21-701: Discrete Mathematics

Lecturer: Wes Pegden

Scribe: Sidharth Hariharan

Carnegie Mellon University - Fall 2026

This version of the notes was compiled on September 16, 2026.

For the latest version of these notes, visit

https://thefundamentaltheor3m.github.io/DiscreteMathNotes/main.pdf.

For comments, errors or suggestions, please raise an issue/make a pull request at

https://github.com/thefundamentaltheor3m/DiscreteMathNotes.

Contents

Chapter 1

Probability

Almost every argument in these notes produces an object by showing that a random one works, so the vocabulary for talking about random objects had better come first. This chapter collects it. Very little of what follows is new—probability spaces, expectation, variance—and none of it is proved at length; what matters is that the two inequalities the rest of the course uses by name, Markov's and Chebyshev's, are written down somewhere rather than assumed. Conditional expectation gets a little more care than the rest, because Chapter 4 leans on it hard enough to notice.

The course itself begins in Chapter 2, and a reader who would rather see a graph than a sigma-algebra can start there and come back when a probability turns up.

1.1 Probability Spaces

Almost everything in these notes is probabilistic, so it is worth setting down what we are going to use before we use it. None of this section goes beyond a first course, and we will take the measure theory on trust.

Recall the following: if $(\Omega, \mathcal{F}, \Pr)$ is a probability space, this means

- Ω is a set of outcomes, called a **sample space**;

- $\mathcal{F}$ is a sigma-algebra on Ω;
- $\Pr : \mathcal{F} \to [0, \infty)$ is a measure that only takes values in $[0, 1]$, with $\Pr(\Omega) = 1$.

An **event** is a measurable set (ie, an element of $\mathcal{F}$), and a random variable is a measurable $\Omega \to \mathbb{R}$ function.

We say two events A and B are **independent** if $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$. We say that random variables $X_1, \ldots, X_n$ are independent if for all Borel sets $A_1, \ldots, A_n \subseteq \mathbb{R}$ the events $X_1^{-1}[A_1], \ldots, X_n^{-1}[A_n]$ (all of which lie in $\mathcal{F}$ because the X_i are measurable) satisfy $\Pr\big(X_1^{-1}[A_1] \cap \cdots \cap X_n^{-1}[A_$ $\Pr\big(X_1^{-1}[A_1]\big) \cdots \Pr(X_n^{-1}[A_n])$, and an infinite family is independent if every finite subfamily is.

In this course, Ω will usually be finite, $\mathcal{F}$ will usually be the $\mathcal{P}(\Omega)$, and all functions $\Omega \to \mathbb{R}$ will be random variables.

1.2 Expectation

The expectation is the integral of a random variable over the probability space with respect to the probability measure.

Two things stand out about independent random variables.

1. If $\eta_1, \eta_2, \eta_3, \ldots$ are independent and $f, g, h, \ldots$ are measurable functions, then $f(\eta_1), g(\eta_2), h(\eta_3, \eta_4), f($ are all independent.
2. If $\eta_1, \ldots, \eta_n$ are independent, then
$$\mathbb{E}\left(\prod_{i=1}^{n} \eta_i\right) = \prod_{i=1}^{n} \mathbb{E}(\eta_i)$$

1.2.1 Conditional Expectation

Often we learn something partway through and want the expectation of what is left. That is an expectation again, taken over a smaller sample space.

Given an event A with $\Pr(A) \neq 0$ and a random variable X, we can define

$$\mathbb{E}(X \mid A) = \frac{1}{\Pr(A)} \sum_{x \in A} X(x) \cdot p(x)$$

where $p(x)$ is the density function. It is easy to show that

$$\mathbb{E}(X \mid A) = \sum_{r \in X[\Omega]} r \cdot \Pr(X = r \mid A)$$

Given *two* random variables X and Y, we denote by $\mathbb{E}(X \mid Y)$ the **random variable**

$$\omega \mapsto \mathbb{E}(X \mid Y = Y(\omega))$$

The point of packaging it as a random variable rather than as a list of numbers is that it can then be averaged again.

Theorem 1.2.1 (Tower Property of Conditional Expectation). *For all random variables X and Y,*

$$\mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}(X)$$

The above type-checks because $\mathbb{E}(X \mid Y)$ is a random variable.

Several random variables at once are no different: $\mathbb{E}(X \mid Y_1, \ldots, Y_k)$ is the random variable $\omega \mapsto \mathbb{E}(X \mid Y_1 = Y_1(\omega), \ldots, Y_k = Y_k(\omega))$, conditioning on the one event that records what all of them did. Both this and $\mathbb{E}(X \mid Y)$ need each of those events to have nonzero probability, which on a finite Ω whose outcomes all have positive probability—the only case we use—every ω supplies.

Section 4.3 leans on all of this rather harder than anything before it does.

1.3 Deviation from the Mean

Knowing what a random variable averages to is not the same as knowing what it does, and most of what we want from a random object is of the second kind. Two inequalities get us most of the way, and the whole of Chapter 4 is what to do when they do not.

The first says that a nonnegative random variable cannot often be much larger than its mean.

Theorem 1.3.1 (Markov's Inequality). *Let X be a nonnegative random variable and let $a > 0$. Then,*

$$\Pr(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Proof. Let A be the event that $X \geq a$. Because X is nonnegative, $X \geq a\mathbf{1}_A$ everywhere on Ω: off A the right-hand side is 0, and on A it is a, which is at most X. Taking expectations gives $\mathbb{E}(X) \geq a \Pr(A)$. □

Applied to X itself this is usually all we want, but applied to some other nonnegative random variable built out of X it is much sharper. The standard choice is the squared deviation.

Definition 1.3.2 (Variance). The **variance** of a random variable X is

$$\mathrm{Var}(X) = \mathbb{E}\Big((X - \mathbb{E}(X))^2\Big) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2$$

Variance is not linear, but it is additive on independent random variables: if $X_1, \ldots, X_n$ are independent then $\mathrm{Var}(\sum_i X_i) = \sum_i \mathrm{Var}(X_i)$, because every cross term $\mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\,\mathbb{E}(X_j)$ with $i \neq j$ vanishes.

Feeding the squared deviation into Theorem 1.3.1 gives Chebyshev's inequality.

Theorem 1.3.3 (Chebyshev's Inequality). *Let X be a random variable and let $a > 0$. Then,*

$$\Pr(|X - \mathbb{E}(X)| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}$$

Proof. The random variable $Y = (X - \mathbb{E}(X))^2$ is nonnegative and has $\mathbb{E}(Y) = \mathrm{Var}(X)$, and $|X - \mathbb{E}(X)| \geq a$ exactly when $Y \geq a^2$. So Theorem 1.3.1, applied to Y at a^2, gives what we want. □

The reason this is worth having is that it says something the expectation cannot. Knowing $\mathbb{E}(X)$ is large tells us nothing about whether X is ever nonzero; knowing in addition that X does not stray far from $\mathbb{E}(X)$ does.

Corollary 1.3.4. *Let $X_1, X_2, \ldots$ be random variables with $\mathbb{E}(X_n) > 0$ for every n. If $\frac{\mathrm{Var}(X_n)}{\mathbb{E}(X_n)^2} \to 0$, then $\Pr(X_n \neq 0) \to 1$.*

Proof. If $X_n = 0$ then $|X_n - \mathbb{E}(X_n)| = \mathbb{E}(X_n)$, so $\Pr(X_n = 0) \leq \Pr(|X_n - \mathbb{E}(X_n)| \geq \mathbb{E}(X_n))$, which is at most $\frac{\mathrm{Var}(X_n)}{\mathbb{E}(X_n)^2}$ by Theorem 1.3.3. $\square$

Here is the smallest example worth doing, and the only one in which we will get a good bound out of Chebyshev alone.

Example 1.3.5. Flip n fair coins and let X be the number of heads, so that X is a sum of n independent random variables each taking the value 1 or 0 with probability $\frac{1}{2}$. Each of them has mean $\frac{1}{2}$ and variance $\frac{1}{4}$, so $\mathbb{E}(X) = \frac{n}{2}$ and, by additivity, $\mathrm{Var}(X) = \frac{n}{4}$. Theorem 1.3.3 at $a = \varepsilon n$ then gives

$$\Pr\left(\left|X - \frac{n}{2}\right| \geq \varepsilon n\right) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2 n^2} = \frac{1}{4\varepsilon^2 n}$$

for every $\varepsilon > 0$, which tends to 0 as $n \to \infty$. So all but a vanishing proportion of the outcomes have between $\left(\frac{1}{2} - \varepsilon\right) n$ and $\left(\frac{1}{2} + \varepsilon\right) n$ heads.

With this we end our terse review.

Chapter 2

Graphs and Colorings

Much of this course will be about graphs, and a surprising amount of what we want to know about a graph can be phrased as a question about coloring it. We begin with the vocabulary, and then settle the first such question completely: which graphs admit a coloring with only two colors. We then ask the same question of hypergraphs, where it becomes considerably harder—even pinning down how many edges a 3-uniform hypergraph needs before two colors are no longer enough takes real work. Turning the question around—how large must a complete graph be before every coloring of its edges is forced to produce something monochromatic—gives us the Ramsey numbers. We then come back to coloring graphs, now with as many colors as it takes, and find that the number needed can be forced up as far as we like without ever creating a short cycle. The hypergraphs get a chapter of their own after that.

Notation.

- $[k]$ will denote $\{1, \ldots, k\}$.
- $\sim$ will denote the adjacency relation on graph vertices.
- K_n will denote the complete graph on n vertices.
- $\mathcal{H}$ will denote the set of hyper-edges of a hypergraph H, and likewise for any other letter; calligraphic letters will denote families of sets more generally.

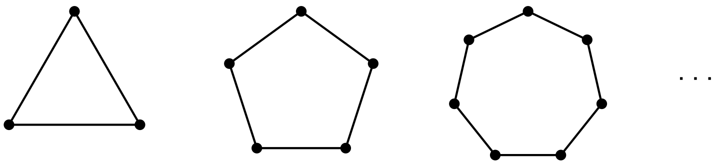
2.1 Colorings

Definition 2.1.1 (Coloring). Given a graph $G = (V, E)$, a k-coloring of G is a map $c : V \to [k]$ such that $u \sim v \implies c(u) \neq c(v)$.

2.1.1 2-Colorings of Graphs

Let's start with a motivating question: which graphs can be 2-colored?

Any odd cycle graph can't be 2-colored. That is, any graph that looks like



Let's study when odd cycles exist.

Lemma 2.1.2. *If G has an odd closed walk, then G has an odd cycle.*

Recall that a closed walk need not be a cycle: walks can contain repetitions. So even line graphs can have closed walks (start from one vertex, go to its neighbor, return).

