Probabilistic Methods

Yanheng Wang

Table of contents

Chapter 1
Probability and Counting

1.1 Intersecting Systems

$\mathscr{A} \subseteq 2^{[n]}$ is called an *intersecting system* if $A \cap A' \neq \emptyset$ for all $A, A' \in \mathscr{A}$. How large can it be? At most 2^{n-1}: otherwise there must be some set $S \subseteq [n]$ such that $S \in \mathscr{A}$ and $\overline{S} \in \mathscr{A}$.

What if we require the system to include only k-element sets? How large can it be then? This is answered by the Erdős-Ko-Rado theorem.

Theorem. If $\mathscr{A} \subseteq \binom{[n]}{k}$ is an intersecting system, then its cardinality is at most $\frac{k}{n}\binom{n}{k}$.

Proof. Rephrased in probabilistic language, the theorem says that a uniform random set $R \in \binom{[n]}{k}$ lands in the system $\mathscr{A}$ with probability at most k/n.

To establish the claim, we refine the probability space as follows: Sample a permutation $\pi : [n] \to [n]$ and a value $r \in [n]$ uniformly, then take $R \coloneqq \{\pi(r), \ldots, \pi(r+k-1)\}$, where the addition "overflows" when exceeding n. Indeed the R is uniform over $\binom{[n]}{k}$, because any condition on r is overridden by the uniform π.

Now let us study $\mathbb{P}(R \in \mathscr{A})$. Conditioned on π, there are n possibilities of r, but at most k of them can lead to $R \in \mathscr{A}$. Suppose to the contrary that $\{\pi(r), \ldots, \pi(r+k-1)\} \in \mathscr{A}$ for $k+1$ distinct r's, then surely two of them are disjoint, contradicting the assumption of $\mathscr{A}$. Therefore we must have $\mathbb{P}(R \in \mathscr{A} \mid \pi) \leqslant k/n$, which of course implies $\mathbb{P}(R \in \mathscr{A}) \leqslant k/n$. □

Remark. Probabilistic argument on a uniform space can always be cast as counting: probability becomes fractions of counts, independence becomes product (or independent sums), conditioning is the same as counting a subspace, etc. That said, counting jargons produce less elegant proofs. Were we to apply it in the proof above, we would have to model the two-stage sampling as a Cartesian product and take care of messy factorials everywhere.

1.2 Kraft's Inequality

In coding theory, a finite set of bit strings $C \subseteq \{0,1\}^*$ is called *prefix-free* if x is not a prefix of y for any distinct $x, y \in C$. We can imagine a binary tree whose nodes are bit strings, with the root being the empty string and every node x having two children $x0$ and $x1$. The set C is prefix-free if the path from any $y \in C$ to the root does not contain another $x \in C$.

Lemma. If C is prefix-free then $\sum_{x \in C} 2^{-|x|} \leqslant 1$, where $|x|$ denotes the length of bit string x.

Probabilistic proof. Let $\ell \coloneqq \max_{x \in C} |x|$. Sample a uniform random bit string $r \in \{0,1\}^\ell$. For each $x \in C$ define event $E_x \coloneqq \{x \text{ is prefix of } r\}$. Observe that the events are disjoint because r, by prefix-freeness, cannot pass two $x, y \in C$ on its way to the root. Hence we have

$$1 \geqslant \sum_{x \in C} \mathbb{P}(E_x) = \sum_{x \in C} 2^{-|x|}. \qquad \square$$

Combinatorial proof. Let $\ell := \max_{x\in C}|x|$. Whenever there is $x \in C$ of length $|x| < \ell$, we remove it and add $x0, x1$ to C. Due to prefix-freeness, $x0$ and $x1$ were not in C, so the sum $\sum_{x\in C} 2^{-|x|}$ shall not change. Moreover C remains prefix-free. So we may repeat the procedure until all $x \in C$ have length ℓ. In this final situation, $\sum_{x\in C} 2^{-|x|} = |C| \cdot 2^{-\ell} \leqslant 2^{\ell} \cdot 2^{-\ell} = 1$. □

1.3 Set-Pair Systems

Theorem. Assume a system of set pairs $\{(A_1, B_1), \ldots, (A_m, B_m)\}$ satisfies $|A_i| = a$, $|B_i| = b$, $A_i \cap B_i = \emptyset$ and $A_i \cap B_j \neq \emptyset$ for all distinct i, j. Then its cardinality is bounded by $m \leqslant \binom{a+b}{a}$.

For convenience in our proofs, we denote the ground set by $[n] := \bigcup_{i=1}^{m} (A_i \cup B_i)$.

Proof. Let us sample a random ordering of the ground set $[n]$. For each i define an event $E_i := \{A_i < B_i\}$, namely "all elements in A_i come before elements in B_i". Notice that the events are disjoint. Say $E_i = \{A_i < B_i\}$ happens, then in particular $(A_i \cap B_j) < (A_j \cap B_i)$. Since both sides are non-empty, they forbid the event $E_j = \{A_j < B_j\}$.

Given disjointness, we can write

$$1 \geqslant \sum_{i\in[m]} \mathbb{P}(E_i) = m \cdot \frac{a! \cdot b!}{(a+b)!}$$

and consequently $m \leqslant \binom{a+b}{a}$. □

1.4 Largest Anti-Chain in Hypercube

What is the lsrgest anti-chain in the poset $(2^{[n]}, \subseteq)$? Intuitively the middle level $\binom{[n]}{n/2}$ is the most effective packing of incomparable sets, so $\binom{n}{n/2}$ should be the answer. This is indeed true. Let us take an arbitrary anti-chain $\mathscr{A} = \{A_1, \ldots, A_m\}$, so that $A_i \not\subseteq A_j$ (equivalently, $A_i \setminus A_j \neq \emptyset$) for all $i \neq j$. We aim to show $m \leqslant \binom{n}{n/2}$.

Probabilistic proof. Similar to the proof for set-pair system. □

Combinatorial proof. The idea is to "lift" light sets and "sink" heavy sets to the middle level without collision. Let $\mathscr{L}$ collect the lightest sets (suppose their size equals $k < n/2$) in our anti-chain. Define

$$\mathscr{R} := \left\{R \in \binom{[n]}{k+1} : L \subseteq R \text{ for some } L \in \mathscr{L}\right\}.$$

Note that $\mathscr{R}$ does not intersect $\mathscr{A}$.

Consider a bipartite graph on $\mathscr{L} \uplus \mathscr{R}$ where edges represent inclusion relation. For any $\mathscr{L}' \subseteq \mathscr{L}$, denote by $\mathscr{R}' \subseteq \mathscr{R}$ its neighbourhood. Then we have

$$(n-k) \cdot |\mathscr{L}'| = \#\text{edges in between} \leqslant (k+1) \cdot |\mathscr{R}'|,$$

thus $|\mathscr{R}'| \geqslant |\mathscr{L}'|$. By Hall's theorem, we may find a matching between $\mathscr{L}$ and $\mathscr{R}$ that saturates the former. This means we may "lift" every set in $\mathscr{L}$ to its partner in $\mathscr{R}$ (which were not yet occupied by $\mathscr{A}$), resulting in a new anti-chain of length m.

We repeat the procedure until all sets in the anti-chain live on or above the middle level. With a symmetric argument, we may "sink" all heavy sets so that they live on the middle level. Then we may conclude $m \leqslant \binom{n}{n/2}$. □

1.5 Two-Colourability of Hypergraphs

Let G be a k-uniform hypergraph. A vertex colouring is proper if no edge is monochromatic. When $k = 2$ (thus G is a graph), we can decide in linear time if G is 2-colourable. In sharp contrast, for all $k \geqslant 3$ the task is NP-complete.

To gain more insights, we study m_k, the minimum number of edges we need to build a non-2-colourable k-uniform hypergraph. Clearly $m_2 = 3$. Below we will apply probabilistic arguments to upper and lower bound m_k in general.

Lemma. $m_k \leqslant 3k^2 2^k$.

Proof. Our goal is finding a non-2-colourable $G = (V, E)$ with at most $m := 3k^2 2^k$ edges. We take the vertex set as $V = [n] = [k^2]$. To construct the edge set E, we include m independent samples from $\binom{V}{k}$, potentially with repetitions.

Next we show that the random G is non-2-colourable with positive probability. Fix a 2-colouring of V, which induces a partition $V_0 \cup V_1$. Without loss of generality assume $|V_0| \geqslant n/2$. A sample $e \in E$ is monochromatic iff $e \subseteq V_0$ or $e \subseteq V_1$, which happens with probability at least $\mathbb{P}(e \subseteq V_0) = \binom{n/2}{k} / \binom{n}{k}$. Using independence, the (fixed) 2-colouring is proper with probability at most $\left(1 - \binom{n/2}{k} / \binom{n}{k}\right)^m$. By a union bound over all 2^n 2-colourings and some calculations, we see that

$$\mathbb{P}(\exists \text{proper 2-colouring of } G) < 1.$$

Hence the random G has positive chance of being non-2-colourable. □

For lower bounds, we need to show that any G with limited number of edges is 2-colourable. Let us start simple:

Lemma. $m_k > 2^{k-1}$.

