MATH 264A: Combinatorics	**Fall 2024**

Lecture 1: Some Basic Tools

October 1, 2024

Lecturer: Lutz Warnke *Scribes: Ray Tsai*

Inductive Approaches

This method is done by changing the size of the problem, e.g. adding an vertex or an edge in a graph.

Example 0.1. *Every n-vertex graph with maximum degree Δ has $\geq \beta^n$ valid vertex colorings with $\leq \lceil \Delta+\beta \rceil$ colors.*

Proof. Color the vertices $v_1, v_2, \ldots, v_n$ sequentially. Since v_i has $\leq \Delta$ neighbors already colored, there are $\geq \lceil \Delta + \beta \rceil - \Delta \geq \beta$ choices to color v_i. Define N_i as # valid colorings of $v_1, \ldots, v_i$. Then, the *Telescoping Product* now yields

$$N_n = \frac{N_n}{N_{n-1}} \cdot \frac{N_{n-2}}{N_{n-1}} \cdots \frac{N_1}{N_0} \cdot N_0 \geq \beta^n,$$

as $N_0 = 1$. □

Despite being an extremely basic technique, induction can prove several advanced theorems if used artfully. The following are some exciting theorems which can be proven by induction:

1. Strengthen Lovász Local Lemma (LLL)
2. Chromatic number of triangle-free graph with max-degree Δ is $\leq (1 + o(1))\frac{\Delta}{\log \Delta}$ as $\Delta \to \infty$.
3. Almost all triangle-free graphs are bipartite.

Double Counting/Switching

Also known as the Pertubation method, e.g. change of location of edges.

Example 0.2. *Find the $\#\Pi \in S_n$ without fix-points, i.e. $\Pi(i) \neq i$ for all i.*

Proof. We prove this by a basic approach which consists of several steps:

Step 1: Define the "Switching Operation." Let $S_{n,k}$ be the set of permutations with k fix-points. Define the switching operation to transform $\pi \in S$ to $\pi' \in S_{n,1}$.

Step 2: Consider the auxiliary bipartite graph. Let $S_{n,0}, S_{n,1}$ be parts of the bipartite graph. Connect $\pi \in S_{n,0}$ with $\pi' \in S_{n,1}$ if π' results from π through the switching operation.

Step 3: Double count the degrees.

$$\sum_{\pi \in S_{n,0}} \deg \pi = \sum_{\pi' \in S_{n,1}} \deg \pi'$$

Step 4: Degree essentially transfers to ratio. Suppose $\deg \pi \approx a$ and $\deg \pi' \approx b$, for all $\pi \in S_{n,0}$ and $\pi' \in S_{n,1}$. Then,

$$\frac{|S_{n,0}|}{|S_{n,1}|} \approx \frac{b}{a}.$$

□

This method can be applied to count d-regular graphs with certain properties, i.e. random model without independence.

Asymptotic Methods

Rather than finding the close form of a discrete function, sometimes it is significantly easier to approximate the function in asymptotic settings.

Bootstrapping

Suppose we have an equation $w(z)e^{w(z)} = z$ and we try to extract $w(z)$. By bootstrapping, $w(z) = \ln z - \ln\ln z + o(1)$.

Integral-Approximation

As the title suggests, this method estimates a summation $\sum_{k\in I} f(k)$ with its integral counterpart $\int_I f(x)\,dx$. For example, the the summation derived from the Fibonacci Tiliing Problem can be estimated by the Laplace-Method, i.e.

$$\sum_{0\leq k\leq \frac{n}{2}} \binom{n-k}{k} \sim \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} \qquad n\to\infty.$$

MATH 264A: Combinatorics	**Fall 2024**

Lecture 2: Inductive Counting Arguments

October 1, 2024

Lecturer: Lutz Warnke *Scribes: Sam Spiro, Jason O'Neill, and Lutz Warnke*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Today's topics:

- Inductive Counting Approach (often improves Lovász Local Lemma[1] applications)
- Examples: Non-repetitive Words and Independent Sets

- *Sources:*
  arXiv:2006.09094 (Another approach to non-repetitive colorings of graphs, by Rosenfeld)
  arXiv:2008.00775 (Exponentially Many Hypergraph Colourings, by Wanless-Wood)

1 Inductive Counting Approaches

- *Goal:* show that "good" object exists (e.g. valid coloring with special properties)
- *Approach:* generalize the problem, and then to use structural induction (e.g. induction on the number of vertices or edges) to show that many such good objects exist, i.e., give a concrete quantitative lower bound.