Proof. Suppose G has an odd closed walk. Let W be such a walk of minimal length. Write $W = (x_1, \dots, x_i, \dots, x_j, \dots, x_\ell)$, with $x_\ell = x_1$. Suppose W is not a cycle. Then there must be some x_i, x_j such that $x_i = x_j$ with $1 \leq i < j \leq \ell - 1$, ie, there must be some repetition beyond the closing one. Then, $(x_1, \dots, x_i, x_{j+1}, \dots, x_\ell)$ and $(x_i, x_{i+1}, \dots, x_j)$ are both shorter walks than W. Moreover, one of them must be of odd length. This contradicts minimality of the length of W. So W must be a cycle. $\square$

Here's a cool result.

Theorem 2.1.3. *A graph $G = (V, E)$ can be 2-colored if and only if G has no odd cycle.*

Proof.

($\Longrightarrow$) We've already seen that odd cycles can't be 2-colored. So if G can be 2-colored, it can't have an odd cycle. We've essentially seen this from the examples above.

($\Longleftarrow$) We may assume G is connected, since otherwise we can color each component separately. Fix some $v_0 \in V$. Define $c : V \to \{1, 2\}$ by

$$c(u) = \begin{cases} 2 & \text{if } \operatorname{dist}(v_0, u) \text{ is odd} \\ 1 & \text{if } \operatorname{dist}(v_0, u) \text{ is even} \end{cases}$$

where $\operatorname{dist}(\cdot, \cdot)$ denotes the shortest number of edges from one vertex to another.

We claim that this c is a valid 2-coloring.

Indeed, suppose it is not. Then, there must be some $u, v \in V$ such that $u \sim v$ and $c(u) = c(v)$. Now let's consider the two possible cases on the color.

Suppose $c(u) = c(v) = 1$ for adjacent u, v. Then, there is a path of even length between v and v_0 and another path of even length between u and v_0. Concatenating the path from v to v_0 with the one from v_0 to u, and closing up with the edge uv, we get a closed walk of length even + even +1. This is odd, so by Theorem 2.1.2, G has an odd cycle.

The case $c(u) = c(v) = 2$ is identical: the two paths are now of odd length, so the closed walk they form with the edge uv has length odd + odd +1, which is odd again.

Either way, G has an odd cycle, contrary to hypothesis. So c must be a valid 2-coloring.

$\square$

2.1.2 Colorings of Hypergraphs

Coloring is not really a question about graphs. All the definition above asks of G is that we know which pairs of vertices may not share a color, and nothing stops us forbidding larger sets instead.

Definition 2.1.4 (Hypergraph). Given a vertex set V, a **hypergraph** is a pair $H = (V, \mathcal{H})$, where $\mathcal{H}$ is a collection of subsets of V. The sets in $\mathcal{H}$ are called **hyper-edges**.

Definition 2.1.5 (Uniformity). A hypergraph H is k-**uniform** if $|f| = k$ for all $f \in \mathcal{H}$.

The following is thus obvious.

Proposition 2.1.6. *A graph is a 2-uniform hypergraph.*

Definition 2.1.7. A proper coloring of a hypergraph H is an assignment $c : V \to [r]$ such that $\forall f \in \mathcal{H}$, $|c(f)| \geq 2$.

We say "no edge is monochromatic".

We can ask ourselves the following question: **"Which 3-uniform hypergraphs are 2-colorable?**

For each k, define $m(k)$ to be the smallest m such that $\exists$ some k-uniform hypergraph with m edges that is not 2-colorable. We can show that $m(2) = 3$. But what is $m(3)$? That is, how many 3-edges do I need to prevent 2-colorability?

2.1.3 The Value of $m(3)$

Let H be a 3-uniform hypergraph that's not 2-colorable. Assume, further, that H has at most 6 edges (each being a subset of size 3).

First, we'll show that we can reduce, without loss of generality, to the case where H has exactly 6 vertices. Indeed, if H has vertices x, y that are not in any common edge, then I can define H' by "merging" x and y. So H' has one fewer vertex than H. This H' cannot be 2-colorable: if it were, then that coloring would extend to H by using the color of the merged vertex at both x and y, which would make H 2-colorable. And when H has more than 6 vertices, such a pair always exists.

An edge has 3 vertices, so it covers the $\binom{3}{2} = 3$ pairs among them, and H has at most 6 edges, so at most

$$6 \cdot \binom{3}{2} = 18$$

pairs are covered in total. But if H has $n \geq 7$ vertices, there are $\binom{n}{2} \geq \binom{7}{2} = 21$ pairs to cover. So some pair is left uncovered, and merging it brings us one vertex closer to 6.

So now assume that H has exactly 6 edges and 6 vertices and is not 2-colorable.

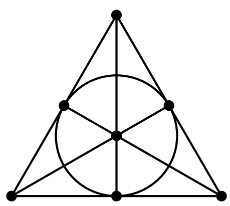
There is no loss in the vertex count in the other direction either: a hypergraph with fewer than 6 vertices can be handed as many extra vertices as it likes, and vertices lying in no edge change nothing about which colorings are proper.

Now count colorings. Among all the ways of coloring 6 vertices with two colors, consider only the balanced ones, which split V into two sets of three; there are $\frac{1}{2}\binom{6}{3} = 10$ of these. Such a coloring fails exactly when some edge is monochromatic, and an edge of H has three vertices, so it is monochromatic under a balanced coloring precisely when it is one of the two color classes. Each edge f therefore rules out exactly one of the ten balanced colorings, namely the one splitting V into f and $V \setminus f$. Six edges rule out at most six of the ten, so at least four survive, and any one of them is a proper 2-coloring of H. This contradicts our assumption, so no such H exists and

$$m(3) \geq 7$$

That bound is attained, and the hypergraph attaining it is a familiar object.

Example 2.1.8 (The Fano Plane). The **Fano plane** is the 3-uniform hypergraph on the seven vertices below, whose seven edges—by convention, its **lines**—are the three sides of the triangle, the three medians, and the circle through the three edge midpoints.



Every vertex lies on exactly three lines, any two vertices lie on exactly one line together, and any two lines meet in exactly one vertex. That last property is what does the work below.

Proposition 2.1.9. *The Fano plane is not 2-colorable. Hence $m(3) = 7$.*

Proof. Of the $\binom{7}{4} = 35$ sets of four vertices, exactly 28 contain a line: a line together with any one of the four vertices off it gives such a set, which is $7 \cdot 4 = 28$ sets in all, and no two of them coincide because four vertices cannot contain two lines—two lines meet in a vertex, so between them they cover five. The seven sets left over are the complements of the seven lines.

Now suppose we had a proper 2-coloring, with color classes A and B, and say $|A| \leq 3$. If $|A| = 3$, then A is not a line, since A is monochromatic; so $B = V \setminus A$ is a set of four vertices that is not the complement of a line, and hence contains one. If $|A| \leq 2$, then the vertices outside B lie on at most $3 + 3 - 1 = 5$ lines between them, so at least two lines avoid them both and lie inside B. Either way some line is monochromatic, which is a contradiction.

So the Fano plane is a 3-uniform hypergraph with seven edges that is not 2-colorable, giving $m(3) \leq 7$. □

2.1.4 Bounds on $m(k)$

Pinning down $m(3)$ took a page of case analysis, and doing the same for $m(4)$ is out of the question. What we can do for every k at once is bound it, and the lower bound falls out of an idea that will keep coming back: color at random, and show that a bad coloring is improbable.

Theorem 2.1.10. $m(k) \geq 2^{k-1}$. *In other words, any k-uniform hypergraph with $< 2^{k-1}$ edges is 2-colorable.*

Proof. Given a hypergraph with $< 2^{k-1}$ edges, randomly color the vertices red and blue, with equal probability $(1/2)$ for each. For each edge f, let $\mathcal{E}_f$ be the event that f is monochromatic. Then,

$$\Pr(\mathcal{E}_f) = \frac{2}{2^k}$$

Taking a simple union bound,

$$\Pr\left(\bigcup_{f \in \mathcal{H}} \mathcal{E}_f\right) \leq \frac{|\mathcal{H}|}{2^{k-1}} < 1$$

where we interpret $\bigcup_{f \in \mathcal{H}} \mathcal{E}_f$ as the event that there is a monochromatic edge. As the probability is less than 1, there must be at least one coloring with no monochromatic edge. □

Note that the above argument can also be written combinatorially instead of probabilistically. In this instance, the probabilistic language essentially conveys the same information as combinatorial language would. But this may not always be true, so it's worth being careful about this.

In the other direction, the truth is only a factor of about k^2 away.

Exercise 2.1.11. Prove that $m(k) = O\left(k^2 \cdot 2^k\right)$

2.1.5 Maker-Breaker Games

Theorem 2.1.10 says that a hypergraph with few enough edges has a proper 2-coloring somewhere among all the colorings of its vertices. It turns out that the very same threshold survives a far more demanding requirement: that one of two players, claiming vertices one at a time, be able to stop the other from ever owning a whole edge. Winning that does not by itself produce a proper 2-coloring, since the winner's own class may contain a monochromatic edge. It does force one to exist all the same: a breaker who can win going second can also win going first, by holding a spare vertex in reserve, and setting those two strategies against each other leaves neither class containing an edge.

Here's a fun fact: a "maker-breaker" game is a game where the "maker" wins if they are able to achieve an objective and the "breaker" wins if they are able to stop the maker from winning. So if you set up a "maker-breaker" version of tic-tac-toe, where only the crosses can win by making three in a row and the noughts can only win by stopping the crosses, the crosses will always win (assuming the crosses play first). This is, of course, different from tic-tac-toe, which, as our own studies from when we were a kid have shown, can always be drawn.

The threshold at which breaker can hold out is exactly the one we have just been looking at.

Theorem 2.1.12 (Erdős-Selfridge). *If H is a k-uniform hypergraph with m edges, and $m < 2^{k-1}$, then breaker wins (if breaker goes second).*

In the language of hypergraphs, the game is played on the vertices of H: the two players alternately claim unclaimed vertices, maker wins by claiming every vertex of some edge, and breaker wins if play ends with no edge wholly maker's.

Proof. Let Δ be the maximum degree of H, that is,

$$\Delta = \max_{v \in V} |\{f \in \mathcal{H} \mid v \in f\}|$$

We'll prove that if $m + \Delta < 2^k$, then breaker will win. Then, since $\Delta \leq m$, it will follow that if $m < 2^{k-1}$, then $m + \Delta \leq 2m < 2^k$.

At the start of move i, we will conceptualize the danger of edge $f \in \mathcal{H}$ by

$$\phi_i(f) = \begin{cases} 2^{\text{number of maker moves in } f} & \text{if there are no breaker moves in } f \\ 0 & \text{if there is a breaker move in } f \end{cases}$$

$$\Phi_i = \sum_{f \in \mathcal{H}} \phi_i(f)$$

Consider the strategy where, once maker has played y on move i, breaker plays the unclaimed vertex carrying the most danger, that is, the x maximizing

$$\sum_{f \ni x} \phi(f)$$

with ϕ measured after maker's move. Two things move Φ on move i. Maker's play doubles the danger of every edge through y that breaker has not already killed, so it raises Φ by $a_i = \sum_{f \ni y} \phi_i(f)$; breaker's play sends the danger of every surviving edge through x to 0, so it lowers Φ by the quantity b_i that breaker has just maximized. Hence

$$\Phi_{i+1} = \Phi_i + a_i - b_i$$

The point of breaker's rule is that $a_i \leq b_{i-1}$ for every $i \geq 2$. Indeed, y is unclaimed right up until maker takes it on move i, so it was one of the vertices breaker was choosing between on move $i - 1$, and breaker chose one carrying at least as much danger as it. Between then and the start of move i nothing happens except breaker's own play, which only ever sends dangers to 0, so the danger through y at the start of move i is no greater than it was when breaker passed y over.