Proof. Take any G with $m \leqslant 2^{k-1}$ edges. We 2-colour each vertex in G independently and uniformly. Then any particular edge e is monochromatic with probability $\mathbb{P}(B_e) = 1/2^{k-1}$. By a union bound, the colouring is improper with probability $\mathbb{P}(\exists e : B_e) < m/2^{k-1} \leqslant 1$. (The union bound was strict because the events intersect.) Hence there exists a proper 2-colouring of G, meaning G is 2-colourable. □

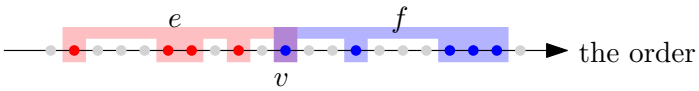
To improve the result, we produce a 2-colouring more strategically so that we succeed with a higher chance.

Lemma. $m_k > \left(\frac{k}{4\pi}\right)^{1/4} 2^k$.

Proof. Take any G with $m \leqslant \left(\frac{k}{4\pi}\right)^{1/4} 2^k$ edges. We try to 2-colour it by the following procedure. First, we randomly order the vertices. Second, we follow this order and assign red/blue to each vertex. Colour red enjoys priority: if assigning red does not cause immediate violation then we do so; otherwise we assign blue.

Upon termination, every vertex is coloured either red or blue, and there is no monochromatic red edge. We will show that, with decent probability, there is no monochromatic blue edge either.

Let B be the event that there is a monochromatic blue edge. What does it imply? Say edge f is blue. Let $v \in f$ be its first vertex that got coloured. The vertex v got blue only because there is another edge $e \ni v$ where every vertex in $e \setminus \{v\}$ was coloured red at an earlier time.



This motivates us to abstract an event $B_{ef} := \{e \text{ comes earliear than } f\}$ for every pair of edges $e, f : |e \cap f| = 1$. Then by our discussion,

$$B \subseteq \bigcup_{|e \cap f| = 1} B_{ef}.$$

Apparently $\mathbb{P}(B_{ef}) = \frac{[(k-1)!]^2}{(2k-1)!}$, so by a union bound and Stirling approximation we have

$$\mathbb{P}(B) < m^2 \cdot \frac{[(k-1)!]^2}{(2k-1)!} \leqslant 1. \qquad \square$$

Somewhat technically, we can squeeze more from the union bound by "folding" some of the overlapping events.

Lemma. $m_k > \sqrt{\frac{k}{\log k}}\, 2^{k-2}$.

Proof. We think of the random order as putting indepedent uniform value $x_v \in [0,1]$ on every vertex v. Those who get smaller values are coloured earlier, using the previous rule.

Fix a parameter $\delta := \frac{\log k}{k}$. For each vertex v we introduce an event

$$C_v := \left\{ \left| x_v - \frac{1}{2} \right| \leqslant \frac{\delta}{2} \right\}$$

stating that x_v is around the centre. For every pair $e \cap f = \{v\}$ we define B_{ef} as before and decompose $B_{ef} = (B_{ef} \setminus C_v) \cup (B_{ef} \cap C_v)$. Observe that $B_{ef} \setminus C_v$ implies one of the following two events:

$$L_e := \left\{ \forall u \in e,\, x_u \leqslant \frac{1-\delta}{2} \right\}, \quad R_f := \left\{ \forall u \in f,\, x_u \geqslant \frac{1+\delta}{2} \right\}.$$

Therefore

$$B \subseteq \left(\bigcup_e L_e \right) \cup \left(\bigcup_f R_f \right) \cup \left(\bigcup_{|e \cap f| = 1} (B_{ef} \cap C_v) \right).$$

This effectively "folds" many overlapping events to the sides e and f. Concretely, we have reduced some union of m^2 terms to only m terms.

It is easy to see that

$$\mathbb{P}(L_e) = \mathbb{P}(R_f) = \left(\frac{1-\delta}{2} \right)^k$$

and

$$\mathbb{P}(B_{ef} \cap C_v) = \int_{\frac{1-\delta}{2}}^{\frac{1+\delta}{2}} \mathbb{P}(B_{ef} \cap C_v \,|\, x_v)\, \mathrm{d}x_v = \int_{\frac{1-\delta}{2}}^{\frac{1+\delta}{2}} x_v^{k-1} (1-x_v)^{k-1}\, \mathrm{d}x_v \leqslant \frac{\delta}{4^{k-1}}.$$

So by a union bound we derive

$$\mathbb{P}(B) \leqslant 2m \left(\frac{1-\delta}{2} \right)^k + m^2 \frac{\delta}{4^{k-1}} \leqslant \sqrt{\frac{k}{4 \log k}} \cdot e^{-\delta k} + \frac{\delta k}{4 \log k} < 1. \qquad \square$$

1.6 List Colouring of Bipartite Graphs

Theorem. Let $G = (V, E)$ be a bipartite graph on n vertices. Every vertex $v \in V$ is associated with a colour list L_v of size at least $\log n$. Then there always exists a proper colouring σ such that $\sigma(v) \in L_v$ for all $v \in V$.

Perhaps the most "obvious" probabilistic argument goes as follows: For each vertex $v \in V$ pick a random colour $\sigma(v) \in L_v$ from its list, and hope that σ is proper with positive probability. Well, it *ought to* be positive after all—assuming the theorem is true. But to *argue* about it is another story. In fact, the probability is too small to be controlled by analytical tools. Consider for instance $G := K_{n/2,n/2}$ and $L_v := [\log n]$ for all v. Even for a fixed $v \in V$ alone, the probability that all its $n/2$ neighbours get a colour different from $\sigma(v)$ is $\left(1 - \frac{1}{\log n}\right)^{n/2}$, a tiny quantity. It is rather unclear how we may lower bound the probability that σ is proper.

The lesson is that we should design the random experiment more wisely. Recall the structure of bipartite graphs: there is no edge inside each part, that is no direct conflict if we colour every vertex in the same part red whenever available. Generally, we should couple the colours inside each part as much as possible in order to "make space" for the other part. This suggests the procedure below:

> Assume $V = A \uplus B$. We permute the colours in $\bigcup_{a \in A} L_a$ randomly, and denote the resulting colour sequence by $(c_1, \dots, c_N)$. We colour each $a \in A$ by c_i where $i := \min\{i : c_i \in L_a\}$, namely the first colour that fits. After we finish part A, we try to colour each $b \in B$ by a candidate in L_b not yet taken by its neighbours.

It is usually tiresome (though possible) to deal with permutations. So our final proof uses a simplified experiment that retains much of the spirit.

Proof. Assume $V = A \uplus B$. We sample a random subset $L \subseteq \bigcup_{v \in V} L_v$ by including each colour independently with probability $1/2$. We try to colour each $a \in A$ by a candidate in $L_a \cap L$, and each $b \in B$ by a candidate in $L_b \setminus L$. If all attempts succeed, then the final colouring must be proper.

Next we bound the failure probability. For any fixed $a \in A$, we have $\mathbb{P}(L_a \cap L = \emptyset) \leqslant 2^{-\log n} = 1/n$. For any fixed $b \in B$, we have $\mathbb{P}(L_b \setminus L = \emptyset) = 2^{-\log n} = 1/n$. So by a union bound, the failure probability is less than

$$\sum_{a \in A} \frac{1}{n} + \sum_{b \in B} \frac{1}{n} = 1.$$

This means the procedure succeeds with positive probability, as desired. □

Chapter 2
Probability Bounds via Concentration

Concentration means that a random variable X takes values from a small range (compared to its support) with high probability. The most common concentration is around the expectation of the variable; that is $\mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) \approx 0$ for a relatively small ε. Below we list several scenarios that admit this type of concentration.

Chebyshev's Inequality. For any random variable X, we have

$$\mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) \leqslant \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

Therefore, X is concentrated around its expectation when $\sqrt{\mathrm{Var}(X)} \ll \varepsilon \ll \mathbb{E}(X)$.

Chernoff's Inequality. If $X = \sum_{i=1}^{n} X_i$ is a sum of independent Bernoulli variables, then

$$\mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) \leqslant 2 \cdot \exp\left(\frac{-\varepsilon^2}{3\,\mathbb{E}(X)}\right).$$

Therefore, X is concentrated around its expectation when $\sqrt{\mathbb{E}(X)} \ll \varepsilon \ll \mathbb{E}(X)$.

Azuma's Inequality. Let $X = X(\omega_1, \ldots, \omega_n)$ be a random variable in an n-dimensional product probability space. Suppose that it is c-Lipschitz, i.e. the value $X(\omega_1, \ldots, \omega_n)$ can change by at most c if we alter a single coordinate ω_i in the outcome. Then

$$\mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) \leqslant 2 \cdot \exp\left(\frac{-\varepsilon^2}{2c\,n}\right).$$

Therefore, X is concentrated around its expectation when $\sqrt{n} \ll \varepsilon \ll \mathbb{E}(X)$.

None of these inequalities are particularly deep; they all depend on the elementary Markov's inequality and can be understood as some kind of central limit theorem. We shall omit the proofs. There also exist other types of concentration: Talagrand's inequality, for example, exhibits concentration around median.

Concentration usually serves as a piece in a longer proof. It may allow us to show that every particular bad event B_i happens with extremely small probability—so small that we can apply union bound to avoid all bad events B_i with high probability.

