by removing vertices;
- H is a **subgraph** if G if H can be obtained from G by removing vertices and edges;
- H is a **subdivision** of G if H can be obtained from a subgraph of G by contracting induced paths to edges;
- H is a **minor** of G if H can be obtained from a subgraph of G by by contracting edges to vertices.

Kuratowski's theorem (1930). Every graph without $K_{3,3}$ and K_5 as subdivisions as subdivision is planar.

Wagner's theorem (1937). Every graph free of $K_{3,3}$ and K_5 as minors is planar.

(There is a short argument shows that Kuratowski and Wagner's theorems are equivalent.)

Four color theorem (Appel and Haken 1977) Every planar graph is 4-colorable.

Corollary: Every graph without $K_{3,3}$ and K_5 as minors is 4-colorable.

The condition on K_5 is clearly necessary, but what about $K_{3,3}$? What is the "real" reason for 4-colorability.

Hadwidger posed the following conjecture, which is one of the biggest open conjectures in graph theory.

Conjecture 5.2.1 (Hadwiger 1936)**.** For every $t \geq 1$, every graph without a K_{t+1} minor is t-colorable.

$t = 1$ trivial

$t = 2$ nearly trivial (if G is K_3-minor-free, then it's a tree)

$t = 3$ elementary graph theoretic arguments

$t = 4$ is equivalent to the 4-color theorem (Wagner 1937)

$t = 5$ is equivalent to the 4-color theorem (Robertson–Seymour–Thomas 1994; this work won a Fulkerson Prize)

$t \geq 6$ remains open

Let us explore a variation of Hadwiger's conjecture:

Hajós conjecture. (1961) Every graph without a K_{t+1}-subdivision is t-colorable.

Hajós conjecture is true for $t \leq 3$. However, it turns out to be found in general. Catlin (1979) constructed counterexamples for all $t \geq 6$ ($t = 4, 5$ are still open).

It turns out that Hajós conjecture is not just false, but very false.

Erdős–Fajtlowicz (1981) showed that almost every graph is a counterexample (it's a good idea to check for potential counterexamples among random graphs!)

To be continued

Theorem 5.2.2. With probability $1 - o(1)$, $G(n, 1/2)$ has no K_t-subdivision with $t = \lceil 10\sqrt{n} \rceil$.

From Theorem 4.3.3 we show that, with high probability, $G(n, 1/2)$ has independence number $\sim 2 \log_2 n$ and hence chromatic number $\geq (1 + o(1)) \frac{n}{2 \log_2 n}$. Thus the above result shows that $G(n, 1/2)$ is whp a counterexample to Hajós conjecture.

Proof. If G had a K_t-subdivision, say with $S \subset V$, $|S| = t$, then at most $n - t \leq n$ of the edges in the subdivision can be paths with at least two edges (since they must use distinct vertices outside S). So S must induce at least $\binom{t}{2} - n \geq \frac{3}{4}\binom{t}{2}$ edges in G.

By Chernoff bound, for fixed t-vertex subset S

$$\mathbb{P}\left(e(S) \geq \frac{3}{4}\binom{t}{2}\right) \leq e^{-t^2/10}.$$

Taking a union bound over all t-vertex subsets S, and noting that

$$\binom{n}{t} e^{-t^2/10} < n^t e^{-t^2/10} \leq e^{-10n + O(\sqrt{n} \log n)} = o(1)$$

we see that whp no such S exists, so that this $G(n, 1/2)$ whp has no K_t-subdivision □

Remark 5.2.3. One can ask the following quantitative question regarding Hadwidger's conjecture:

Can we show that every graph without a K_{t+1}-minor can be properly colored with a small number of colors?

Wagner (1964) showed that every graph without K_{t+1}-minor is 2^{t-1} colorable.

Here is the proof: assume that the graph is connected. Take a vertex v and let L_i be the set of vertices with distance exactly i from v. The subgraph induced on L_i has no K_t-minor, since otherwise such a K_t-minor would extend to a K_{t+1}-minor with v. Then by induction L_i is 2^{t-2}-colorable (check base cases), and using alternating colors for even and odd layers L_i yields a proper coloring of G.