Summing $\Phi_{i+1} = \Phi_i + a_i - b_i$ over the first n moves and canceling each a_i against b_{i-1}, everything past the first two terms disappears and we are left with

$$\Phi_{n+1} \leq \Phi_1 + a_1$$

Stopping the same sum one term earlier gives the same bound halfway through a move, after maker has played and before breaker has replied. Now $\Phi_1 = m$, since at the start every edge contains a move of neither player and so contributes $2^0 = 1$, and a_1 is the number of edges through maker's opening vertex, which is at most Δ. So the danger never exceeds $m + \Delta$ at any point in the game.

But if maker ever completes an edge f, then at that moment f contains all k of its vertices from maker and none from breaker, so $\phi(f) = 2^k$ and hence $\Phi \geq 2^k$. Since $m + \Delta < 2^k$, this never happens, and breaker wins. $\square$

Let's consider higher-dimensional maker-breaker tic-tac-toe.

Example 2.1.13. Turns out that $8 \times 8 \times 8$ maker-breaker tic-tac-toe is breaker-winnable: we can show that $m = \frac{10^3 - 8^3}{2} = 244$ and $\Delta = 7$, meaning $m + \Delta = 251 < 2^k = 256$.

Both numbers come out of the geometry. For m, pad the board with one extra cell in every direction, giving a $10 \times 10 \times 10$ cube, and run each winning line one step past each of its two ends. The two cells we land on lie in the shell we just added, and every cell of the shell is the end of exactly one line: a coordinate of the cell equal to 0 tells the line to increase along that axis, one equal to 9 tells it to decrease, and the rest stay fixed. So the lines are in bijection with those pairs of shell cells that arise as the two ends of a line, and since every shell cell is the end of exactly one line there are $\frac{10^3 - 8^3}{2}$ of them. For Δ, a line through a given cell is determined by its direction, and a corner cell has the 3 lines along the axes, the 3 face diagonals and the one long diagonal through it, which is 7; no cell of an $8 \times 8 \times 8$ board does better. And it is the $m + \Delta < 2^k$ form of the criterion doing the work here, since 244 is a long way above $2^{k-1} = 128$.

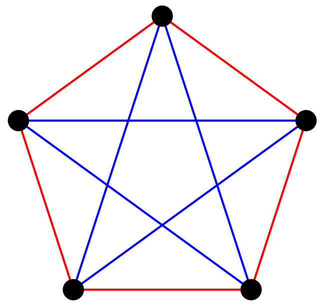
We won't say any more about tic-tac-toe in this course, but it was a fun aside.

2.2 Ramsey Numbers

Everything so far has asked when a coloring exists. The opposite question is at least as natural: how large does a structure have to be before every coloring of it is forced to contain something monochromatic? Let's step back into "graph land" from "hypergraph land" and ask it of the complete graphs, coloring edges this time rather than vertices.

Definition 2.2.1 (Ramsey Number)**.** The kth Ramsey number, denoted $R(k)$, is the smallest $n \in \mathbb{N}$ such that any 2-coloring of edges of K_n gives a monochromatic K_k.

One immediate result we can see is that $R(3) \geq 6$, which we can see simply by drawing a picture of K_5 and studying what happens if we move our attention to K_6.



Each color carries a 5-cycle—the outer pentagon in red, the inner pentagram in blue—and a 5-cycle contains no triangle at all, so neither color gives us a monochromatic K_3. Nothing forces a monochromatic triangle at $n = 5$, and so $R(3) \geq 6$. The picture also shows how little room there is to spare: every vertex of K_5 meets exactly two edges of each color, and the moment we pass to K_6 each vertex has five edges to distribute between two colors, so one of the colors is forced to take three of them.

We can actually prove that $R(3) = 6$, but this is rather nontrivial.

Other known Ramsey numbers are $R(2) = 2$ and $R(4) = 18$. The rest are unknown, though we do have bounds for them. The best (or close to the best) known bounds are

$$2^{\frac{k}{2}} \leq R(k) \leq 3.8^k$$

2.3 Chromatic Number

Back to coloring vertices, and this time we are not rationed to two colors. Chapter 2 opened by asking which graphs can be colored with two of them; the question now is how many colors a graph needs in general, and how far that number can be pushed away from the obvious lower bound.

Definition 2.3.1 (Chromatic Number)**.** The **chromatic number** of a graph G, denoted $\chi(G)$, is the smallest number of colors I need to color G.

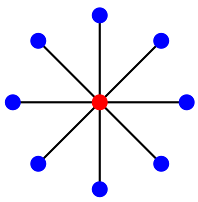
2.3.1 Trivial Bounds

There are some trivial bounds on $\chi(G)$.

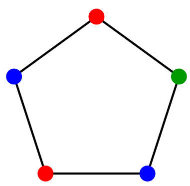
Definition 2.3.2 (Clique)**.** A **clique** in G is a complete subgraph, ie, a set of vertices any two of which are adjacent. The **clique number** of a graph, denoted $\omega(G)$, is the number of vertices in the largest clique.

It isn't hard to see that $\omega(G) \leq \chi(G)$ always. Similarly, it is clear that $\chi(G) \leq \Delta + 1$, where Δ is the maximum degree.

The upper-bound is absolutely not tight, for instance we can have a "star graph" with one vertex in the middle connected to a number of other vertices. This has high Δ but $\chi(G) = 2$.



Similarly, the lower-bound is also not tight. For example, any odd cycle on more than 3 vertices (eg, the 5-cycle) has clique number 2 but chromatic number 3.



In fact, we will show that $\chi(G)$ can be much bigger than $\omega(G)$.

2.3.2 Perfect Graphs

First, the graphs for which the lower bound is exact—and stays exact when we throw vertices away. That needs a word for what is left when we do.

Definition 2.3.3 (Induced Subgraph). Given a graph $G = (V, E)$ and a subset $W \subseteq V$, the subgraph of G **induced** by W, denoted $G[W]$, is the graph with vertex set W whose edges are precisely the edges of G with both endpoints in W. An **induced subgraph** of G is any subgraph of the form $G[W]$.

Definition 2.3.4 (Perfect Graph). A graph G is **perfect** if $\chi(H) = \omega(H)$ for all induced subgraphs H of G.

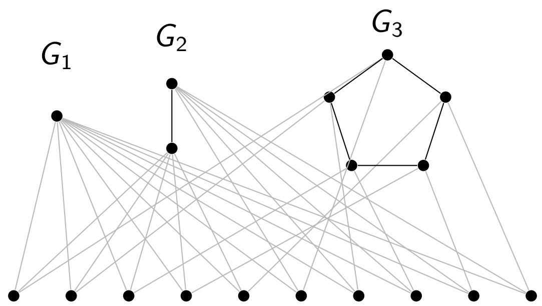
2.3.3 Zykov's Construction

At the opposite extreme are graphs with $\omega = 2$ and χ as large as we like, and we can build them by hand.

Let's now construct a sequence $(G_n)_{n\in\mathbb{N}}$. Start with G_1 a single vertex. Given $G_1, \ldots, G_{k-1}$, take disjoint copies of all of them, and for every way of choosing one vertex v_i from each G_i add a new vertex adjacent to $v_1, \ldots, v_{k-1}$ and to nothing else; the result is G_k. In the pictures below the copies of the earlier graphs sit on top and the new vertices along the bottom, so that each bottom vertex sends exactly one edge into each of the graphs above it. There are $|G_1| \cdots |G_{k-1}|$ new vertices, so G_2 has $1 + 1 = 2$ vertices and is a single edge, G_3 has $1 + 2 + 2 = 5$ and is the 5-cycle, and G_4 has $1 + 2 + 5 + 10 = 18$ (the construction is due to Zykov).

G_1 G_2 G_3

G_4 is already hard to take in, but the shape of the construction survives in it: the three earlier graphs across the top, and below them one new vertex for each of the $1 \cdot 2 \cdot 5 = 10$ ways of picking one vertex from each. The edges into the copies are drawn faintly, so that the copies themselves stay visible.



These have exactly the property we want.

Proposition 2.3.5. *For every $k \geq 1$, G_k has no triangles and $\chi(G_k) = k$.*

Proof. Both by induction on k, and both trivial for G_1. The new vertices of G_k form an independent set, and the neighbors of any one of them lie one in each of $G_1, \ldots, G_{k-1}$, so no two of them are adjacent. A triangle can therefore contain at most one new vertex, and if it contains one then its other two vertices are adjacent neighbors of that vertex, which do not exist. So every triangle of G_k lies inside some G_i, and by induction there are none.

For $\chi(G_k) \leq k$, color each G_i with $i \leq k-1$ colors, which we can by induction, and give every new vertex color k. For $\chi(G_k) \geq k$, suppose c is a proper $(k-1)$-coloring of G_k. Its restriction to G_i is a proper coloring of G_i, and so uses at least $\chi(G_i) = i$ colors. Pick $v_1 \in G_1$; G_2 uses at least two colors, so pick $v_2 \in G_2$ with $c(v_2) \neq c(v_1)$; G_3 uses at least three, so pick $v_3 \in G_3$ with $c(v_3) \notin \{c(v_1), c(v_2)\}$; and so on, at each step choosing $v_i \in G_i$ to avoid the $i-1$ colors chosen so far. After $k-1$ steps, $c(v_1), \ldots, c(v_{k-1})$ are $k-1$ distinct colors, which is all of them, and the new vertex attached to $v_1, \ldots, v_{k-1}$ has no color left. □

2.3.4 Graphs of Large Girth

The graphs of Theorem 2.3.5 have no triangles, but nothing stops them having cycles of length four or five—G_3 is one. It is natural to ask whether forbidding all short cycles finally holds the chromatic number down, and the answer is that it does not.

Definition 2.3.6 (Girth)**.** The **girth** of G, denoted girth(G), is the length of the shortest cycle in G.

The girth of a graph with no cycles is infinite.

Theorem 2.3.7 (Erdős)**.** *For any k, there are graphs of girth at least k and chromatic number at least k.*

The proof runs through independent sets rather than through colorings directly.

Definition 2.3.8 (Independence Number)**.** A set S of vertices of G is **independent** if no two of its elements are adjacent. The **independence number** of G, denoted $\alpha(G)$, is the size of the largest independent set in G.

We know we can construct graphs of average degree d with $\frac{nd}{2}$ edges. We want to prove that such a graph has a "large" independent set.