In the following, however, we present two examples where concentration plays central and direct role.

2.1 Sum Collision

Theorem. Suppose $A = \{a_1, \ldots, a_m\} \subseteq [n]$. If $m \geqslant \log n + \frac{1}{2}\log\log n + \Theta(1)$ then there exist distinct $I, I' \subseteq [m]$ such that $\sum_{i \in I} a_i = \sum_{i \in I'} a_i$.

Proof. Sample independent variables $X_1,\dots,X_m \sim \mathrm{Bern}(1/2)$, which encode a random index set $I \subseteq [m]$. Consider the subset sum $X \coloneqq \sum_{i\in I} a_i = \sum_{i=1}^m a_i X_i$. Clearly

$$\mathrm{Var}(X) = \sum_{i=1}^m \frac{a_i^2}{4} < \frac{m n^2}{4},$$

so by Chebyshev's inequality,

$$\mathbb{P}\left(|X - \mathbb{E}(X)| > \sqrt{m}\, n\right) \leqslant \frac{1}{4}.$$

That is, at least 75% of the total 2^m subset sums land in the range $\mathbb{E}(X) \pm \sqrt{m}\, n$.

On the other hand, we compare

$$\frac{3}{4} \cdot 2^m > 2\sqrt{m}\, n$$

provided $m \geqslant \log n + \frac{1}{2}\log\log n + \Theta(1)$. So there must be a sum collision by the pigeon-hole principle. □

2.2 Intersection Size Collision

Theorem. Let $A_1,\dots,A_m$ be subsets of $[n]$. If $m \leqslant \frac{n}{\log n}$ then there exist distinct $S, S' \subseteq [n]$ such that $|S \cap A_i| = |S' \cap A_i|$ for all $i \in [m]$.

Proof. Sample a uniform random set $S \subseteq [n]$ and consider its size vector $(Y_1,\dots,Y_m)$ where $Y_i \coloneqq |S \cap A_i|$. Note that $Y_i \sim \mathrm{Bin}(|A_i|, 1/2)$, so by Chernoff's inequality,

$$\mathbb{P}\left(|Y_i - \mathbb{E}(Y_i)| > \sqrt{2n\log n}\right) \leqslant 2 \cdot \exp\left(\frac{-2n\log n}{3|A_i|/2}\right) \leqslant \frac{2}{n^{4/3}}.$$

Applying a union bound over all $i \in [m]$, we see that $(Y_1,\dots,Y_m)$ is confined in a box of size $\left(2\sqrt{2n\log n}\right)^m$ with probability at least $1 - \frac{2m}{n^{4/3}} \gg \frac{1}{2}$. It is straightforward to verify that

$$\frac{1}{2} \cdot 2^n > \left(2\sqrt{2n\log n}\right)^m$$

provided $m \leqslant \frac{n}{\log n}$. Hence, by pigeon-hole principle, there must exist two sets whose size vectors collide. □

Chapter 3
Probability Bounds via Locality

In many contexts we are searching for a structure that avoids certain bad events $B_1, \dots, B_n$. To employ a probabilistic argument, we may sample a structure at random and show that $\mathbb{P}(\bigcap_{i=1}^n \overline{B_i}) > 0$. The previous chapters extensively applied the union bound:

$$\mathbb{P}\left(\bigcap_{i=1}^n \overline{B_i}\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^n B_i\right) \geqslant 1 - \sum_{i=1}^n \mathbb{P}(B_i).$$

When the number of events n and the probabilities $\mathbb{P}(B_i)$ are small, it produces a positive result. We also saw refined techniques that reduce the number n and thus the loss.

But generally the union bound is rather clumsy as we are competing a long summation with 1. Let us say $\mathbb{P}(B_i) \leqslant p$ for all i and we are interested in asymptotics when $n \to \infty$. Only in very rare situations can we argue $np \to c \in (0,1)$; typically np either diverges or converges to 0. This means that the union bound either is useless, or proves something excessively strong.

Can we develop better tools to show $\mathbb{P}(\bigcap_{i=1}^n \overline{B_i}) > 0$? As a warm-up, suppose $B_1, \dots, B_n$ are mutually independent and $\mathbb{P}(B_i) \leqslant p < 1$ for all i, then

$$\mathbb{P}\left(\bigcap_{i=1}^n \overline{B_i}\right) = \prod_{i=1}^n \mathbb{P}(\overline{B_i}) \geqslant (1-p)^n > 0.$$

Sounds good. But independence is boring because then the existence of desired structure is trivial even without a probabilistic argument.

What if we relax independence to "almost independence"? Intuitively the argument should still work (though subject to some loss). The intuition is captured by the famous lemma below, whose proof is as natural as one could think of.

Lovász Local Lemma. Suppose we have events $B_1, \dots, B_n$ where each B_i is mutually independent of all but d other events. Assume $\mathbb{P}(B_i) \leqslant \frac{1}{\mathrm{e}(d+1)}$ for all i. Then

$$\mathbb{P}\left(\bigcap_{i=1}^n \overline{B_i}\right) > \left(1 - \frac{1}{d+1}\right)^n > 0.$$

Proof. By chain rule,

$$\mathbb{P}\left(\bigcap_{i=1}^n \overline{B_i}\right) = \prod_{i=1}^n \mathbb{P}(\overline{B_i} \mid \overline{B_1} \cap \cdots \cap \overline{B_{i-1}}),$$

so our goal is to lower bound each of the conditional probabilities. We claim that, for all i,

$$\mathbb{P}(B_i \mid \overline{B_1} \cap \cdots \cap \overline{B_{i-1}}) < \frac{1}{d+1}.$$

In other words, the conditions, though very long, can only bloat the bound on $\mathbb{P}(B_i)$ by a factor of at most e. This is roughly because only d of the conditions "matter".

We show the claim by induction on the number of conditions. Assume, by proper renaming, that B_i is mutually independent with $B_{d+1},\dots,B_{i-1}$. Let us realise the conditions in two steps: First condition on the "irrelevant" part $\overline{B_{d+1}}\cap\cdots\cap\overline{B_{i-1}}$, and then impose the extra condition $\overline{B_1}\cap\cdots\cap\overline{B_d}$. In formula,

$$\mathbb{P}\left(B_i \mid \overline{B_1}\cap\cdots\cap\overline{B_{i-1}}\right) = \frac{\mathbb{P}\left(B_i\cap\overline{B_1}\cap\cdots\cap\overline{B_d} \mid \overline{B_{d+1}}\cap\cdots\cap\overline{B_{i-1}}\right)}{\mathbb{P}\left(\overline{B_1}\cap\cdots\cap\overline{B_d} \mid \overline{B_{d+1}}\cap\cdots\cap\overline{B_{i-1}}\right)}.$$

Its numerator is at most

$$\mathbb{P}\left(B_i \mid \overline{B_{d+1}}\cap\cdots\cap\overline{B_{i-1}}\right) = \mathbb{P}\left(B_i\right) < \frac{1}{\mathrm{e}\,(d+1)}$$

by assumption. And its denominator is, using chain rule,

$$\prod_{j=1}^{d} \mathbb{P}\left(\overline{B_j} \mid \overline{B_1}\cap\cdots\cap\overline{B_{j-1}}\cap\overline{B_{d+1}}\cap\cdots\cap\overline{B_{i-1}}\right) > \left(1-\frac{1}{d+1}\right)^d > \frac{1}{\mathrm{e}}$$

where the first inequality follows from induction hypothesis. Taking their quotient completes the induction of the claim and finishes the lemma. □

3.1 Two-Colourability of Hypergraphs

Lemma. If $G=(V,E)$ is a k-uniform hypergraph where every edge intersects at most $d\leqslant \mathrm{e}^{-1}2^{k-1}-1$ other edges, then G is two-colourable.

Proof. We 2-colour each vertex in G uniformly and independently. Let B_e be the event that edge e is monochromatic. Then B_e is independent with all events $B_{e'}$ unless $e\cap e'\neq\emptyset$; thus the level of dependency is d. Since $\mathbb{P}\left(B_e\right)=2^{1-k}\leqslant\frac{1}{\mathrm{e}\,(d+1)}$, Lovász local lemma applies and shows the existence of a proper 2-colouring. □

3.2 Directed Cycles Modulo d

Theorem. Any Δ-regular digraph contains a directed cycle of length divisible by d, provided $d\leqslant\frac{\Delta}{2\log(\Delta+1)}$.

Proof. Take $G=(V,E)$ to be any Δ-regular digraph. We sample a colouring $\sigma:V\to\mathbb{Z}_d$ uniformly at random and define a subgraph $H\coloneqq(V,F)$, where $F\coloneqq\{uv\in E:\sigma(u)+1=\sigma(v)\}$. We claim that $\deg^+_H(u)>0$ for all $u\in V$ with positive probability. Note that it would imply the existence of a cycle isomorphic to $\mathbb{Z}_d$: Just start from any vertex and follow the F-edges until hitting a vertex visited before.