Proposition 1.1 (Toy Example)**.** *Any graph $G = (V, E)$ with maximum degree Δ has a valid coloring with at most $\lceil \Delta + \beta \rceil$ colors for any $\beta \geqslant 1$.*

Proof. Our first step is to generalize the problem to set us up for induction: set $V = \{v_1, \dots, v_n\}$ and define

$N_i :=$ number of valid colorings of the graph induced by $\{v_1, \dots, v_i\}$

If we sequentially color v_i after starting with one of the N_{i-1} valid colorings for $v_1, \dots, v_{i-1}$, then at most Δ colors are forbidden for v_i. With this we see

$$N_i \geqslant N_{i-1} \cdot (\lceil \Delta + \beta \rceil - \Delta) \geqslant N_{i-1} \cdot \beta,$$

which in turn gives the ratio bound

$$\frac{N_i}{N_{i-1}} \geqslant \beta.$$

Using induction it follows that $N_i \geqslant \beta^i$ for all i, which for $i = n$ gives $N_n \geqslant \beta^n \geqslant 1$, proving the result. $\square$

Two remarks regarding proof of this toy example:

1. *It shows a connection between local and global properties*: to prove $N_i \geqslant \beta^i$, all we need to understand is the effect of adding a single vertex.
2. *Structural induction over number of vertices or edges often improves classical Lovász Local Lemma (LLL) type applications*, which implicitly rely on induction over the number of 'bad' events (often yielding either better bounds and/or simpler proofs).

In the following we shall illustrate the power and simplicity of the inductive counting approach via two concrete applications.

[1] To understand these notes, you do not need to know what the Lovász Local Lemma (LLL) is. We mainly mention it several times, since this is a classical approach from probabilistic combinatorics that serves as a benchmark for our results (the induction approach often allows us to improve the LLL results, or yields simpler proofs).

1.1 Non-repetitive Words

A *word* w is just a sequence of symbols from some finite alphabet.
We say that a word w is *non-repetitive* if within w, no word appears twice consecutively.

Here are some examples of repetitive words: aa, $abab$, $abcbc$, and $abcabca$.
Here are some examples of non-repetitive words: ab, aba (note that here the word a does appear twice, but not consecutively) and $abcab$.

Theorem 1.2 (Thue 1906)**.** *Let A be a 3-symbol alphabet. Then, for any $n \geqslant 1$, there exists a non-repetitive word $a_1 \cdots a_n$ such that $a_i \in A$ for all i.*

We will not prove Theorem 1.2 in class, but instead remark that the following stronger variant[2] is open (which is inspired by list-coloring variants of the normal coloring problem).

Conjecture 1.3 (Open)**.** *Let $L_1, \dots, L_n$ be alphabets with $|L_i| = 3$ for all i. Then, for any $n \geqslant 1$, there exists a non-repetitive word $a_1 \cdots a_n$ with $a_i \in L_i$ for all i.*

We will now prove that this conjecture holds if we allow for alphabets of size 4 (instead of size 3).

Theorem 1.4. *Let $L_1, \dots, L_n$ be alphabets with $|L_i| = 4$ for all i. Then, for any $n \geqslant 1$, there exists a non-repetitive word $a_1 \cdots a_n$ with $a_i \in L_i$ for all i (in fact at least 2^n many).*

One can prove this result by using (somewhat more technical) variants of LLL. We will use an inductive approach, which guarantees the existence of many such words and is the simplest known proof of the result.

Proof (using inductive approach). As before, our first step will be to generalize our problem: we define

$N_i :=$ to be the number of non-repetitive words $a_1 \cdots a_i$ with $a_j \in L_j$ for all $j \leqslant i$

Our goal is then to show that

$$N_i \geqslant 2^i \quad \text{for all } 0 \leqslant i \leqslant n,$$

which give us our desired result by taking $i = n$. We will do this by *proving by induction* on $i \geqslant 1$ that

$$\frac{N_i}{N_{i-1}} \geqslant \beta \qquad \text{for some } \beta,$$

and it will turn out (at the end of the proof) that we can take $\beta := 2$.
<u>For the base case</u>, we have $N_1 = |L_1| = 4$ and $N_0 = 1$, so $N_1/N_0 \geqslant \beta$ as long as $\beta \leqslant 4$.
<u>For the induction step</u>, the goal is to try to use a symbol from L_i to *extend* each of the N_{i-1} non-repetitive words of length $i-1$. To this end, consider all words of the special form $a_1 \cdots a_{i-1} b$ where $a_1 \cdots a_{i-1}$ is non-repetitive and $b \in L_i$. If B denotes the number of repetitive words of this special form (which we think of as 'bad' words), then