This bound has been improved over time. The best current bound was proved this past summer. Postle (2020+) showed that if every graph with no K_t-minor is $O(t(\log \log t)^6)$-colorable.

For more on Hadwiger's conjecture, see Seymour's survey (2016).

6 Lovász local lemma

The Lovász local lemma (LLL), introduced in the paper of Erdős and Lovász (1975) is a powerful tool in the probabilistic method. It is some form of interpolation between the following two extreme (easy) scenerios

- Complete independence: if we have an arbitrary number of independent bad events, each occurring with probability < 1, then it is possible to avoid all of them (although with tiny probability)
- Union bound: if we have a collection of bad events whose total probability is < 1 (but usually much smaller), then it is possible to avoid all of them (often with high probability)

The local lemma deals with the case when each bad event is independent with most other bad events, but possibly dependent with a small number of other events.

We saw an application of the Lovász local lemma back in Section 1.1, where we used it to lower bound Ramsey numbers. This chapter we will explore the local lemma and its applications in depth.

6.1 Statement and proof

Here is the **setup** for the local lemma:

- We have "bad events" $A_1, A_2, \ldots, A_n$
- For each i there is some subset $N(i) \subseteq [n]$ such that A_i is independent from $\{A_j : j \notin N(i) \cup \{i\}\}$.

Here we say that event A_0 is **independent** from $\{A_1, \ldots, A_m\}$ if A_0 is independent of every event of the form $B_1 \wedge \cdots \wedge B_m$ where each B_i is either A_i or $\overline{A}_i$, i.e.,

$$\mathbb{P}(A_0 B_1 \cdots B_m) = \mathbb{P}(A_0)\mathbb{P}(B_1 \cdots B_m),$$

or, equivalently, using Bayes's rule: $\mathbb{P}(A_0 | B_1 \cdots B_m) = \mathbb{P}(A_0)$. (Here $\wedge$ = 'and' and $\vee$ = 'or', and we may omit $\wedge$ symbols, similar to multiplication)

We can represent the above relations by a **dependency (di)graph** whose vertices are indexed by the events (or equivalently $V = [n]$), and the (out-)neighbors of i are $N(i)$. (Mostly we'll just work with undirected dependency graphs for simplicity, but in general it may be helpful to think of them as directed—hence digraphs.)

Remark 6.1.1 (Important!). **Independence $\neq$ pairwise independence**
The dependency graph is *not* made by joining $i \sim j$ whenever A_i and A_j are not independent (i.e., $\mathbb{P}(A_iA_j) \neq \mathbb{P}(A_i)\mathbb{P}(A_j)$).

Example: suppose one picks $x_1, x_2, x_3 \in \mathbb{Z}/2\mathbb{Z}$ uniformly and independently at random and set, for each $i = 1, 2, 3$ (indices taken mod 3), A_i the event that $x_{i+1} + x_{i+2} = 0$. Then these events are pairwise independent but not independent. So the empty graph on three vertices is not a valid dependency graph (on the other hand, having at least two edges makes it a valid dependency graph).

A related note: there could be more than one choices for dependency graphs. So we speak of "a dependency graph" instead of "the dependency graph."

Remark 6.1.2 (**Random variable model** / **hypergraph coloring**). Many common applications of the local lemma can be phrased in the following form:

- A collection of independent random variables $x_1, \ldots, x_N$
- Each event A_i only depends on $\{x_j : j \in S_i\}$ for some subset $S_i \subseteq [N]$

In this case, valid dependency graph can be formed by placing an edge $i \sim j$ whenever $S_i \cap S_j \neq \emptyset$.

We can also view the above as coloring a hypergraph with vertices labeled by $[N]$, using independent random colors $x_1, \ldots, x_N$ for each vertex, so that various constraints on edges $S_1, S_2, \cdots \subseteq [N]$ are satisfied.

An example of such a problem is the **satisfiability problem (SAT)**: given a **CNF formula** (conjunctive normal norm = *and-of-or*'s), e.g.,

$$(x_1 \vee x_2 \vee x_3) \wedge (\overline{x_1} \vee x_2 \vee x_4) \wedge (\overline{x_2} \vee x_4 \vee x_5) \wedge \cdots$$

the problem is to find a satisfying assignment with boolean variables $x_1, x_2, \ldots$. Many problems in computer science can be modeled using this way.