Now it remains to prove the claim. For each vertex $u\in V$ we associate a bad event $B_u\coloneqq\{\deg^+_H(u)=0\}$. It relies on, and only on, the colours $\sigma(v)$ for $v\in N^+_G(u)\cup\{u\}$. Therefore, B_u is independent with all other $B_{u'}$ unless $N^+_G(u)\cup\{u\}$ and $N^+_G(u')\cup\{u'\}$ intersect. So the level of dependency is at most $\Delta\left(\Delta+1\right)$. On the other hand,

$$\mathbb{P}\left(B_u\right)=\left(1-\frac{1}{d}\right)^{\Delta}<\mathrm{e}^{-2\log(\Delta+1)}<\frac{1}{\mathrm{e}\left[\Delta\left(\Delta+1\right)+1\right]},$$

so we may apply Lovász local lemma to conclude. □

3.3 Linear Forest Decomposition

A forest is *linear* if all its constituents are paths. Given a graph $G=(V,E)$ with maximum degree Δ, we want to partition its edges into a small number of linear forests: $E=F_1\cup\cdots\cup F_t$. How small can t be?

An easy lower bound is supplied by $t \geqslant \left\lceil \frac{\Delta}{2} \right\rceil$, since every linear forest can cover at most two incident edges to a Δ-degree vertex. When G is Δ-regular, we actually need $t \geqslant \left\lceil \frac{\Delta+1}{2} \right\rceil$, since each forest can consume up to $n-1$ edges out of a total of $\frac{n\Delta}{2}$ edges.

Linear Arboricity Conjecture. Any graph with maximum degree Δ can be partitioned into $t \leqslant \left\lceil \frac{\Delta+1}{2} \right\rceil$ linear forests.

The conjecture remains open to date. But a fruitful direction looks into its *directed* variant. Let us call a directed forest *linear* if its constituents are directed paths.

Conjecture. Any Δ-regular digraph can be partitioned into $\Delta+1$ linear directed forests.

This latter conjecture implies the linear arboricity conjecture for even Δ:

1. Given a graph G with even maximum degree Δ, we "complete" it into a Δ-regular supergraph as follows. Whenever a vertex v has $\deg(v) \leqslant \Delta-2$, we create two companion vertices v_1, v_2 and a $K_{\Delta-2}$. Each v_i $(i=1,2)$ is connected with v and all vertices in the $K_{\Delta-2}$. The operation increases $\deg(v)$ by two, yet all the new vertices have degree Δ. Eventually every vertex would have degree $\Delta-1$ or Δ. Then we double the graph and connect those vertices with degree $\Delta-1$ in pairs.
2. Now we have a Δ-regular graph. Since Δ is even, we can find an Eulerian tour and orient the edges accordingly, and get a $(\Delta/2)$-regular digraph. We apply the latter conjecture to partition it into $\frac{\Delta}{2}+1 = \left\lceil \frac{\Delta+1}{2} \right\rceil$ directed linear forests. They of course induce the same amount of linear forests in the original graph G.

It turns out the latter conjecture is quite close to a full resolution:

Theorem. Any Δ-regular digraph $G=(V,E)$ with $\operatorname{girth}(G) \geqslant 4\,\mathrm{e}\Delta$ can be partitioned into $\Delta+1$ linear directed forests.

Proof. First we partition $E = \mathscr{C}_1 \cup \cdots \cup \mathscr{C}_\Delta$ where each $\mathscr{C}_i$ is a cycle cover. This is a consequence of Hall's theorem. Consider a bipartite graph with one copy of V on each side, where we connect the left-u to the right-v iff $uv \in E$. Clearly the bipartite graph is Δ-regular, so Hall's theorem guarantees a perfect matching, which corresponds to a cycle cover of the digraph. We remove these edges and repeat the procedure (on the regular subgraph). Eventually we shall partition the entire edge set E into cycle covers $\mathscr{C}_1, \ldots, \mathscr{C}_\Delta$. We collect all these cycles into $\mathscr{C} \coloneqq \bigcup_{i=1}^{\Delta} \mathscr{C}_i$. Keep in mind that every edge is covered exactly once.

Next we will "break" all cycles into paths by one "cut", thereby producing linear forests. Recall that each cycle $C \in \mathscr{C}$ contains at least $\operatorname{girth}(G) \geqslant 4\,\mathrm{e}\Delta$ edges. For our argument we assume that it contains exactly $4\,\mathrm{e}\Delta$; otherwise we just ignore the extra. Choose one edge from C uniformly at random. Do this independently for all cycles. We claim that F, the set of chosen edges, is a matching with positive probability. The theorem would immediately follow since $F, (\mathscr{C}_1 \setminus F), \ldots, (\mathscr{C}_\Delta \setminus F)$ partitions E into $\Delta+1$ linear forests.

In order to show the claim, for every pair of incident edges $e, f \in E$ that belong to different cycles, define a bad event $B_{ef} \coloneqq \{e, f \in F\}$. Note that

$$\mathbb{P}(F \text{ is a matching}) = \mathbb{P}\left(\bigcap_{e,f} \overline{B_{ef}} \right).$$

Let us analyse the dependency of these events. Assume $e \in C$ and $f \in D$, then B_{ef} is only related to the events $B_{e'f'}$ where $e' \in C \cup D$ or $f' \in C \cup D$. In the first case there are $|C \cup D| \leqslant 8\,\mathrm{e}\Delta$ choices for e' and then at most Δ choices for an incident f', thus amounting to $8\,\mathrm{e}\Delta^2$ combinations. A symmetric calculation works for the second case. Therefore, B_{ef} is indepedent of all but at most $d \coloneqq 16\,\mathrm{e}\Delta^2 - 1$ other events (where the minus 1 excludes itself).

Since $\mathbb{P}(B_{ef}) = \frac{1}{(4\,\mathrm{e}\Delta)^2} = \frac{1}{16\,\mathrm{e}^2\Delta^2} \leqslant \frac{1}{\mathrm{e}(d+1)}$, the claim is established by Lovász local lemma. □

Chapter 4
Probability Bounds via Correlation

Locality is not always present in combinatorial problems. Think of events $\{\chi(G) \geqslant 5\}$ and $\{G \text{ is Hamiltonian}\}$ in the random graph model $G \sim G(n,p)$ that are notoriously global and tricky to handle. Still they seem to be positively correlated, in the sense that both a high chromatic number and the existence of Hamiltonian cycles "favour" dense graphs. Intuitively, the probability that the two events happen simultaneously should be larger than the product probability that each happens individually.

Prior to a formal proof, let us reflect on the notion of "positive correlation". Why do we *feel* that two random variables (or events) f and g are positively correlated? One thesis is that the underlying probability space is partially ordered, and both f and g respect that order. So when we go up in the probability space, the values of f and g shall increase. This creates an illusion that the variables are working synchronously.

In the random graph example, the partial order is the "supergraph relation", respected by both events $\{\chi(G) \geqslant 5\}$ and $\{G \text{ is Hamiltonian}\}$.

Now we work our way to a formal proof. First we consider one-dimensional and totally ordered probability space. Then we will generalise to product space where each dimension is totally ordered. (That includes the random graph model.)

Chebyshev's Sum Inequality. Let x be any random variable, and $f, g : \mathbb{R} \to \mathbb{R}$ be increasing functions. Then $\mathbb{E}[f(x)\,g(x)] \geqslant \mathbb{E}[f(x)] \cdot \mathbb{E}[g(x)]$.

Proof. By monotonicity we have $(f(a) - f(b)) \cdot (g(a) - g(b)) \geqslant 0$ for all $a, b \in \mathbb{R}$. Let us sample independent a, b as distributed as x. Then

$$\begin{aligned} 0 &\leqslant \mathbb{E}[(f(a) - f(b)) \cdot (g(a) - g(b))] \\ &= \mathbb{E}[f(a)\,g(a)] + \mathbb{E}[f(b)\,g(b)] - \mathbb{E}[f(a)] \cdot \mathbb{E}[g(b)] - \mathbb{E}[f(b)] \cdot \mathbb{E}[g(a)] \\ &= 2 \cdot \mathbb{E}[f(x)\,g(x)] - 2 \cdot \mathbb{E}[f(x)] \cdot \mathbb{E}[g(x)] \end{aligned}$$

and the claim follows. □

FKG Inequality. Let $\boldsymbol{x} := (x_1, \ldots, x_n)$ be independent random variables. Suppose $f, g : \mathbb{R}^n \to \mathbb{R}$ are coordinate-wise increasing functions. Then $\mathbb{E}[f(\boldsymbol{x})\,g(\boldsymbol{x})] \geqslant \mathbb{E}[f(\boldsymbol{x})] \cdot \mathbb{E}[g(\boldsymbol{x})]$.

Proof. By induction on n. The base case $n = 1$ is simply Chebyshev's sum inequality. For the induction step, let us consider $n+1$ independent variables $(\boldsymbol{x}, y) = (x_1, \ldots, x_n, y)$ and condition on y. Both functions $\boldsymbol{x} \mapsto f(\boldsymbol{x}, y)$ and $\boldsymbol{x} \mapsto g(\boldsymbol{x}, y)$ are coordinate-wise increasing on $\mathbb{R}^n$, so by induction hypothesis we have

$$\begin{aligned} \mathop{\mathbb{E}}_{\boldsymbol{x}}[f(\boldsymbol{x}, y)\,g(\boldsymbol{x}, y) \,|\, y] &\geqslant \mathop{\mathbb{E}}_{\boldsymbol{x}}[f(\boldsymbol{x}, y) \,|\, y] \cdot \mathop{\mathbb{E}}_{\boldsymbol{x}}[g(\boldsymbol{x}, y) \,|\, y] \\ &=: \varphi(y)\,\gamma(y). \end{aligned}$$

Observe that $\varphi, \gamma : \mathbb{R} \to \mathbb{R}$ are increasing due to monotonicity of f, g and independence of the variables. Independence is crucial because otherwise the conditional space of $\boldsymbol{x}$ shall depend on y, making the conditional expectations incomparable across different y's.