$$|L_i| \cdot N_{i-1} = N_i + B, \tag{1.1}$$

since $|L_i| \cdot N_{i-1}$ is the total number of words of this special form and each non-repetitive word counted by N_i is generated in a unique way by this process. Our next goal then is to upper bound B. Note that by construction, any repetitive word counted by B must have its repetition at the end of the word. That is, the word will be of the form

$$a_1 \cdots a_{i-2j} w' w',$$

where w' is some word of length $j \geqslant 1$. The key insight here is that for such a repetitive word, if we are told the first $i - j$ symbols of this word (and j), then we can uniquely determine what the word is. Moreover, the $i - j$ symbols form a non-repetitive word. It follows that

$$B \leqslant \sum_{1 \leqslant j < i/2} N_{i-j} \leqslant \sum_{1 \leqslant j < i/2} \beta^{1-j} N_{i-1} \leqslant \sum_{j \geqslant 1} \beta^{1-j} N_{i-1} = \frac{\beta}{\beta - 1} N_{i-1},$$

[2] Remark: Conjecture 1.3 is certainly true if all L_i are disjoint, and by Theorem 1.2 it is also true when all the L_i are equal. Furthermore, as many of you pointed out, it appears as if the extreme case where all the L_i are equal is hardest, so one might think Theorem 1.2 trivially implies Conjecture 1.3. While this might seem plausible on first sight, this is hard to convert into a proof, since things are trickier than one might think (in particular, the analogous 'plausible' reasoning for list-coloring is wrong).

where the first inequality follows by first specifying the length of the repetition w' and then choosing the first $i-j$ symbols of the word, and the second inequality uses repeated applications of the inductive hypothesis. Combining this with (1.1) and $|L_i| = 4$ gives

$$4N_{i-1} \leqslant N_i + \frac{\beta}{\beta-1} N_{i-1} \implies \frac{N_i}{N_{i-1}} \geqslant 4 - \frac{\beta}{\beta - 1}.$$

To finish the induction, we need to guarantee that the least quantity is at least β, which works for $\beta = 2$. $\square$

Let us take a step back and reflect on the key steps of the above inductive counting proof, which are typical features of this approach. We first generalized the problem, by counting the total number of non-repetitive words of length i, and aimed at a stronger inductive bound (which guarantees many words). By extending non-repetitive words of length $i-1$, we then constructed all non-repetitive words of length i together with some 'bad' words. Furthermore, we used structural results/properties together with our inductive bounds to efficiently upper bound the number of these bad words. Finally, if we can ensure that a certain 'consistency' equation (for β) is satisfied at the end of the induction step, then the desired counting result follows.

1.2 Independent Sets (ISETs)

Given a graph $G = (V, E)$ with maximum degree Δ, a simple greedy algorithm shows that there exists an independent set of size at least $|V|/(\Delta + 1)$ (namely, by adding an arbitrary vertex v to the independent set, then deleting v together with its neighborhood, then iterate on the remaining graph). By decreasing the size of the independent set by a constant factor, the following results guarantees the existence of many "nice" independent sets, i.e., where we can specify the location of the vertices (to see this, consider $V = V_1 \cup \cdots \cup V_s$ with $|V_i| = 4\Delta$, in which case this result gives an independent set of size at least $|V|/(4\Delta)$ containing one vertex from each V_j).

Theorem 1.5. *Let $G = (V, E)$ be a graph with maximum degree Δ. If $V_1, \ldots, V_s \subseteq V$ are disjoint sets of vertices of a graph with $|V_i| \geqslant 4\Delta$ for all i, then there exists at least $(2\Delta)^s$ independent sets $I \subseteq V$ with exactly one vertex from each V_i.*

Remark 1.6. One can prove a similar result using LLL, but with 4Δ replaced by $2e\Delta \approx 5.44\Delta$ (which is slightly worse) and with less control over how many such independent sets exist.

Proof (using inductive approach). Again we start by generalizing our problem: for any $T \subseteq [s]$, we define

$N_T :=$ number of independent sets $I \subseteq V_T := \bigcup_{j \in T} V_j$ such that I has exactly one vertex from each V_j (such independent sets are called *valid*)

Our goal is then to show

$$N_T \geqslant (2\Delta)^{|T|} \qquad \text{for all } T \subseteq [s],$$

which will in particular give us our desired result by taking $T = [s]$. Specifically, we will *show by induction* that for all $T \subseteq [s]$ and $x \in T$, we have

$$\frac{N_T}{N_{T \setminus \{x\}}} \geqslant \beta \qquad \text{for some } \beta,$$

and we will eventually show $\beta := 2\Delta$ works.