As a first attempt, let's choose a set S randomly by including each vertex of G independently with probability p. Write $e(S)$ for the number of edges of G with both endpoints in S, so that S is independent exactly when $e(S) = 0$. Then,

$$\mathbb{E}(|S|) = pn$$
$$\mathbb{E}(e(S)) = \frac{p^2 nd}{2}$$

the second line because each of the $\frac{nd}{2}$ edges lands inside S with probability p^2. We want S large and $e(S)$ small, and the two pull against each other: the expected size grows like p but the expected number of edges inside like p^2. So we take p as large as we can while keeping the expected number of edges inside S down to about one. I would want $p \simeq \sqrt{\frac{2}{dn}}$. In that case we would get something like $\mathbb{E}(|S|) \approx \sqrt{\frac{n}{d}}$. This is a poor return: a disjoint union of copies of K_{d+1} has average degree d and an independent set of size $\frac{n}{d+1}$, and the trouble is that we have insisted on S coming out independent all by itself.

Now, let's make another attempt. Still choose S randomly, including each vertex with probability p.

Let $X = |S|$ and $Y = e(S)$.

Now consider $\mathbb{E}(X - Y)$. This is $pn - \frac{p^2nd}{2}$ by linearity of expectation. So if S has X vertices and Y edges inside it, then G contains an independent set of size at least $X - Y$, and using $p = \frac{1}{d}$ gives $\mathbb{E}(X - Y) = \frac{n}{2d}$.

This needs $d \geq 1$ for $p = \frac{1}{d}$ to be a probability at all, and then some S has $X - Y \geq \mathbb{E}(X - Y) = \frac{n}{2d}$. Now alter that S: for each of the Y edges lying inside it, throw out one of the two endpoints. At most Y vertices go, and no edge of G keeps both its endpoints, so what is left is an independent set of size at least $X - Y \geq \frac{n}{2d}$.

What we have shown is worth recording.

Lemma 2.3.9. *If G is a graph with n vertices and average degree $d \geq 1$, then G has an independent set of size at least $\frac{n}{2d}$.*

The above is called the **method of alterations**—constructing something randomly and then changing based on how it goes.

The proof of Theorem 2.3.7 plays the same trick, but with a random graph in place of a random set of vertices.

Definition 2.3.10 (Random Graph)**.** For $n \in \mathbb{N}$ and $p \in [0, 1]$, the **random graph** $G_{n,p}$ is the graph on the vertex set $[n]$ in which each of the $\binom{n}{2}$ possible edges is present independently with probability p.

Proof of Theorem 2.3.7. Fix $k \geq 3$; for $k \leq 2$ a single edge already does. Consider the graph $G_{n,p}$ for $p = \frac{n^\varepsilon}{n}$. For some small $\varepsilon > 0$ (in particular, $\varepsilon \ll \frac{1}{k}$), we see that

$$\begin{aligned}\Pr(\alpha(G_{n,p}) \geq a) &\leq \binom{n}{a}(1-p)^{\binom{a}{2}} \\ &\leq n^a e^{-p\binom{a}{2}} = \exp\left(a\left(\ln n - \frac{p(a-1)}{2}\right)\right)\end{aligned}$$

using $\binom{n}{a} \leq n^a$ and $1 - p \leq e^{-p}$. Take $a = \left\lceil \frac{n}{2k} \right\rceil$, so that $a - 1 \geq \frac{n}{2k} - 1$ and hence $\frac{p(a-1)}{2} \geq \frac{n^\varepsilon}{4k} - \frac{1}{2}$. Any positive power of n eventually beats $\ln n$, so the bracket is negative for n large enough and the whole thing tends to 0: that is, $\Pr\left(\alpha(G_{n,p}) \geq \frac{n}{2k}\right) \to 0$ as $n \to \infty$.

Now, let X be the number of short cycles. That is, cycles of length $\leq k$. Then,

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{\ell=3}^{k} \frac{\binom{n}{\ell} \cdot \ell!}{2\ell} p^{\ell} \\
&\leq \frac{1}{2} \sum_{\ell=3}^{k} n^{\ell} p^{\ell} \\
&\leq (k-2) \frac{1}{2} n^{k} p^{k} \\
&\leq \frac{k}{2} n^{\varepsilon k}
\end{aligned}$$

By Markov's inequality (Theorem 1.3.1),

$$\Pr\left(X \geq \frac{n}{2}\right) \leq \frac{\mathbb{E}(X)}{n/2} \to 0$$

as $n \to \infty$. So, there is some graph G with n vertices, for sufficiently large n, such that $X \leq \frac{n}{2}$ and $\alpha(G) \leq \frac{n}{2k}$. Construct a new graph G' by deleting one vertex from each short cycle. This new graph has at least $\frac{n}{2}$ vertices, independence number at most $\frac{n}{2k}$, and thus chromatic number at least $\frac{n/2}{n/2k} \geq k$. $\square$

Chapter 3

Set Systems

Strip the coloring question away from a hypergraph and what is left is a family of subsets of a finite set, and this chapter asks what such a family can look like. Three questions, each with a theorem for an answer: whether one vertex can be chosen out of each edge, all of them different, which is Hall's marriage theorem; how many edges there can be once no edge is allowed to sit inside another, which is Sperner's; and how many there can be once no two edges are allowed to miss each other altogether, which is Erdős-Ko-Rado. The last two are answered by the same device, a permutation chosen uniformly at random, and it is worth watching for it the second time round.

3.1 Matchings in Bipartite Graphs

Hypergraphs came up in Theorem 2.1.4 as the setting in which the coloring question gets hard, and Chapter 2 set them aside again once it had its answer. The first thing this chapter asks of them is this: given a hypergraph, can we pick one vertex out of each edge, all of them different? Getting to the answer means first noticing that a hypergraph, a $0, 1$ matrix and a bipartite graph are three views of the same data.

3.1.1 Adjacency Matrices

We can represent the data of a hypergraph H in a matrix, known as the adjacency matrix.

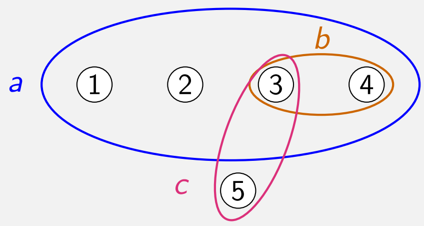
Definition 3.1.1 (Adjacency Matrix)**.** If H is a hypergraph with n vertices and m hyper-edges, both taken in some fixed order, then the adjacency matrix of H, denoted $\mathrm{Adj}(H)$, is the unique $n \times m$ matrix (a_{ij}) where

$$a_{ij} = \begin{cases} 1 & \text{if vertex } i \text{ lies in hyper-edge } j \\ 0 & \text{otherwise} \end{cases}$$

Here are some examples of constructing adjacency matrices.

Example 3.1.2 (Some Adjacency Matrices of Hypergraphs)**.**

1. Consider the hypergraph



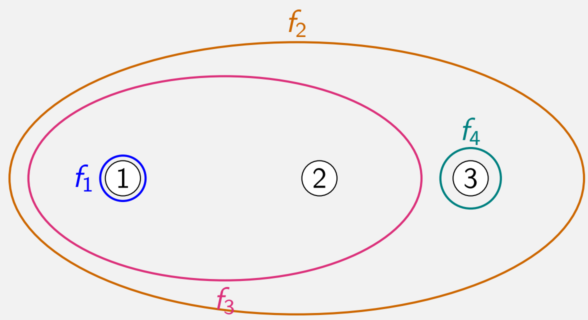
   Its adjacency matrix is

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

2. Consider the hypergraph

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

   Reading the columns off one at a time, this is the hypergraph on three vertices with the four hyper-edges $f_1 = \{1\}$, $f_2 = \{1, 2, 3\}$, $f_3 = \{1, 2\}$ and $f_4 = \{3\}$.



It turns out that there is not just a correspondence between hypergraphs and $0, 1$ adjacency matrices, we can also complete the trifecta with bipartite graphs.

$$\begin{tikzcd}
 & \text{Hypergraphs} \arrow[dl, leftrightarrow] \arrow[dr, leftrightarrow] & \\
0,1 \text{ matrices} \arrow[rr, leftrightarrow] & & \text{Bipartite graphs}
\end{tikzcd}$$

The bipartite graph attached to H has the vertices of H on one side and the hyper-edges of H on the other, with v joined to f exactly when $v \in f$, which is to say exactly when the adjacency matrix has a 1 in position (v, f).

3.1.2 Hall's Marriage Theorem

Recall, now, the most important result about bipartite graphs.

Theorem 3.1.3 (Hall's Marriage Theorem). *Given a bipartite graph with parts A and B, with A finite, there is a complete matching of A iff $\forall S \subseteq A$,*

$$|\Gamma(S)| \geq |S|$$

where $\Gamma(S) = \{v \in B \mid \exists u \in S \text{ s.t. } u \sim v\}$.

Proof.

($\Longrightarrow$) Let M be a complete matching of A and let $S \subseteq A$. The partners under M of the elements of S are $|S|$ distinct vertices of B, each adjacent to something in S, so $\Gamma(S)$ has at least $|S|$ elements.

($\Longleftarrow$) We induct on $|A|$. If $|A| = 1$, then Hall's condition applied to A itself says that the one vertex of A has a neighbor, and matching it to one is a complete matching. So suppose $|A| > 1$, and let's do some casework.

- <u>Case 1:</u> Assume that in fact, $\forall S \subsetneq A$ nonempty, $|\Gamma(S)| \geq |S| + 1$. Arbitrarily match some pair: pick any $a \in A$ and any $b \in \Gamma(\{a\})$, which is nonempty by hypothesis. After deleting this edge—and with it the vertices a and b and every other edge at either of them—the rest satisfies the desired result, since a nonempty $T \subseteq A \setminus \{a\}$ is a nonempty proper subset of A and so had at least $|T| + 1$ neighbors to begin with, of which deleting b can take away only one. What is left has $|A| - 1$ vertices on the left, so we are done by induction: it has a complete matching of $A \setminus \{a\}$, and adding the edge ab to that matching completes it to one of A.

- <u>Case 2:</u> Now assume the opposite, ie, that $\exists S \subsetneq A$ nonempty such that $|\Gamma(S)| = |S|$. We will apply induction again, on $|A|$ as before.

  If we restrict our attention to S, obviously we have a perfect matching between S and $\Gamma(S)$: every $T \subseteq S$ has $\Gamma(T) \subseteq \Gamma(S)$, so the graph they span satisfies Hall's condition and induction gives a complete matching of S, which the fact that they're of equal cardinality makes perfect. (There may be more than one, but there is one.)

  Define $A' = A \setminus S$, $B' = B \setminus \Gamma(S)$. We claim that the bipartite graph induced on A' and B' also has a complete matching of A', in which case we'd be done (by induction).