Finally, we take expectation over y and derive

$$\begin{aligned}\mathbb{E}[f(\boldsymbol{x},y)\,g(\boldsymbol{x},y)] &= \mathop{\mathbb{E}}_{y}\Big[\mathop{\mathbb{E}}_{\boldsymbol{x}}[f(\boldsymbol{x},y)\,g(\boldsymbol{x},y)\,|\,y]\Big]\\ &\geqslant \mathop{\mathbb{E}}_{y}[\varphi(y)\,\gamma(y)]\\ &\geqslant \mathop{\mathbb{E}}_{y}[\varphi(y)]\cdot\mathop{\mathbb{E}}_{y}[\gamma(y)]\\ &= \mathbb{E}[f(\boldsymbol{x},y)]\cdot\mathbb{E}[g(\boldsymbol{x},y)]\end{aligned}$$

where the third line is, again, due to Chebyshev's sum inequality. □

Corollary. Let $\boldsymbol{x} \coloneqq (x_1,\dots,x_n)$ be independent random variables. Suppose $A,B\subseteq\mathbb{R}^n$ are upward closed sets. That is, if $\boldsymbol{a}\in A$ then every $\boldsymbol{a}'\geqslant\boldsymbol{a}$ are contained in A as well; similar for B. Then $\mathbb{P}(\boldsymbol{x}\in A\cap B)\geqslant\mathbb{P}(\boldsymbol{x}\in A)\cdot\mathbb{P}(\boldsymbol{x}\in B)$.

Proof. Take indicator functions $f(\boldsymbol{x})\coloneqq\mathbb{1}\{\boldsymbol{x}\in A\}$ and $g(\boldsymbol{x})\coloneqq\mathbb{1}\{\boldsymbol{x}\in B\}$. Both are coordinate-wise increasing since A,B are upward closed. □

There are several remarks in order:

- The FKG inequality (and its corollary) inductively applies to more than two functions (events), since the product of two increasing functions is still increasing.
- Nothing will break if both functions are decreasing—just turn the partial order upside down! On the other hand, if one function is increasing whereas the other is decreasing, then the inequality shall change direction.

Can we further generalise to any partially ordered space that are not necessarily product space? Both yes and no: there is a generalisation to distributive lattices equipped with log-supermodular probability measure; but not beyond.

4.1 Random Graphs Properties

Applying the corollary, our intuition that

$$\mathbb{P}(\chi(G)\geqslant 5 \quad\text{and}\quad G\text{ is Hamiltonian})\geqslant\mathbb{P}(\chi(G)\geqslant 5)\cdot\mathbb{P}(G\text{ is Hamiltonian})$$

is true for $G\sim G(n,p)$.

4.2 Intersecting Systems

Recall that $\mathscr{A}=\{A_1,\dots,A_m\}$ is an intersecting system if it contains pairwise intersecting sets. Denote $\mathscr{A}'\coloneqq\{\overline{A_1},\dots,\overline{A_m}\}$.

Theorem. Assume $\mathscr{A}\subseteq 2^{[n]}$. If both $\mathscr{A}$ and $\overline{\mathscr{A}}$ are intersecting systems, then $|\mathscr{A}|\leqslant 2^{n-2}$.

Proof. Let $\mathscr{B}$ and $\mathscr{C}$ be the upward and downward closures of $\mathscr{A}$, respectively. Observe that $\mathscr{B}$ is an intersecting system as $\mathscr{A}$ is. Similarly, $\mathscr{C}'$ is an intersecting system as $\mathscr{A}'$ is.

Now we sample a random set $S\subseteq[n]$ by including each element independently with probability $1/2$. Clearly S is uniformly distributed over all subsets of $[n]$. We have $\mathbb{P}(S\in\mathscr{B})=|\mathscr{B}|/2^n\leqslant 1/2$, and $\mathbb{P}(S\in\mathscr{C})=\mathbb{P}(S\in\mathscr{C}')=|\mathscr{C}'|/2^n\leqslant 1/2$. Note that the event $\{S\in\mathscr{B}\}$ is upward closed (with regard to the product space), whereas the event $\{S\in\mathscr{C}\}$ is downward closed. So by the corollary we know

$$\mathbb{P}(S\in\mathscr{A})=\mathbb{P}(S\in\mathscr{B}\cap\mathscr{C})\leqslant\mathbb{P}(S\in\mathscr{B})\cdot\mathbb{P}(S\in\mathscr{C})\leqslant\frac{1}{4}.$$

This immediately translates to $|\mathscr{A}|\leqslant 2^n/4$. □

Chapter 5
Expectation: Beyond Probability

It is not uncommon that we fail to obtain decent probability estimate after all. Expectation comes to aid: If a random variable has expectation t, then there must *exist* an outcome in which the variable evaluates to $\leqslant t$, as well as an outcome in which it evaluates to $\geqslant t$. They are sufficient for *existential* proofs, even though the probability that such outcomes occur is unknown or potentially very small.

5.1 Sets without Sum Capture

Theorem. For any finite $A \subseteq \mathbb{Z} \setminus \{0\}$, there exists a subset $B \subseteq A$ of size $|B| > |A|/3$ free from sum capture, i.e. $b + b' \neq b''$ for all $b, b', b'' \in B$.

Proof. Take prime $p > 2\max_{a \in A} |a|$ so that we may treat $A \subseteq \mathbb{Z}_p^*$. Define $I \coloneqq \left[\frac{p}{3}, \frac{2p}{3}\right)$, which is a set without sum capture. Consider the random set $B \coloneqq A \cap rI$ for a uniformly sampled $r \in \mathbb{Z}_p^*$ where we performed the multiplication modulo p. Observe that rI, thus also B, are free from sum capture because r is invertible.

What is the expected size of B? For any fixed element $a \in A$ we have

$$\mathbb{P}(a \in rI) = \mathbb{P}(r \in aI^{-1}) = \frac{|I|}{p-1} > \frac{1}{3}.$$

So $\mathbb{E}(|B|) > |A|/3$, meaning that there is a desired B of size at least $|A|/3$. □

5.2 Independence Number

Any graph on n vertices with maximum degree Δ has an independent set of size $\frac{n}{\Delta+1}$. We can construct it by iteratively picking a vertex v and deleting $v \cup N(v)$ from the graph.

However, if the graph contains only a few high-degree vertices, the bound is a gross underestimate. In such cases the average degree would be a more robust parameter.

Lemma. If $G = (V, E)$ has n vertices and average degree d, then $\alpha(G) \geqslant \frac{n}{2d}$.

Proof. Define a random set $S \subseteq V$ by independently including each vertex $v \in V$ with probability $p \coloneqq 1/d$. Let $F \coloneqq \{uv \in E : u, v \in S\}$ collect all the "conflicting" edges. For each $e \in F$ we remove one of its two ends from S so that the conflict is resolved. Let S' be the set S after the removal. Clearly it is an indepedent set of size $|S'| \geqslant |S| - |F|$. Since

$$\mathbb{E}(|S'|) \geqslant \mathbb{E}(|S|) - \mathbb{E}(|F|) = np - mp^2 = \frac{n}{d} - \frac{n}{2d} = \frac{n}{2d},$$

there must be an independent set of that size. □

It might be worth mentioning a deterministic argument for a weaker bound. Let $U \subseteq V$ collect all vertices with degree less than $2d$. Observe that $|U| \geqslant n/2$, for otherwise the total degree would exceed $(n/2) \cdot (2d) = nd$ which contradicts the average degree. Now we could find an independent set of size $\frac{|U|}{2d} \geqslant \frac{n}{4d}$ of the induced subgraph $G[U]$ by the deterministic procedure; it is also an independent set of G.

Arguably, the expectation approach is more powerful as it adapts to the graph stochastically. Let us take a closer look. The first half of the proof samples a standard random set regardless of the graph structure. The $\mathbb{E}(|F|)$ can be viewed as a union bound of the probability that "there is a conflicting edge". Well, this is at most mp^2, which is useless from a probability aspect. In fact, deriving a good probability bound might just be out of current reach. The ingenuity and power comes at the second half where we admit and remove those blemishes. This half heavily depends on F and thus the graph structure, and the output S' is highly customised. We basically understand nothing but the expectation $\mathbb{E}(|S'|)$; fortunately that is enough for an existential proof.

Using a suitably crafted probability space, we can gain finer control in the bound.

Lemma. Any graph $G = (V,E)$ satisfies $\alpha(G) \geqslant \sum_{v \in V} \frac{1}{\deg(v)+1}$.