The following formulation of the local lemma is easiest to apply and is the most commonly used.

Theorem 6.1.3 (Lovász local lemma; symmetric form)**.** Let $A_1, \ldots, A_n$ be events, with $\mathbb{P}[A_i] \leq p$ for all i. Suppose that each A_i is independent from a set of all other A_j except for at most d of them. If

$$ep(d + 1) \leq 1,$$

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

Remark 6.1.4. The constant e is best possible (Shearer 1985).

Theorem 6.1.5 (Lovász local lemma; general form)**.** Let $A_1, \dots, A_n$ be events. For each $i \in [n]$, let $N(i)$ be such that A_i is independent from $\{A_j : j \notin \{i\} \cup N(i)\}$. If $x_1, \dots, x_n \in [0, 1)$ satisfy

$$\mathbb{P}(A_i) \le x_i \prod_{j \in N(i)} (1 - x_j) \quad \forall i \in [n],$$

then with probability $\ge \prod_{i=1}^{n}(1 - x_i)$, none of the events A_i occurs.

Proof that the general form implies the symmetric form. Set $x_i = 1/(d+1) < 1$ for all i. Then

$$x_i \prod_{j \in N(i)} (1 - x_j) \ge \frac{1}{d+1}\left(1 - \frac{1}{d+1}\right)^d > \frac{1}{(d+1)e} \ge p$$

so the hypothesis of general local lemma holds. □

Here is another corollary of the general form. It says that the local lemma works if the total probability of any neighborhood in a dependency graph is small.

Corollary 6.1.6. In the setup of the Lovász local lemma, if $\mathbb{P}(A_i) < 1/2$ and $\sum_{j \in N(i)} \mathbb{P}(A_j) \le 1/4$ for all i, then with positive probability none of the events A_i occurs.

Proof. In Theorem 6.1.5, set $x_i = 2\mathbb{P}(A_i)$ for each i. Then

$$x_i \prod_{j \in N(i)} (1 - x_j) \ge x_i \left(1 - \sum_{j \in N(i)} x_j\right) = 2\mathbb{P}(A_i)\left(1 - \sum_{j \in N(i)} 2\mathbb{P}(A_i)\right) \le \mathbb{P}(A_i).$$

(The first inequality is by "union bound.") □

Proof of Lovász local lemma (general case). We will prove that

$$\mathbb{P}\left(A_i \,\middle|\, \bigwedge_{j \in S} \overline{A}_j\right) \le x_i \quad \text{whenever } i \notin S \subseteq [n] \tag{6.1}$$

Once (6.1) has been established, we then deduce that

$$\begin{aligned}\mathbb{P}(\overline{A}_1 \cdots \overline{A}_n) &= \mathbb{P}(\overline{A}_1)\mathbb{P}\left(\overline{A}_2 \mid \overline{A}_1\right)\mathbb{P}\left(\overline{A}_3 \mid \overline{A}_1\overline{A}_2\right)\cdots\mathbb{P}\left(\overline{A}_n \mid \overline{A}_1 \cdots \overline{A}_{n-1}\right)\\ &\ge (1 - x_1)(1 - x_2)\cdots(1 - x_n),\end{aligned}$$

which is the conclusion of the local lemma.

Now we prove (6.1) by induction on $|S|$. The base case $|S| = 0$ is trivial.

Let $i \notin S$. Let $S_1 = S \cap N(i)$ and $S_2 = S \setminus S_1$. We have

$$\mathbb{P}\left(A_i \,\middle|\, \bigwedge_{j \in S} \overline{A}_j\right) = \frac{\mathbb{P}\left(A_i \bigwedge_{j \in S_1} \overline{A}_j \,\middle|\, \bigwedge_{j \in S_2} \overline{A}_j\right)}{\mathbb{P}\left(\bigwedge_{j \in S_1} \overline{A}_j \,\middle|\, \bigwedge_{j \in S_2} \overline{A}_j\right)} \tag{6.2}$$