  Writing Γ' for the neighborhood taken in that smaller graph, we have $\Gamma'(T) = \Gamma(T) \setminus \Gamma(S)$ for every $T \subseteq A'$, and $\Gamma(S \cup T) = \Gamma(S) \cup \Gamma(T)$ as it does for any two sets, so

  $$|\Gamma'(T)| = |\Gamma(S) \cup \Gamma(T)| - |\Gamma(S)| = |\Gamma(S \cup T)| - |S| \geq |S \cup T| - |S| = |T|$$

  where the inequality is Hall's condition applied to $S \cup T$, and the last step is that $T \subseteq A'$ is disjoint from S. (If $\Gamma(T)$ is infinite the display needs no reading at all: $\Gamma(S)$

is finite, so $\Gamma'(T)$ is infinite and certainly at least $|T|$.) So the smaller graph satisfies Hall's condition, and $|A'| < |A|$ because S is nonempty, so induction gives a complete matching of A' into B'. That matching touches no vertex of S and no vertex of $\Gamma(S)$, so laying it alongside the matching of S onto $\Gamma(S)$ gives a complete matching of A.

$\square$

Given a hypergraph H, a family of distinct representatives is a collection of distinct $x_1, \ldots, x_m$ such that $x_i \in f_i$, where we have distinct hyper-edges $f_1, \ldots, f_m$ in $\mathcal{H}$. Such a collection exists iff for any subcollection $\mathcal{C}$ of $f_1, \ldots, f_m$,

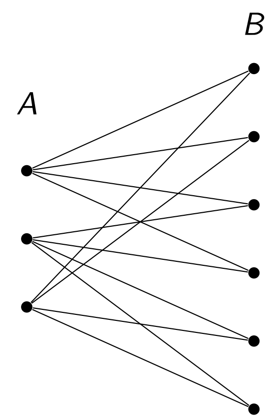
$$\left| \bigcup_{f \in \mathcal{C}} f \right| \geq |\mathcal{C}|$$

In other words, the number of vertices in all the edges in the collection should be at least the number of edges in the collection. This is what Hall's Marriage Theorem tells us about hypergraphs.

Hall's condition asks something of every subset of A, which is a lot to check. There is a consequence of it that only asks about degrees.

Theorem 3.1.4. *If G is a bipartite graph with parts A and B, if all degrees in A are d_1 and all degrees in B are d_2, with $1 \leq d_2 \leq d_1$, then there is a complete matching of A.*

Such a graph is a good deal more rigid than a bipartite graph with Hall's condition alone; here is one with $d_1 = 4$ and $d_2 = 2$.



Proof. Given a set $S \subseteq A$, there are $d_1 |S|$ edges coming out of S and going to the right. Each vertex of $\Gamma(S)$ absorbs at most d_2 of them, so

$$|\Gamma(S)| \geq \frac{d_1 |S|}{d_2} \geq |S|$$

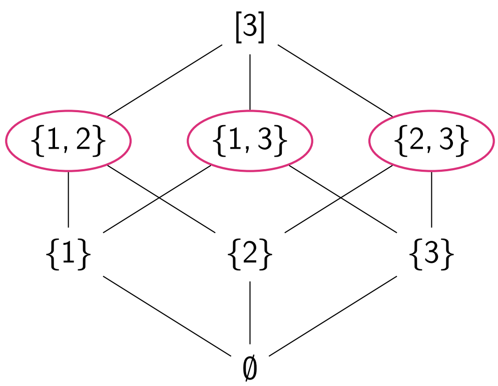
so we can just apply Theorem 3.1.3. □

3.2 Sperner Systems

A hypergraph is a vertex set together with a family of subsets of it, and the next question this chapter asks about one is how large that family can be. Unrestricted the question is dull—all 2^n subsets of $[n]$ form a hypergraph—so we forbid the cheapest way of making a family large, which is to pile sets up inside one another.

Definition 3.2.1 (Sperner System). A collection of subsets of $[n]$ such that none is a subset of another is called a Sperner system.

Example 3.2.2. The three two-element subsets of [3], circled below, form a Sperner system.



The picture is the whole collection of subsets of [3] ordered by inclusion, arranged in layers by size, with an edge drawn from each set to every set one size up that contains it. Containment between two subsets is then visible as an upward path, so a Sperner system is precisely a collection of vertices no two of which are joined by such a path—and drawing the diagram turns a question about set inclusions into a question about which vertices we are allowed to pick. The diagram as a whole is emphatically not a Sperner system; each individual layer is, since two distinct sets of the same size never contain one another. That gives us two Sperner systems of size 3 here, the singletons and the two-element subsets, and the top and bottom layers give two more, though useless ones: any system containing $\emptyset$ or $[3]$ can contain nothing else.

How large can a Sperner system be?

Theorem 3.2.3 (Sperner's Theorem). *A Sperner system on $[n]$ can have at most $\binom{n}{\lfloor n/2 \rfloor}$ sets.*

An intuitive way of arguing is to find complete matchings of the kth layer into the $(k+1)$th layer (cf. Theorem 3.1.4), but at some point, this starts to break down. And at what point is that? Before k exceeds $n/2$.

Proof. Let $\mathcal{H}$ be a Sperner system on $[n]$. Define

$$\mathcal{F} = \{\emptyset, \{1\}, \{1,2\}, \{1,2,3\}, \dots, \{1,2,\dots,n\}\}$$

Choose a random permutation $\pi \in S_n$, and define

$$\mathcal{H}_\pi = \{\pi(h) \mid h \in \mathcal{H}\}$$

Define the random variable

$$X = |\{f \in \mathcal{F} \mid f \in \mathcal{H}_\pi\}|$$

Note that we can express

$$X = \sum_{h \in \mathcal{H}} \xi_{h,\pi}$$

where $\xi_{h,\pi}$ is 1 if $\pi(h) \in \mathcal{F}$ and 0 otherwise.

Computing the expected value of X, we get

$$\mathbb{E}(X) = \sum_{h \in \mathcal{H}} \mathbb{E}(\xi_{h,\pi})$$

Let's now compute the expectation of $\xi_{h,\pi}$. This is the probability that the permutation $\pi \in S_n$ sends $h \in \mathcal{H}$ to some element of $\mathcal{F}$. Since you cannot modify sizes, the only possible element in $\mathcal{F}$ to which π can send h is precisely the set $\{1, 2, \dots, |h|\}$. How many such π can do this? We have to consider the way π acts on h, and the way it acts on the other $n - |h|$ elements of $[n]$, because we know it sends h to $\{1, 2, \ldots, |h|\}$. There are $|h|!\,(n-|h|)!$ such π, so the probability is

$$\frac{|h|!\,(n-|h|)!}{n!} = \frac{1}{\binom{n}{|h|}}$$

The conclusion now comes from bounding $\mathbb{E}(X)$ on the other side. Any π is a bijection and so preserves containment, which makes $\mathcal{H}_\pi$ a Sperner system; and $\mathcal{F}$ is a chain, any two of its members being comparable. A chain and a Sperner system share at most one set, so $X \le 1$ for every π, and hence $\mathbb{E}(X) \le 1$. Since $\binom{n}{k}$ is largest at $k = \lfloor n/2 \rfloor$, this gives

$$1 \ge \mathbb{E}(X) = \sum_{h \in \mathcal{H}} \frac{1}{\binom{n}{|h|}} \ge \sum_{h \in \mathcal{H}} \frac{1}{\binom{n}{\lfloor n/2 \rfloor}} = \frac{|\mathcal{H}|}{\binom{n}{\lfloor n/2 \rfloor}}$$

and rearranging is exactly the bound we wanted. □

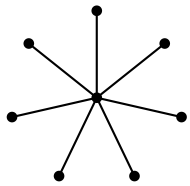
3.3 Intersecting Families

Sperner's theorem came out of forbidding one hyper-edge to sit inside another. Forbidding the opposite—a pair of hyper-edges that miss each other entirely—turns out to have a much more interesting answer, and it is the one we spend this section on.

Definition 3.3.1 (Intersecting). We say that a k-uniform hypergraph H is **intersecting** if $\forall h_1, h_2 \in \mathcal{H},\ h_1 \cap h_2 \neq \emptyset$.

The question is: how big can an intersecting k-uniform hypergraph on n vertices be? Ie, how many hyper-edges can it have? Small k is worth doing by hand, and as long as $n \ge 2k$ every case points the same way—until the last one, where dropping that assumption is exactly the point.

- $k = 1$: we can see that the maximum number of edges must be 1.
- $k = 2$: the maximum number of edges is $n - 1$, achieved by the 'star graph'



- $k = 3$: $\binom{n-1}{2}$, achieved by fixing one vertex that lies in everything and choosing 2 more.
- $k \le \frac{n}{2}$: $\binom{n-1}{k-1}$ by exactly the same strategy.
- $k > \frac{n}{2}$: we can do better! $\binom{n}{k}$ works because of something involving pigeons and holes.

3.3.1 The Erdős-Ko-Rado Theorem

The middle of that list is the general answer, and the bottom of it is why $n \ge 2k$ has to be assumed to get one.

Theorem 3.3.2 (Erdős-Ko-Rado). *If $n \ge 2k$, and H is an intersecting k-uniform hypergraph on n vertices, then the number of hyper-edges of H is $\le \binom{n-1}{k-1}$. Moreover, if $n \ge 2k+1$, then the ("unique") extremal example is given by all k-subsets containing a fixed vertex.*

As a warmup case, let's assume that $k \mid n$. Suppose H is intersecting. Let $\mathcal{F}$ denote some partition of $[n]$ into n/k subsets of size k (so $\mathcal{F}$ is fundamentally a collection of subsets).

$$X = |\{h \in \mathcal{H} \mid h = \pi(f) \text{ for some } f \in \mathcal{F}\}|$$

where $\pi \in S_n$ is a permutation chosen uniformly at random.

Note that $X \le 1$. Let's now compute its expected value.

$$\begin{aligned}\mathbb{E}(X) &= \sum_{h\in\mathcal{H}}\sum_{f\in\mathcal{F}} \Pr(h = \pi(f)) \\ &= |\mathcal{H}| \cdot \frac{n}{k} \cdot \frac{1}{\binom{n}{k}}\end{aligned}$$

So since $X \le 1$, we conclude that

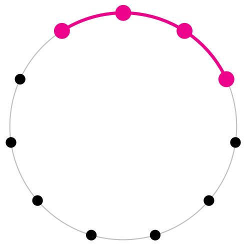
$$|\mathcal{H}| \cdot \frac{n}{k} \cdot \frac{1}{\binom{n}{k}} \le \mathbb{E}(X) \le 1$$

Rearranging gives us that

$$|\mathcal{H}| \leq \binom{n}{k} \cdot \frac{k}{n} = \binom{n-1}{k-1}$$

Now let's prove the result in full generality. The random permutation survives, but the partition into n/k blocks does not, and what replaces it is a cycle.

Local Notation. Lay the n vertices out around a circle. A k-**interval** of that cycle is a set of k vertices occupying consecutive positions on it, and $\mathcal{F}_\pi$ will denote the collection of all of them when the vertices are laid out in the order $\pi(1), \pi(2), \ldots, \pi(n)$.



There are n intervals in $\mathcal{F}_\pi$, one starting at each vertex, and—this is the whole point of moving to a cycle—very few of them can pairwise intersect.