Proof. Put a uniform random value $X_v \in [0,1]$ on each vertex $v \in V$, independent of each other. Define $S \coloneqq \{v \in V : \forall uv \in E, X_u > X_v\}$, namely the vertices whose values are locally minimal. By definition S is an indepedent set. Moreover $\mathbb{E}(|S|) = \sum_{v \in V} \mathbb{E}(X_v) = \sum_{v \in V} \frac{1}{\deg(v)+1}$. □

It directly implies the famous Turán theorem:

Corollary. If graph $G = (V,E)$ has n vertices and m edges, then $\alpha(G) \geqslant \frac{n^2}{2m+n}$.

Proof. Denote $a_v \coloneqq \sqrt{\deg(v)+1}$. By Cauchy-Schwarz, we have

$$\left(\sum_{v \in V} a_v^2\right)\left(\sum_{v \in V} \frac{1}{a_v^2}\right) \geqslant \left(\sum_{v \in V} a_v \cdot \frac{1}{a_v}\right)^2.$$

or just

$$(2m+n)\left(\sum_{v \in V} \frac{1}{\deg(v)+1}\right) \geqslant n^2.$$

Therefore

$$\alpha(G) \geqslant \left(\sum_{v \in V} \frac{1}{\deg(v)+1}\right) \geqslant \frac{n^2}{2m+n}.$$ □

As a remark, the bound is tight for the complement of Turán graphs. Basically these are just $\frac{n}{d+1}$ copies of K_{d+1} where $d \coloneqq 2m/n$ is the average degree. Some rounding details arise when the numbers are not integers, but we shall leave them out.

One naturally asks what happens if we ban certain dense substructure in the graph. For example, if we disallow occurrence of triangles, thereby forbidding cliques and so on, can we improve the bound on $\alpha(G)$?

Theorem. If a triangle-free graph $G = (V,E)$ has n vertices and maximum degree Δ, then $\alpha(G) > \frac{n \log \Delta}{4\Delta}$.

Proof. Let $I \subseteq V$ be a uniformly random independent set of G; we will show that $\mathbb{E}(|I|) > \frac{n \log \Delta}{4\Delta}$. For each vertex $v \in V$ we define random variables

$$\begin{aligned} X_v &\coloneqq \mathbb{1}\{v \in I\} \\ Y_v &\coloneqq |N(v) \cap I| = \sum_{u \in I} \mathbb{1}\{uv \in E\} \\ Z_v &\coloneqq X_v + \frac{Y_v}{\Delta}. \end{aligned}$$

Basically Z_v counts the contribution of neighbourhood $\{v\} \cup N(v)$, where $1/\Delta$ factor neutralises repeating counts. This two-part complication is beneficial because it balances the contention between v and $N(v)$, which allows us to apply a "meet in the middle" argument later. For now, observe that

$$\sum_{v \in V} Z_v = \sum_{v \in V} X_v + \frac{\sum_{v \in V} Y_v}{\Delta} = |I| + \frac{\sum_{u \in I} \deg(u)}{\Delta} \leqslant 2|I|,$$

so it suffices to lower bound $\mathbb{E}(Z_v) > \frac{\log \Delta}{4\Delta}$ for each individual $v \in V$.

Fix a vertex v. Define subset $J := I \setminus (\{v\} \cup N(v))$ to capture the far-away region. What is the conditional expectation $\mathbb{E}(X_v \mid J)$? Under the condition, it remains to expose which vertices in $\{v\} \cup N(v)$ are selected by I. We can

- either choose v but none of its neighbours;
- or choose *any* subset of $N(v) \setminus N(J)$ but not v.

Note that the second option allows *any* subset because the graph is triangle-free and thus all neighbours are non-adjacent.

There are $1 + 2^{|N(v) \setminus N(J)|} =: 1 + 2^s$ possibilities in total, all happening with equal probability. Therefore

$$\mathbb{E}(X_v \mid J) = \frac{1}{1+2^s} \geqslant \frac{1}{2^{s+1}}$$

and

$$\begin{aligned} \mathbb{E}(Y_v \mid J) &= \mathbb{E}(Y_v \mid v \in I, J) \cdot \mathbb{P}(v \in I \mid J) + \mathbb{E}(Y_v \mid v \notin I, J) \cdot \mathbb{P}(v \notin I \mid J) \\ &= 0 \cdot \frac{1}{1+2^s} + \frac{s}{2} \cdot \frac{2^s}{1+2^s} \\ &\geqslant \frac{s}{4}. \end{aligned}$$

They together imply

$$\mathbb{E}(Z_v \mid J) \geqslant \frac{1}{2^{s+1}} + \frac{s}{4\Delta} =: f(s).$$

Notice that $f(s+1) - f(s) = \frac{1}{4}\left(\frac{1}{\Delta} - \frac{1}{2^s}\right)$, so f is initially decreasing and then increasing. Its minimum is attained when $s = \log \Delta$, with value $\frac{2 + \log \Delta}{4\Delta}$. This means $\mathbb{E}(Z_v \mid J) > \frac{\log \Delta}{4\Delta}$ under any condition J. Hence $\mathbb{E}(Z_v) > \frac{\log \Delta}{4\Delta}$. □

5.3 Large Girth and Chromatic Number

Large girth means that the graph has a locally tree structure, and trees have chromatic number 2. It seems plausible to conjecture that large girth implies small chromatic number. Surprisingly, the intuition is very wrong, and long-term correlations do exist:

Theorem. For any $g, c \in \mathbb{N}$, there is a graph G with $\operatorname{girth}(G) > g$ yet $\chi(G) > c$.

Proof. Sample $G \sim G(n,p)$ where $p := n^{1/(2g)-1}$. We shall show that G contains very few short cycles, but at the same time has small independence number. Once we establish that, we could remove one vertex from each short cycle to enforce a high girth, yet not increasing the independence number. Finally we can use the well-known relation $\chi(G) \geqslant |V(G)| / \alpha(G)$ to conclude the proof.

We start off by analysing the number of short cycles. Let N_l count the number of cycles of length l in graph G. Clearly

$$\mathbb{E}\left(\sum_{3 \leqslant l \leqslant g} N_l\right) = \sum_{3 \leqslant l \leqslant g} \frac{n(n-1)\cdots(n-l+1)}{2l!} p^l < g \cdot (np)^l \leqslant g\sqrt{n}.$$

So with high probability, the total number of short cycles is smaller than, say, $n/2$.

Next we analyse the independence number. The probability that a $U \subseteq V : |U| = n/(2c)$ forms an independent set in G is

$$(1-p)^{\binom{n/(2c)}{2}} \leqslant \exp\left(-p\binom{n/(2c)}{2}\right) < \exp\left(-\frac{n^{1+1/(2g)}}{16c^2}\right)$$

for sufficiently large n. By a union bound over all U, we have

$$\mathbb{P}(\alpha \geqslant n/(2c)) < \binom{n}{n/(2c)} \cdot \exp\left(-\frac{n^{1+1/(2g)}}{16c^2}\right) < 2^n \cdot \exp\left(-\frac{n^{1+1/(2g)}}{16c^2}\right) \to 0.$$

To summarise, with high probability the random graph G has less than $n/2$ short cycles and, at the same time, independence number $\alpha(G) < n/(2c)$. Now we may remove one vertex for each short cycle, resulting in a graph G' such that $\operatorname{girth}(G') > g$ and $\alpha(G') \leqslant \alpha(G) < n/(2c)$. The latter bound implies $\chi(G') \geqslant (n - n/2)/\alpha(G') > c$. □

5.4 Sharing Neighbours

For graph $G = (V,E)$ and subset $R \subseteq V$, let us denote the common neighbours of vertices in R as $\operatorname{Com}(R) := \bigcap_{v \in R} N_G(v)$. Note that $R \subseteq \operatorname{Com}(\operatorname{Com}(R))$.

Theorem. Let $r,s \in \mathbb{N}$ be any given parameters. In any graph $G = (V,E)$ with average degree at least d, it is possible to find a "large" subset $U \subseteq V$ such that

- every $R \in \binom{U}{r}$ satisfies $|\operatorname{Com}(R)| > s$;
- $|U| \geqslant n\left(\frac{d}{n}\right)^t - \binom{n}{r} \cdot \left(\frac{s}{n}\right)^t$ for all $t \in \mathbb{N}$.

Proof. Sample vertices $v_1, \dots, v_t \in V$ independently and uniformly. Consider their common neighbours $U := \operatorname{Com}\{v_1, \dots, v_t\}$. This U is almost what we are looking for. Let us calculate

$$\begin{aligned}
\mathbb{E}(|U|) &= \sum_{v \in V} \mathbb{P}(v \in U) \\
&= \sum_{v \in V} \left(\frac{\deg(v)}{n}\right)^t \\
&= n^{1-t} \cdot \frac{\sum_{v \in V} (\deg(v))^t}{n} \\
&\geqslant n^{1-t} \cdot \left(\frac{\sum_{v \in V} \deg(v)}{n}\right)^t \\
&= n \cdot \left(\frac{d}{n}\right)^t
\end{aligned}$$

where we used Jensen's inequality on the convex function $x \mapsto x^t$.