For the RHS of (6.2),

$$\text{numerator} \le \mathbb{P}\left(A_i \,\middle|\, \bigwedge_{j \in S_2} \overline{A}_j\right) = \mathbb{P}(A_i) \le x_i \prod_{j \in N(i)} (1 - x_i)$$

and, writting $S_1 = \{j_1, \dots, j_r\}$,

$$\begin{aligned}
\text{denominator} &= \mathbb{P}\left(\overline{A}_{j_1} \,\middle|\, \bigwedge_{j \in S_2} \overline{A}_j\right) \mathbb{P}\left(\overline{A}_{j_2} \,\middle|\, \overline{A}_{j_1} \bigwedge_{j \in S_2} \overline{A}_j\right) \cdots \mathbb{P}\left(\overline{A}_{j_r} \,\middle|\, \overline{A}_{j_1} \cdots \overline{A}_{j_{r-1}} \bigwedge_{j \in S_2} \overline{A}_j\right) \\
&\ge (1 - x_{j_1}) \cdots (1 - x_{j_r}) \qquad \text{[by induction hypothesis]} \\
&\ge \prod_{j \in N(i)} (1 - x_i)
\end{aligned}$$

Thus (6.2) $\le x_i$, thereby finishing the induction proof of (6.1). □

6.2 Algorithmic local lemma

The local lemma tells you that some good configuration exists, but the proof is non-constructive. The probability that a random sample avoids all the bad events is often very small (usually exponentially small, e.g., in the case of a set of independent bad events). It had been an open problem for a long time whether there exists some efficient algorithm to sample a good configuration in applications of the local lemma.

Moser (2009), during his PhD, achieved a breakthrough by coming up with the first efficient algorithmic version of the local lemma. Later, in a beautiful paper by Moser and Tardos (2010) extended the algorithm to a general framework for the local lemma.

The Moser–Tardos algorithm considers problems in the random variable model (Re-

mark 6.1.2). The algorithm is surprisingly simple.

Algorithm: Moser–Tardos local lemma algorithm

Initialize all the random variables;
while *there are violated events* **do**
    Pick an arbitrary violated event and resample its variables;

Theorem 6.2.1 (Moser and Tardos 2010)**.** If there are $x_1, \dots, x_n \in [0, 1)$ such that

$$\mathbb{P}(A_i) \le x_i \prod_{j \in N(i)} (1 - x_j) \quad \forall i \in [n],$$

then the above randomized algorithm resamples each A_i at most $x_i/(1 - x_i)$ times in expectation for each i.

Remark 6.2.2. The above theorem shows that the Moser–Tardos algorithm is an *Las Vegas* algorithm with polynomial expected runtime. A Las Vegas algorithm is a randomized algorithm that always terminates a successful result, but it might take a long time to terminate. Contrast this to a *Monte Carlo* algorithm, which runs in bounded time but may return a bad result with some small probability, and there may not be an efficient way to check whether the output is correct–e.g., randomly 2-coloring the edges of K_n to avoid a monochromatic $2 \log_2 n$-clique. A Las Vegas algorithm can be converted into a Monte Carlo algorithm by cutting off the algorithm after some time (significantly larger than the expected running time) and applying Markov's inequality to bound the probability of failure. On the other hand, there is in general no way to convert a Monte Carlo algorithm to a Las Vegas algorithm unless there is an efficient way to certify the correctness of the output of the algorithm.

Remark 6.2.3. The Moser–Tardos algorithm assumes the random variable model. Some assumption on the model is necessary since the problem can be computationally hard in general.

For example, let $q = 2^k$, and $f \colon [q] \to [q]$ be some fixed bijection. Let $y \in [q]$ be given. The goal is find x such that $f(x) = y$.

For each $i \in [k]$, let A_i be the event that $f(x)$ and y disagree on i-th bit. Then $A_1, \dots, A_k$ independent (check!). Also, $f(x) = y$ if and only if no event A_i occurs.

A trivial version of the local lemma (with empty dependency graph) guarantees the existence of some x such that $f(x) = y$.

However, finding x may be computationally hard for certain functions f. In fact, the existence of such one-way functions (easy to compute but hard to invert) is the bedrock of cryptography. A concrete example is $f \colon \mathbb{F}_q \to \mathbb{F}_q$ is given by $f(0) = 0$, and for $x \neq 0$, set

$f(x) = g^x$ for some multiplicative generator. Then inverting f is the **discrete logarithm problem**, which is believed to be computationally difficult.