Lemma 3.3.3. *Let $n \geq 2k$. Any collection $\mathcal{C} \subseteq \mathcal{F}_\pi$ of pairwise intersecting k-intervals has $|\mathcal{C}| \leq k$. Moreover, if $n \geq 2k+1$ and $|\mathcal{C}| = k$, then $\mathcal{C}$ is the collection of all the k-intervals containing some one vertex.*

Proof. Assume $\mathcal{C}$ is nonempty and fix some $I \in \mathcal{C}$, occupying positions $a, a+1, \ldots, a+k-1$. For $i = 1, \ldots, k-1$, let A_i be the interval ending at position $a+i-1$ and let B_i be the one starting at position $a+i$. Every interval other than I that meets I is one of these $2k-2$: an interval meets I exactly when it starts at one of the $2k-1$ positions $a-k+1, \ldots, a+k-1$, and the A_i take the $k-1$ of those before a while the B_i take the $k-1$ after it. Now A_i and B_i between them occupy the $2k$ consecutive positions $a+i-k, \ldots, a+i+k-1$, which are distinct because $n \geq 2k$, so $A_i \cap B_i = \emptyset$ and $\mathcal{C}$ can contain at most one of the two. Taking I together with one interval from each of the $k-1$ pairs gives $|\mathcal{C}| \leq k$.

Suppose now that $n \geq 2k+1$ and $|\mathcal{C}| = k$, so that $\mathcal{C}$ holds exactly one of A_i, B_i for every i. It cannot hold both A_i and B_{i+1}: those occupy positions $a+i-k, \ldots, a+i-1$ and $a+i+1, \ldots, a+i+k$, which lie among $2k+1$ consecutive positions and so are distinct, leaving the two intervals disjoint. So as soon as some B_j lies in $\mathcal{C}$ so does every B_i with $i < j$, and hence there is an m with

$$\mathcal{C} = \{A_{k-1}, \ldots, A_{m+1}, I, B_1, \ldots, B_m\}$$

Reading off where each of those starts, they are the intervals starting at positions $a+m+1-k, \ldots, a+m$, which is a run of k consecutive positions—and the intervals starting there are precisely the ones containing the vertex at position $a+m$. $\square$

The hypothesis $n \geq 2k+1$ does real work in that second half: on C_6 the three 3-intervals $\{1,2,3\}$, $\{3,4,5\}$ and $\{5,6,1\}$ pairwise intersect, and there are $k = 3$ of them, but no vertex lies in all three.

Proof of the bound in Theorem 3.3.2. Choose $\pi \in S_n$ uniformly at random and set

$$X = |\{h \in \mathcal{H} \mid h \in \mathcal{F}_\pi\}|$$

The intervals being counted here are hyper-edges of H, so they pairwise intersect, and by Theorem 3.3.3 it is always true that $X \leq k$. Now, let's compute the expectation of X.

$$\mathbb{E}(X) = \sum_{h \in \mathcal{H}} \Pr(h \in \mathcal{F}_\pi)$$

A given h is an interval of the randomly ordered cycle exactly when π lays its k elements out consecutively, and we can count the π that do so. There are n places for the interval to start, $k!$ ways of arranging h within it and $(n-k)!$ ways of arranging the other $n-k$ vertices, and no π gets counted twice, since an interval determines where it starts. So

$$\Pr(h \in \mathcal{F}_\pi) = \frac{n \cdot k!\,(n-k)!}{n!} = \frac{n}{\binom{n}{k}}$$

and hence

$$|\mathcal{H}| \cdot \frac{n}{\binom{n}{k}} = \mathbb{E}(X) \leq k$$

Rearranging,

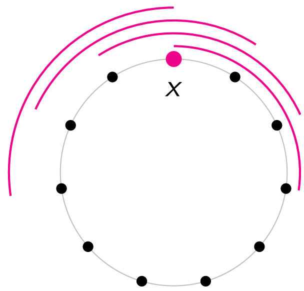
$$|\mathcal{H}| \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}$$

□

3.3.2 The Extremal Case

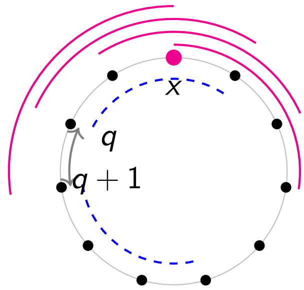
That argument has slack in exactly one place, and at $|\mathcal{H}| = \binom{n-1}{k-1}$ there is none left: X is then equal to k for every ordering rather than merely on average. Once $n \geq 2k+1$, so that the second half of Theorem 3.3.3 is available, every ordering of the cycle hands us a run of k intervals all of which are hyper-edges—and that pins $\mathcal{H}$ down completely. The following is the "moreover" of Theorem 3.3.2, made precise.

Theorem 3.3.4. *Let H be an intersecting k-uniform hypergraph on n vertices. If $|\mathcal{H}| = \binom{n-1}{k-1}$ and $n \geq 2k+1$, then the extremal example is the obvious one: there is a vertex x such that $\mathcal{H}$ consists of precisely the k-subsets of $[n]$ that contain x.*



Whatever order we lay the vertices out in, the picture looks like the one above: the hyper-edges we can see on the cycle are the k intervals through one distinguished vertex, drawn here as concentric arcs. All the work is in showing that the distinguished vertex does not move when the order changes.

Here is the picture to keep in mind. Swapping two neighbors on the cycle disturbs only two of its n intervals—the one ending at the first of them and the one starting at the second, each of which holds exactly one of the pair—and every other interval keeps its vertices. So at least $k-1$ of the k arcs through x are still hyper-edges and still intervals after the swap, and $k-1$ consecutive arcs leave the new fan almost no room to sit anywhere else. The proof below is the bookkeeping that turns "almost" into "none".



Here $n = 11$ and $k = 4$. The two dashed intervals are the ones the swap of positions q and $q+1$ disturbs; one of them is among the four hyper-edges through x and the other is not, and the remaining three arcs are untouched.

Proof of Theorem 3.3.4. Run the argument above again, this time at equality. If $|\mathcal{H}| = \binom{n-1}{k-1}$, then

$$\mathbb{E}(X) = \binom{n-1}{k-1} \cdot \frac{n}{\binom{n}{k}} = k$$

and we already know that $X \le k$ whatever π is. There are finitely many π and each is equally likely, so a mean of k under a ceiling of k leaves nothing to spare: $X = k$ for every single π. However we lay the n vertices out around a cycle, then, exactly k of that cycle's k-intervals are hyper-edges of H, and by Theorem 3.3.3 those k intervals are the intervals through a single vertex—one vertex and no more, since the outermost two of them share only it. Write x_π for that vertex. For $k = 1$ an intersecting family is a single hyper-edge and the theorem is immediate, so assume from here that $k \ge 2$.

It remains to see that x_π does not depend on π, and one adjacent transposition at a time is enough. Fix an ordering, put $x = x_\pi$ at position 0 and number the other positions $1, \dots, n-1$ round the circle, and let I_j be the interval occupying *positions* $j, j+1, \dots, j+k-1$, indices read modulo n. The intervals through position c are $I_{c-k+1}, \dots, I_c$, so the hyper-edges of this ordering are exactly $I_{-(k-1)}, \dots, I_0$. Transposing the vertices at adjacent positions q and $q+1$, neither of them position 0, changes which vertices occupy I_j exactly when I_j holds one of those two positions and not the other—which happens for $j = q+1$ and for $j = q-k+1$, and for no other j.

So the new ordering's hyper-edges, again the intervals through a single position c, agree with $I_{-(k-1)}, \ldots, I_0$ outside those two indices. But two runs of k consecutive indices whose right-hand ends sit c apart differ in $2\min(|c|, k)$ places, where $|c|$ is the distance from c to 0 measured around the circle, so $c \neq 0$ would force $|c| = 1$, and the two indices they differ in are then $\{-(k-1), 1\}$ if $c = 1$ and $\{-k, 0\}$ if $c = -1$. Either set would have to be $\{q-k+1, q+1\}$. Both pairs differ by k, and matching them in the only order that does not ask for $-k \equiv k \pmod{n}$ (which is impossible, as $n > 2k$) gives $q = 0$ in the first case and $q = -1$ in the second. Both are excluded, so $c = 0$: the distinguished vertex is the one at position 0, which the transposition left where it was.

Adjacent transpositions avoiding position 0 generate every arrangement of the other $n-1$ vertices around the circle, and each one leaves the distinguished vertex at position 0 for the next to start from, so the paragraph above applies afresh along the whole chain. Rotating or reflecting a layout leaves its intervals, and hence x_π, alone—so every π can first be turned until x sits at position 0, and $x_\pi = x$ for all of them. Take any k-subset h of $[n]$ containing x, and lay the vertices out with the elements of h consecutive: then h is an interval containing x, so $h \in \mathcal{H}$. There are $\binom{n-1}{k-1}$ such h, which is exactly $|\mathcal{H}|$, so they are all of it. $\square$

Chapter 4

Moments and Concentration

The probabilistic method has been a tool in the two chapters before this one—a random coloring for the lower bound on $m(k)$, a random permutation for Sperner's theorem and for Erdős-Ko-Rado, a random set and then a random graph for Theorem 2.3.7—and every one of those arguments has the same shape: compute an expectation, and read a structure off it. This chapter is about what to do when that is not enough. Knowing $\mathbb{E}(X)$ is large says nothing on its own about whether X is ever nonzero, so we bring in the variance, which is enough to settle when a random graph has a triangle and, more generally, when it has a copy of any fixed graph we like. When the variance in turn is too crude—and for the degrees of a random graph it is hopeless—we go past it to the exponential moments, and get a bound that beats every power of $X - \mathbb{E}(X)$ at once.

The last section takes that idea off sums altogether. A quantity that is not a sum of anything—the chromatic number of a random graph, say—concentrates just as sharply provided it can be revealed a little at a time, and the sequence of guesses we make as it is revealed is a martingale.

Everything here is built on Chapter 1, and on Theorem 1.3.1 and Theorem 1.3.3 in particular; nothing else from outside is needed.

4.1 Second Moment Methods

Let X be the number of triangles in the random graph $G_{n,p}$ of Theorem 2.3.10. We get

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

If $p = \frac{\omega(1)}{n}$ then $\mathbb{E}(X) \to \infty$, so $G_{n,p}$ is likely to have a triangle.

If $p = o\left(\frac{1}{n}\right)$, then $\mathbb{E}(X) \to 0$, so the probability that $G_{n,p}$ has a triangle goes to 0.

None of this actually tells us whether $G_{n,p}$ actually has a triangle or not. We need something more sophisticated than the expectation, so we turn to the variance, and Theorem 1.3.4 turns the computation into exactly the one we want.