But there could be a few bad subsets in U that violate the first property. Next we bound the expected number of them. Fix any "potentially bad" subset $R \in \binom{V}{r}$ such that $|\operatorname{Com}(R)| \leqslant s$. If the bad event $\{R \subseteq U\}$ happens, then $\{v_1, \dots, v_t\} \subseteq \operatorname{Com}(U) \subseteq \operatorname{Com}(R)$; in other words, every vertex v_i must have landed in $\operatorname{Com}(R)$, a set of size $\leqslant s$. Therefore $\mathbb{P}(R \subseteq U) \leqslant (s/n)^t$ and

$$\mathbb{E}(\#\text{bad } R) \leqslant \binom{n}{r} \cdot \left(\frac{s}{n}\right)^t.$$

Finally, we modify the set U by deleting one vertex from each bad R. After this step, the new U does not include any $R \in \binom{V}{r} : |\operatorname{Com}(R)| \leqslant s$, so the first property is satisfied. On the other hand $\mathbb{E}(|U|) \geqslant n\left(\frac{d}{n}\right)^t - \binom{n}{r} \cdot \left(\frac{s}{n}\right)^t$, so the second property also holds. This proves the existence of the desired object. □

Corollary. Let $A \uplus B$ be a bipartite graph where all vertices in B have degree $\leqslant r$. Then we may embed it into any graph G on n vertices of average degree d such that

$$n\left(\frac{d}{n}\right)^t - \binom{n}{r} \cdot \left(\frac{|A|+|B|}{n}\right)^t \geqslant |A|$$

for some $t \in \mathbb{N}$.

Proof. Take a $U \subseteq V = V(G)$ provided by the theorem with parameters r and $s \coloneqq |A|+|B|$, so that $|U| \geqslant |A|$ by assumption.

We will incrementally build an embedding $\gamma : A \uplus B \to V$. For each $a \in A$ we injectively assign a vertex $\gamma(a) \in U$. Then, for each remaining vertex $b \in B$, we consider the set

$$R \coloneqq \{\gamma(a) : a \in N_H(b)\} \subseteq U,$$

namely the image of its neighbourhood. By assumption $|R| \leqslant r$. Applying the property of U, we know $|\mathrm{Com}(R)| > |A|+|B|$, so there is some vacant vertex in $\mathrm{Com}(R)$ not yet embedded. We may simply take it as $\gamma(b)$. □

Appendix A
Permanents of 0-1 Matrices

Let $A \in \{0,1\}^{n\times n}$ be a binary matrix, and denote r_i to be the number of ones in row i. A trivial upper bound for permanent is $\operatorname{Per} A \leqslant \prod_{i=1}^n r_i$, but it neglects the interactions between different rows. The Minc-Bregman inequality adds that back into the picture:

Theorem. $\operatorname{Per} A \leqslant \prod_{i=1}^n (r_i!)^{1/r_i}$. The bound is tight for any block diagonal matrix where each block is an all-one square matrix.

Proof. Let S_n collect all permutations of $[n]$. Define $S_n^+ \coloneqq \{\pi \in S_n : \prod_{i=1}^n a_{i,\pi(i)} = 1\}$ to be all permutations that contribute to the permanent. We build a randomised estimator of $\operatorname{Per} A$:

- Sample a uniform order $\sigma \in S_n$;
- Sample a uniform $\pi \in S_n^+$;
- Process the rows one by one, as specified by the order σ. At step i we look at row $\sigma(i)$. We use the random variable $R_{\sigma(i)}$ to record the number of (remaining) ones in this row. Then we remove column $\tau(i) \coloneqq \pi(\sigma(i))$ from the matrix, which crudely simulates the interaction between different rows.
- Our estimate is just $N \coloneqq \prod_{i=1}^n R_{\sigma(i)} = \prod_{i=1}^n R_i$, which can be seen as a randomised, adaptive version of the trivial estimate.

Depending on the choices of σ, π, the estimate N may vary a great deal, but on average it is an overestimate—which is no surprise at all. To get the idea, fix any order σ and consider what was done in the first step: We looked into row $\sigma(1)$, and selected one random branch $\pi(\sigma(i))$ to expand the permanent. We did not explore other branches; we just treated them all the same as our selected branch, and multiplied our sub-estimate by the number of branches, $R_{\sigma(1)}$. Now, if we did select each branch with equal probability, then the estimate should be unbiased. However, since we specified $\pi \in S_n^+$, our selection is biased towards a branch with heavier sub-determinant, so we must have overestimated overall.

More formally, we claim that

$$\mathbb{E}(\log N) \geqslant \log(\operatorname{Per} A).$$

We prove this by gradual conditioning. Conditional on $\sigma(1)$ and $\tau(1)$, namely the first row that we inspect and the column that we remove, we can calculate

$$\mathbb{E}[\log N \mid \sigma(1), \tau(1)] = \log(r_{\sigma(1)}) + \mathbb{E}(\log N').$$

Here N' is the random estimator of $\operatorname{Per} A_{\tau(1)}$, for $A_{\tau(1)}$ being A without row $\sigma(1)$ and column $\tau(1)$. By induction we assume $\mathbb{E}(\log N') \geqslant \log(\operatorname{Per} A_{\tau(1)})$, thus

$$\mathbb{E}[\log N \mid \sigma(1), \tau(1)] \geqslant \log(r_{\sigma(1)}) + \log(\operatorname{Per} A_{\tau(1)}).$$

Next we take expectation over $\tau(1)$ (which is independent of $\sigma(1)$). The key observation is that $\mathbb{P}(\tau(1) = j) = \frac{\operatorname{Per} A_j}{\operatorname{Per} A}$ due to our uniform sampling $\pi \in S_n^+$. Therefore,

$$\begin{aligned}
\mathbb{E}[\log N \mid \sigma(1)] &\geqslant \log(r_{\sigma(1)}) + \sum_{j:\operatorname{Per} A_j > 0} \frac{\operatorname{Per} A_j}{\operatorname{Per} A} \log(\operatorname{Per} A_j) \\
&= \log(r_{\sigma(1)}) + \log(\operatorname{Per} A) + \sum_{j:\operatorname{Per} A_j > 0} \frac{\operatorname{Per} A_j}{\operatorname{Per} A} \log\left(\frac{\operatorname{Per} A_j}{\operatorname{Per} A}\right) \\
&\geqslant \log(r_{\sigma(1)}) + \log(\operatorname{Per} A) - \log(r_{\sigma(1)}) \\
&= \log(\operatorname{Per} A).
\end{aligned}$$

The third line is due to the entropy inequality on support size $\leqslant r_{\sigma(1)}$. Finally, taking expectation over $\sigma(1)$ establishes the claim.

The remaining proof is easy as we could calculate $\mathbb{E}(\log N) = \sum_{i=1}^{n} \mathbb{E}(\log R_i)$ explicitly. Let us condition on π. For any row $i \in [n]$ (which contains $r_i - 1$ ones excluding the one at column $\pi(i)$), there are exactly other $r_i - 1$ rows that, whenever processed earilier than row i, would "kill" a one in row i due to column removal. Hence $R_i = r_i - d + 1$ iff row i is processed at the d-th position among these r_i rows. So the distribution of R_i is just uniform over $[r_i]$, and

$$\mathbb{E}(\log R_i \mid \pi) = \sum_{k=1}^{r_i} \frac{1}{r_i} \cdot \log k = \frac{\log(r_i!)}{r_i}.$$

Taking expectation over π shows $\mathbb{E}(\log R_i) = \frac{\log(r_i)!}{r_i}$, and consequently

$$\log(\operatorname{Per} A) \leqslant \mathbb{E}(\log N) = \sum_{i=1}^{n} \frac{\log(r_i)!}{r_i}.$$

Bringing both sides to the exponent finishes the proof. □

Appendix B
Random Graph $G(n, 1/2)$

The random graph model $G(n, 1/2)$ is a central figure in random graph theory as it uniformly produces a graph on n vertices. This appendix studies two important parameters of random graph $G \sim G(n, 1/2)$: its independence number $\alpha(G)$ and chromatic number $\chi(G)$.

B.1 Independence number

Let Z_i count the number of independent sets of size i in G. Denote

$$\mu_i := \mathbb{E}(Z_i) = \binom{n}{i} 2^{-\binom{i}{2}}.$$

Note that

$$\frac{\mu_{i+1}}{\mu_i} = \frac{n-i}{i+1} \cdot 2^{-i},$$

so the sequence $(\mu_i)_{0 \leqslant i \leqslant n}$ is first increasing and then decreasing, with the "turning point" roughly at $i \approx \log n$.

Theorem. There exists some $i^\star \approx 2\log n$ such that $\alpha(G) \in \{i^\star - 1, i^\star\}$ with high probability. In other words, the independence number of G concentrates within two points!

Proof. Let $i^\star$ be the largest integer such that $\mu_{i^\star}$ stays above $n^{-1/2}$. One may easily find $i^\star = (1 + o(1))\, 2\log n$, and in this region the sequence depreciates by about $1/n$ in each step. We thus have

$$\begin{aligned} \mu_{i^\star} &\geqslant n^{-1/2} \\ \mu_{i^\star+1} &< n^{-1/2} \\ \mu_{i^\star-1} &\gtrsim n^{1/2}. \end{aligned}$$

Next we apply the standard second-moment method. On one hand $\mu_{i^\star+1} \to 0$, so with high probability $Z_{i^\star+1} = 0$. On the other hand $\mu_{i^\star-1} \to \infty$, so if we managed to show

$$\frac{\operatorname{Var}(Z_{i^\star-1})}{\mu_{i^\star-1}^2} \to 0, \tag{B.1}$$

then $Z_{i^\star-1} > 0$ with high probability. Combining the two parts, we can conclude that $\alpha(G) \in \{i^\star - 1, i^\star\}$ with high probability.