Theorem 4.1.1. *If $np \to \infty$, then with high probability $G_{n,p}$ contains a triangle.*

Proof. For a triple T of vertices write I_T for the indicator of the event that T spans a triangle, so that $X = \sum_T I_T$ and $\mathbb{E}(X) = \binom{n}{3} p^3$. Expanding the variance,

$$\mathrm{Var}(X) = \sum_{T,T'} \left(\mathbb{E}(I_T I_{T'}) - \mathbb{E}(I_T)\,\mathbb{E}(I_{T'})\right)$$

where the sum runs over ordered pairs of triples. A pair with $|T \cap T'| \leq 1$ contributes nothing: T and T' then have no pair of vertices in common, so I_T and $I_{T'}$ are determined by disjoint sets of edges and are independent. There are at most n^3 pairs with $T = T'$, each contributing $p^3 - p^6 \leq p^3$. And a pair with $|T \cap T'| = 2$ shares exactly one edge, so $T \cup T'$ spans 5 edges and $\mathbb{E}(I_T I_{T'}) = p^5$; there are at most $3n^4$ such pairs, since T' is determined by T, which of the 3 pairs inside T is shared, and the remaining vertex. Hence

$$\mathrm{Var}(X) \leq n^3 p^3 + 3n^4 p^5$$

For $n \geq 6$ we have $\binom{n}{3} \geq \frac{n^3}{12}$, so $\mathbb{E}(X)^2 \geq \frac{n^6 p^6}{144}$ and

$$\frac{\mathrm{Var}(X)}{\mathbb{E}(X)^2} \leq \frac{144}{(np)^3} + \frac{432}{n \cdot np}$$

Both terms tend to 0 when $np \to \infty$, and Theorem 1.3.4 finishes it. $\square$

Can we apply this probabilistic approach in the following more general situation: what is the threshold for a random graph $G_{n,p}$ to contain H as a subgraph, where H is an arbitrary fixed graph?

What controls the answer is not the size of H but the density of its densest part.

Definition 4.1.2 (Maximum Density)**.** Let H be a graph with at least one edge, and for a graph J write $e(J)$ and $v(J)$ for its numbers of edges and of vertices. The **maximum density** of H is

$$\varepsilon(H) := \max_{J \subseteq H,\ v(J)>0} \frac{e(J)}{v(J)}$$

the maximum being over all subgraphs J of H with at least one vertex.

This is some sort of "edge density" or whatever. Think Szemerédi.

Markov's inequality already settles one side of the threshold.

Theorem 4.1.3. *Let H be a fixed graph with at least one edge. If $p = o\left(\frac{1}{n^{\frac{1}{\varepsilon(H)}}}\right)$, then with high probability*[a] *$G_{n,p}$ contains no H subgraph.*

[a] for instance, the limit of the probability goes to 1 as $n \to \infty$

Proof. Let $J \subseteq H$ attain the maximum in the definition of $\varepsilon(H)$, and write $v = v(J)$ and $e = e(J)$, so that $\frac{1}{\varepsilon(H)} = \frac{v}{e}$. Every copy of H in $G_{n,p}$ contains a copy of J, so it is enough to show that with high probability there is no copy of J.

Let X be the number of copies of J in $G_{n,p}$. A copy is determined by an injection $V(J) \to [n]$, and different injections may give the same copy, so there are at most n^v of them; each is present precisely when its e edges all are, which happens with probability p^e. Hence

$$\mathbb{E}(X) \leq n^v p^e$$

Now $p = o\left(n^{-v/e}\right)$, so $p^e = o(n^{-v})$ and $\mathbb{E}(X) = o(1)$. By Theorem 1.3.1, $\Pr(X \geq 1) \leq \mathbb{E}(X) \to 0$. □

The bound is sharp in the sense that at $p = n^{-1/\varepsilon(H)}$ itself the first moment stops deciding anything, which is where the second moment has to come in.

Example 4.1.4. Let J be K_4 without one edge, so $v(J) = 4$ and $e(J) = 5$. Its subgraphs are sparser than it is—a triangle has density 1, a single edge $\frac{1}{2}$—so $\varepsilon(J) = \frac{5}{4}$ and the threshold sits at $p = \frac{1}{n^{4/5}}$. At exactly that p, each 4-set of vertices carries $\frac{4!}{|\mathrm{Aut}(J)|} = 6$ copies of J, so the number X of copies has

$$\mathbb{E}(X) = \binom{n}{4} \cdot 6 \cdot p^5 = \binom{n}{4} \cdot 6 \cdot \frac{1}{n^4} \to \frac{1}{4}$$

Markov's inequality bounds $\Pr(X \geq 1)$ by about $\frac{1}{4}$ and says nothing at all about whether X is ever positive.

Our goal is to show that once the expected number of Js goes to infinity (J being the densest part of H), we have a copy of H.

4.2 Chernoff's Bound

Consider a random graph $G_{n,p}$ with n vertices and edges chosen uniformly at random with probability p, and take $p = \frac{1}{2}$ for this discussion.

The expected degree of a vertex is $\frac{n-1}{2}$.

How many vertices have degrees lying in $(0.49n, 0.51n)$? We can actually show that

$$\lim_{n\to\infty} \Pr(\text{every vertex in } G_{n,p} \text{ has degree in } (0.49n, 0.51n)) = 1$$

Indeed, fix a vertex x. $\deg(x)$ is a binomial random variable, counting the $n-1$ possible edges at x, each present independently. $\mathbb{E}(\deg(x)) = \frac{n-1}{2}$ and $\mathrm{Var}(\deg(x)) = \frac{n-1}{4}$.

Applying Chebyshev gives us

$$\Pr\left(\left|\deg(x) - \frac{n-1}{2}\right| \geq \delta n\right) \leq \frac{\frac{n-1}{4}}{(\delta n)^2} \approx \frac{1}{4\delta^2 n}$$

But this is a rather terrible bound: for $\delta = 0.01$, this is already $\frac{2{,}500}{n}$. Saying a probability is less than or equal to this is essentially useless!

We can consider taking higher moment bounds: bounding using not the second power of $X - \mathbb{E}(X)$, but the fourth or sixth or eighth power. In this instance, a fourth power bound would actually be enough. But that won't always be the case.

What's one thing that beats *all* of these? An exponential!

Consider the following setup. Let $\xi_1, \dots, \xi_n$ be independent Bernoulli random variables take values ± 1 with probability $\frac{1}{2}$. Define the binomial random variable $\eta := \sum_{i=1}^n \xi_i$.

Note that because $x < y \iff e^x < e^y$, we get that

$$\Pr(\eta \geq a) = \Pr(\beta^\eta \geq \beta^a) \leq \frac{\mathbb{E}(\beta^\eta)}{\beta^a}$$

by Markov's inequality (where β denotes e^t for $t > 0$, a bit like how q denotes $e^{2\pi i z}$).

Optimizing over t turns that into a bound that decays in a faster than any power.

Theorem 4.2.1 (Chernoff's Bound). *Let $\xi_1, \dots, \xi_n$ be independent random variables each taking the values ± 1 with probability $\frac{1}{2}$, and let $\eta = \sum_{i=1}^n \xi_i$. Then for every $a \geq 0$,*

$$\Pr(\eta \geq a) \leq e^{-\frac{a^2}{2n}}$$

Proof. At $a = 0$ the bound reads $\Pr(\eta \geq 0) \leq 1$ and there is nothing to prove, so take $a > 0$, and fix $t > 0$ for the moment. Notice that

$$\mathbb{E}(e^{t\eta}) = \mathbb{E}\left(e^{t\sum_{i=1}^n \xi_i}\right) = \mathbb{E}\left(\prod_{i=1}^n e^{t\xi_i}\right) = \prod_{i=1}^n \mathbb{E}\left(e^{t\xi_i}\right) = \left(\frac{e^{-t} + e^t}{2}\right)^n$$

the third equality by independence. Comparing the series termwise, $\frac{e^t + e^{-t}}{2} = \sum_{k \geq 0} \frac{t^{2k}}{(2k)!} \leq \sum_{k \geq 0} \frac{t^{2k}}{2^k k!} = e^{\frac{t^2}{2}}$, since $(2k)! \geq 2^k k!$. We therefore conclude that

$$\Pr(\eta \geq a) = \Pr(e^{t\eta} \geq e^{ta}) \leq \frac{\mathbb{E}(e^{t\eta})}{e^{ta}} \leq e^{\frac{t^2 n}{2} - ta}$$

Can we optimize this? Ie, can we find the t that minimizes this? It turns out, we can: by some calculus nonsense, we get that the optimal value is $e^{-\frac{a^2}{2n}}$. $\square$

The same bound holds if the ξ_i are independent and for all i, $|\xi_i| \leq 1$ and $\mathbb{E}(\xi_i) = 0$.

For instance, consider the case where

$$\xi_i = \begin{cases} \frac{1}{3} & \text{with probability } \frac{3}{4} \\ -1 & \text{with probability } \frac{1}{4} \end{cases}$$

and as before, let η denote their sum. By the same sort of technology we used above,

$$\mathbb{E}(e^{t\eta}) = \prod_{i=1}^{n} \mathbb{E}\left(e^{t\xi_i}\right)$$

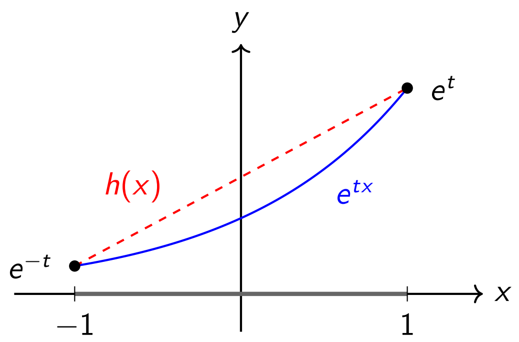
Define

$$h(x) = \frac{e^t + e^{-t}}{2} + x\frac{e^t - e^{-t}}{2}$$

We can show that

$$\mathbb{E}\left(e^{t\xi_i}\right) \le \mathbb{E}(h(\xi_i)) = h(\mathbb{E}(\xi_i)) = h(0) = \frac{e^t + e^{-t}}{2}$$

The first inequality is the whole content of the extension: h is the chord of $x \mapsto e^{tx}$ across $[-1, 1]$, and since e^{tx} is convex it lies below that chord on the whole interval, which is where $|\xi_i| \le 1$ is used. The first equality is linearity, since h is affine, and the second is where $\mathbb{E}(\xi_i) = 0$ is used.



4.3 Martingales

Theorem 4.2.1 got its strength from a very particular structure: a sum of independent contributions, each of them small. Plenty of quantities we care about have no such structure—the chromatic number of a random graph is not a sum of anything—and concentrate just as sharply anyway. What they do have is that they can be revealed a little at a time, with each revelation moving the expectation only slightly, and that is enough.

Throughout this section, fix random variables $X_0, X_1, X_2, \ldots$ on some probability space $(\Omega, \mathcal{P}(\Omega), \Pr)$.

The sequence to have in mind is the one where X_k is our best guess at some quantity after k things have been revealed, so that each term is the average of the term after it.

Definition 4.3.1 (Martingale). We say $X_0, X_1, \ldots$ is a **martingale** if the following all hold:

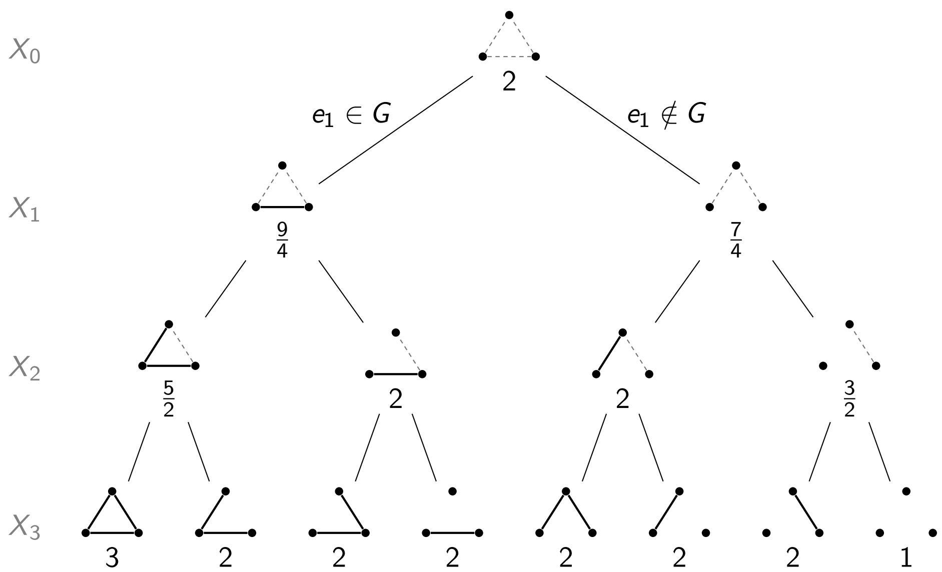
1. for all integers $n \geq 0$, $|X_{n+1} - X_n| \leq 1$
2. the condition below is imposed only where $\Pr(X_0 = x_0, \ \ldots, \ X_n = x_n) \neq 0$, the conditional expectation being undefined otherwise
3. for all integers $n \geq 0$ and all $x_0, \ldots, x_n \in \mathbb{R}$,

$$\mathbb{E}(X_{n+1} \mid X_0 = x_0, \ X_1 = x_1, \ \cdots, \ X_n = x_n) = x_n$$

4.3.1 Exposing a Random Graph

Let's study the expected value of chromatic numbers (cf. Theorem 2.3.1). Consider the family of random graphs $G_{3,\frac{1}{2}}$ of graphs on 3 vertices with the probability of an edge between them being $\frac{1}{2}$ (cf. Theorem 2.3.10). This makes $\chi\left(G_{3,\frac{1}{2}}\right)$ a random variable. What would we expect its expectation to be?

We can heuristically see that the expected chromatic number is two based on the following figure:



Fix an order e_1, e_2, e_3 on the three possible edges and look at them one at a time. An edge not yet looked at is drawn dashed, since it is there with probability $\frac{1}{2}$; one already looked at is drawn solid if it is there and rubbed out if it is not. So the root has all three dashed, its two children split on e_1, theirs on e_2, and the eight leaves are the eight graphs on three vertices.

Under each picture is the expected chromatic number of a graph from the family it stands for—at a leaf, just that graph's: 1 for the empty graph, 2 for one or two edges, 3 for the triangle. Each is the average of the two below it, so the 2 at the root is $\mathbb{E}\left(\chi\left(G_{3,\frac{1}{2}}\right)\right)$.

Write G for $G_{3,\frac{1}{2}}$ and $e_1, \ldots, e_m$ for the $m = \binom{3}{2}$ possible edges.

Define the following martingale:

$$
\begin{aligned}
X &= \chi(G) \\
Y_1 &= \mathbf{1}_{e_1 \in G} \\
&\vdots \\
Y_m &= \mathbf{1}_{e_m \in G} \\
X_0 &= \mathbb{E}(X) \\
X_1 &= \mathbb{E}(X \mid Y_1) \\
X_2 &= \mathbb{E}(X \mid Y_1, Y_2) \\
&\vdots
\end{aligned}
$$

Then, $X_0, X_1, X_2, \ldots$ is a martingale, and it is exactly what the tree draws: X_k is the number written under the picture at the node reached by exposing $e_1, \ldots, e_k$. Its increments are never more than $\frac{1}{2}$, so the first condition is met with room to spare; X_0 is the 2 at the root, and $X_m = X$ is the chromatic number itself, every X_k past that being X over again with nothing left to expose.

We can adapt this to the general $G_{n,\frac{1}{2}}$ case. Define Y_i and X_i as above. Then, we can compute the conditional expectation of $\chi\left(G_{n,\frac{1}{2}}\right)$ on observing edges incident on vertices $1, \ldots, k$: write X for $\chi\left(G_{n,\frac{1}{2}}\right)$, let Y_k record the edges from vertex k back to $1, \ldots, k-1$, and set $X_k = \mathbb{E}(X \mid Y_1, \ldots, Y_k)$. Since Y_1 records nothing, $X_0 = X_1 = \mathbb{E}(X)$; and $X_n = X$, with $X_k = X$ beyond that as before.

We are now exposing a vertex at a time rather than an edge at a time, which is what bounds the increments. Two graphs differing only at one vertex have chromatic numbers differing by at most 1: give that vertex a color of its own. Pairing each unexposed configuration with itself under a different value of Y_{k+1}, the possible values of X_{k+1} differ pairwise by at most 1, and X_k is their average, so $|X_{k+1} - X_k| \leq 1$.

4.3.2 Azuma's Inequality

That condition is what buys us the following: a martingale cannot get far from where it started, and by the same margin Theorem 4.2.1 gives a ± 1 walk.

Theorem 4.3.2 (Azuma's Inequality). *Let $X_0, X_1, \ldots$ be a martingale and let m be a positive integer. Then for every $\lambda > 0$,*

$$\Pr\left(X_m - X_0 \geq \lambda\sqrt{m}\right) \leq e^{-\frac{\lambda^2}{2}}$$

We won't prove it: it is the argument behind Theorem 4.2.1, with the chord bound applied to each increment conditionally on what has been exposed—which is what the bound on the increments is there to license.

So $\chi\left(G_{n,\frac{1}{2}}\right)$ sits in a window of width about $\sqrt{n}$ around its mean—much narrower than χ itself, which grows like $\frac{n}{\log n}$ (not something we will prove).

Corollary 4.3.3. *For every $n \geq 1$ and every $\lambda > 0$,*

$$\Pr\left(\left|\chi\left(G_{n,\frac{1}{2}}\right) - \mathbb{E}\left(\chi\left(G_{n,\frac{1}{2}}\right)\right)\right| \geq \lambda\sqrt{n}\right) \leq 2e^{-\frac{\lambda^2}{2}}$$

Proof. The martingale above has $X_0 = \mathbb{E}\left(\chi\left(G_{n,\frac{1}{2}}\right)\right)$ and $X_n = \chi\left(G_{n,\frac{1}{2}}\right)$, so Theorem 4.3.2 at $m = n$ bounds $\Pr\left(X_n - X_0 \geq \lambda\sqrt{n}\right)$ by $e^{-\frac{\lambda^2}{2}}$. Applying it to $-X_0, -X_1, \ldots$, a martingale for the same reasons, bounds the other tail. $\square$

It turns out that the martingale we constructed for the chromatic number problem is a specific instance of a more general martingale, known as the Doob Martingale.

Definition 4.3.4 (Doob Martingale). For random variables $X, Y_1, Y_2, \ldots$ on some probability space, define

$$X_0 := X \qquad \forall i \geq 1,\ X_i := \mathbb{E}(X \mid Y_1, \ldots, Y_i)$$

It is possible to show that $X_0, X_1, X_2, \ldots$ is a martingale, called the **Doob Martingale**.

Definition 4.3.5 (Coordinatewise Lipschitz). Given metric spaces $(X_i, d_i)_{i=1}^n$ and (Y, u), we say a function $f : X_1 \times \cdots \times X_n \to Y$ is k-**coordinatewise-Lipschitz** if $u(f(\mathbf{a}), f(\mathbf{b})) \leq k$ if $\mathbf{a}$ and $\mathbf{b}$ differ in exactly one coordinate.

Proposition 4.3.6 (McDiarmid's Inequality). *Suppose $Y_1, \ldots, Y_n$ are independent random variables on domains $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$. Suppose $f : \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n \to \mathbb{R}$ is 1-coordinatewise-Lipschitz. Denoting $X = f(Y_1, \ldots, Y_n)$, we have that*

$$\Pr(X \geq \mathbb{E}(X) + a) \leq e^{-\frac{a^2}{2n}}$$

for all $a \in \mathbb{R}$.

Proof. Consider the Doob Martingale $X_i = \mathbb{E}(X \mid Y_1, \ldots, Y_i)$. Suppose we've fixed $y_1, \ldots, y_{i-1}$. Define

$$g_i(y) = \mathbb{E}(f(y_1, \ldots, y_{i-1}, y, Y_{i+1}, \ldots, Y_n))$$

sorry □

One application of McDiarmid's Lemma is something called the Smiley Face Lemma, from a paper of Allen Frieze and Wes.

For a finite set S, any $\varepsilon > 0$, and any $D \subseteq \mathbb{R}$, an ε-D copy of S in $\mathbb{R}^2$ is some embedding of S into D such that distinct points can be separated by ε-balls.

Lemma 4.3.7 (Smiley Face Lemma, sorry). *For any fixed S, ε, D as above, there is some $\delta > 0$ such that n random points ina $\sqrt{n} \times \sqrt{n}$ square have δn many ε-D copies of S (with* *high probability).*

4.3.3 Lovasz Local Lemma

Suppose $B_1, \dots, B_n$ are independent events with $\Pr(B_i) \leq 1$. It's pretty easy to see that $\Pr(\bigcap B_i^C) > 0$. Essentially, we think of the B_i as *bad events*: events we *don't* want to see happen.

We want to extend this. In what way exactly? Here's a motivating question.

In a k-uniform hypergraph where each edge intersects at most D others, what condition on k makes the hypergraph 2-colourable?

What we'd want to do is randomly colour vertices: each edge of f is monochromatic with probability $\Pr(B_f) = \frac{1}{2^{k-1}}$. We would then like to apply some version of the fact that $\Pr(\bigcap B_i^C) > 0$.

Definition 4.3.8 (Dependency Graph). Given a collection $\mathcal{A}$ of events, a **dependency graph on** $\mathcal{A}$ is a graph with vertices $\{v_A \mid A \in \mathcal{A}\}$ such that $v_A \not\sim v_{B_1}, \dots, v_{B_n}$ implies that A is independent of $B_1, \dots, B_n$.

Another way of phrasing this is that adjacency represents dependence (ie, lack of independence).

Local Notation (Dependence). Borrowing from our use of $\sim$ for adjacency, we will denote that events A and B are dependent (ie, not independent) by $A \sim B$.

Theorem 4.3.9 (Lovasz Local Lemma). *Given a collection $\mathcal{A}$ of events and real numbers $0 < x_A \leq 1$ such that for all $A \in \mathcal{A}$,*

$$\Pr(A) \leq x_A \prod_{B \sim A} (1 - x_B)$$

we can conclude that

$$\Pr\left(\bigcap_{A \in \mathcal{A}} A^C\right) > 0$$

Proof. sorry □