It remains to establish (B.1). Denote $i := i^\star - 1$ for convenience. Decompose $Z_i = \sum_{S \in \binom{V}{i}} X_S$ where X_S indicates if S forms an independent set. We calculate

$$\begin{aligned} \operatorname{Var}(Z_i) &< \sum_{S,T} \mathbb{E}(X_S X_T) \\ &= \mu_i + \sum_{2 \leqslant |S \cap T| \leqslant i-1} \mathbb{E}(X_S X_T) \\ &= \mu_i + \sum_{j=2}^{i-1} \binom{n}{i} \binom{i}{j} \binom{n-i}{i-j} 2^{-2\binom{i}{2} + \binom{j}{2}}, \end{aligned}$$

hence

$$\frac{\mathrm{Var}(Z_i)}{\mu_i^2} < \frac{1}{\mu_i} + \sum_{j=2}^{i-1} \frac{\binom{i}{j}\binom{n-i}{i-j}}{\binom{n}{i}} 2^{\binom{j}{2}}.$$

Put $f_j := 2^{\binom{j}{2}}\binom{i}{j}\binom{n-i}{i-j}/\binom{n}{i}$ and compute

$$\frac{f_{j+1}}{f_j} = \frac{(i-j)^2}{(j+1)\,(n-2i+j+1)} \cdot 2^j,$$

thus the f_j first decreases and then increases. The maximum value is attained at the boundaries $j=2$ and $j=i-1$. With careful computation we can bound $f_j \leqslant O(1/n)$. So

$$\frac{\mathrm{Var}(Z_i)}{\mu_i^2} = o(1) + O(\log n)\cdot O\left(\frac{1}{n}\right) \to 0$$

which is as we wanted. □

In fact we can prove very sharp probability bound (specifically, exponential instead of polynomial decay) by replacing the second moment method with a smart application of Azuma's inequality.

Lemma. Let i be the largest integer such that $\mu_i \geqslant n^3$; note that $i \approx 2\log n$ again. We have $\mathbb{P}(\alpha(G) < i) \leqslant \exp(-n^{2-o(1)})$.

Proof. We say a collection $\mathcal{I}$ of sets is *even* if $|I \cap J| \leqslant 1$ for all distinct $I, J \in \mathcal{I}$. Let random variable X record the cardinality of a largest even collection of size-i independent sets in G. Apparently $\mathbb{P}(\alpha(G) < i) = \mathbb{P}(X = 0)$. We will upper bound the latter by Azuma's inequality.

First of all, observe that X is 1-Lipschitz with respect to edge exposure. Indeed,

(a) X is monotonically decreasing with edge insertion.

(b) Assume $\mathcal{I}$ is a largest even collection of size-i independent sets in G. Now we add an arbitrary edge uv to G. Any $I \in \mathcal{I}$ containing both u and v is no longer independent in the new graph. Fortunately there could be at most one such I because $\mathcal{I}$ is even. So $\mathcal{I} \setminus \{I\}$ forms a (not necessarily largest) even collection of size-i independent sets in the new graph. This means that the value X can drop by at most 1.

Hence we may apply Azuma's inequality to obtain

$$\mathbb{P}(X=0) \leqslant \exp\left(\frac{-\mathbb{E}^2(X)}{2\binom{n}{2}}\right) \leqslant \exp\left(\frac{-\mathbb{E}^2(X)}{n^2}\right).$$

It remains to lower bound $\mathbb{E}(X)$ by roughly n^2. To this end, we cleverly produce a specific even collection $\mathcal{I}$ of size-i independent sets, and use its cardinality as a lower bound for X.

- Initialise $\mathcal{I}$ as the collection of all size-i independent sets.
- Keep each $I \in \mathcal{I}$ independently with probability q (to be specified later).
- Whenever there still exist distinct $I, J \in \mathcal{I}$ such that $|I \cap J| \geqslant 2$, remove I from $\mathcal{I}$.

Upon termination of the procedure, the $\mathcal{I}$ is indeed even. We analyse its cardinality below.

Let N count the pairs of size-i independent sets $\{I, J\}$ in G such that $|I \cap J| \geqslant 2$. Then by definition,

$$\mathbb{E}(|\mathcal{I}| \,|\, G) = q Z_i - q^2 N$$

where the randomness originates from the procedure itself. (Under condition G, both Z_i and N are just constants.) Therefore,

$$\begin{aligned} \mathbb{E}(X) \geqslant \mathbb{E}(|\mathcal{I}|) &= q\,\mathbb{E}(Z_i) - q^2\,\mathbb{E}(N) \\ &= q\,\mu_i - q^2\,\mathbb{E}(N) \\ &= \frac{\mu_i^2}{4\,\mathbb{E}(N)} \end{aligned} \tag{B.2}$$

with parameter $q \coloneqq \frac{\mu_i}{2\,\mathbb{E}(N)}$. It will soon become clear that $q \in (0,1)$.

Computing $\mathbb{E}(N)$ exactly follows the second-moment calculation before. One can derive

$$\mathbb{E}(N) = \frac{\mu_i^2}{2} \sum_{j=2}^{i-1} \frac{\binom{i}{j}\binom{n-i}{i-j}}{\binom{n}{i}} 2^{\binom{j}{2}} \geqslant \Omega(n^6)\, \Omega\left(\frac{\log^4 n}{n^2}\right) = \Omega\left((n\log n)^4\right) \gg \mu_i$$

which means $q \in (0,1)$ indeed. Plugging the expression for $\mathbb{E}(N)$ into (B.2), one can bound

$$\mathbb{E}(X) \geqslant \left(2 \sum_{j=2}^{i-1} \frac{\binom{i}{j}\binom{n-i}{i-j}}{\binom{n}{i}} 2^{\binom{j}{2}}\right)^{-1} \geqslant (1-o(1))\, \frac{n^2}{\log^5 n},$$

thus finishing the proof. □

B.2 Chromatic number

Unlike the independence number, the chromatic number is a much more complicated creature due to its lack of locality. We shall study it through the well-known connection

$$\frac{n}{\alpha(G)} \leqslant \chi(G) \leqslant 1 + n - \alpha(G).$$

It might be instructive to mention the proofs. For the left-hand side, note that any optimal proper colouring partitions the vertices into $\chi(G)$ independent sets, each one having size at most $\alpha(G)$, hence $n \leqslant \chi(G)\,\alpha(G)$. For the right-hand side, we can paint a maximum independent set by one colour, and the rest of vertices by distinct colours, thus producing a proper $(1+n-\alpha(G))$-colouring.

By the concentration of $\alpha(G)$, we immediately have

$$(1-o(1))\, \frac{n}{2\log n} \leqslant \chi(G) \leqslant 1 + n - (1+o(1))\, 2\log n$$

with high probability. The upper bound is far from decent though. Can we strengthen it to something comparable to the lower bound? Here is our "dream proof": repeatedly pull out a maximum independent set from the remaining graph and paint it by a new colour, until the graph becomes empty (or until the graph contains few vertices so that we may colour them distinctly). In other words,

1. pull out an $I_1 \subseteq V(G)$ of size $\alpha(G)$;
2. pull out an $I_2 \subseteq V(G - I_1)$ of size $\alpha(G - I_1)$;
3. pull out an $I_3 \subseteq V(G - I_1 - I_2)$ of size $\alpha(G - I_1 - I_2)$;
4. …

If we managed to show that $\alpha(G-I_1), \alpha(G-I_1-I_2), \ldots$ are relatively close to $\alpha(G) \approx 2\log n$, then the procedure should terminate in about $\frac{n}{2\log n}$ steps, thus witnessing a proper colouring of that size.

In fact our dream can readily come true, given the sharp lemma in the previous section.

Lemma. $\chi(G) \leqslant (1+o(1))\, \frac{n}{2\log n}$ with high probability.

Proof. Denote $n' \coloneqq n/\log^2 n$. For any subset $U \in \binom{V(G)}{n'}$, the induced subgraph $G[U]$ is a sample from model $G(n', 1/2)$, hence

$$\mathbb{P}(\alpha(G[U]) < i) \leqslant \exp(-n^{2-o(1)})$$

for some $i \approx 2\log n' \approx 2\log n$ by the lemma in the previous section. Applying a union bound over all subsets U, we conclude with high probability that $\alpha(G[U]) \geqslant i$ for *all* $U \in \binom{V(G)}{n'}$.

Once this happens, we can repeatedly pull out independent sets of size $\geqslant i$ until there are less than n' vertices left. We colour all the remaining vertices by distinct colours. Overall

$$\chi(G) \leqslant \frac{n}{i} + n' = (1+o(1))\,\frac{n}{2\log n}. \qquad \square$$

Corollary. $\chi(G) = (1+o(1))\,\frac{n}{2\log n}$ with high probability.