<u>The base case</u> $T = \{x\}$ has $N_T = |V_x| \geqslant 4\Delta$ and $N_\emptyset = 1$, so this will be fine provided $\beta \leqslant 4\Delta$.

<u>For the induction step</u>, we use a vertex from V_x to *extend* each valid independent set $I \subseteq V_{T \setminus \{x\}}$ counted by $N_{T \setminus \{x\}}$. This gives

$$|V_x| \cdot N_{T \setminus \{x\}} = N_T + B, \tag{1.2}$$

where B is the number of independent sets generated in this way which are not valid (which we think of as 'bad' sets). As before, it remains to upper bound B in an efficient way. To this end observe that[3] any element

[3] For those familiar with the usual LLL proof of Theorem 1.5 (assuming $|V_i| \geqslant 2e\Delta$), we remark that the way we count the number 'bad' sets in (1.3) is closely related to the way one usually bounds the maximum degree in the LLL dependency graph.

counted by B consists of a vertex $v \in V_x$, some neighbor w of v which lies in some V_z set with $z \in T \setminus \{x\}$, together with a valid independent set with respect to $T \setminus \{x, z\}$. Using this and our inductive hypothesis,

$$B \leqslant |V_x| \cdot \Delta \cdot \max_{z \in T \setminus \{x\}} N_{T \setminus \{x,z\}} \leqslant |V_x| \cdot |\Delta| \cdot N_{T \setminus \{x\}}/\beta. \tag{1.3}$$

Combining this with (1.2) and $|V_i| \geqslant 4\Delta$ gives

$$\frac{N_T}{N_{T \setminus \{x\}}} \geqslant |V_x| \cdot (1 - \Delta/\beta) \geqslant 4\Delta(1 - \Delta/\beta),$$

which concludes the inductive proof if the last quantity is at least β, which works for $\beta = 2\Delta$. □

MATH 264A: Combinatorics	**Fall 2024**

Lecture 3: Kleitman–Rothschild Method

October 3, 2024

Lecturer: Lutz Warnke *Scribes: Kalen Patton (writing), Tolson Bell & Erlang Surya (editing)*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Today's topics:

- Kleitman–Rothschild Method: Inductive method to show asymptotic equality of sets
- Illustrated via example: Almost all triangle triangle-free graphs are 2-colorable (Erdős–Kleitman–Rothschild '76).

1 Estimating the Number of Triangle-Free Graphs

Theorem 1.1 (Almost all triangle-free graphs are 2-colorable)**.** *If we define*

$$\mathcal{T}(n) := \{\textit{triangle-free graphs on } n \textit{ vertices}\},$$
$$\mathcal{C}ol_2(n) := \{\textit{2-colorable graphs on } n \textit{ vertices}\},$$

then we have

$$\lim_{n\to\infty} \frac{|\mathcal{T}(n)|}{|\mathcal{C}ol_2(n)|} = 1.$$

The main difficulty in estimating $|\mathcal{T}(n)|$ is that there is no obvious way to construct/count triangle-free graphs (no known generating function). This means classical methods fail to give a good estimate.

Theorem 1.1 is useful, since $\mathcal{C}ol_2(n) \subset \mathcal{T}(n)$ has a much simpler structure than $\mathcal{T}(n)$, and thus is easier to enumerate. For instance, we get the following corollary from the known asymptotics of $|\mathcal{C}ol_2(n)|$.

Corollary 1.2. $|\mathcal{T}(n)| = (1 + o(1)) \cdot |\mathcal{C}ol_2(n)| = (1 + o(1)) \cdot 2^{n^2/4+n-(1/2)\log_2 n}$.

Theorem 1.1 also yields useful structural information, namely that a random triangle-free graph is with high probability 2-colorable (i.e., with probability tending to 1 as $n \to \infty$).

1.1 Sketch of the Proof

Noting that $\mathcal{C}ol_2(n) \subseteq \mathcal{T}(n)$, it suffices to show that $|\mathcal{T}(n)| \le (1 + o(1))|\mathcal{C}ol_2(n)|$. To do so, we apply the *Kleitman-Rothschild method*, which consists of the following three parts.

I. **Define "bad" sets** $A(n)$, $B(n)$, and $D(n)$ such that

$$\mathcal{T}(n) \subseteq \mathcal{C}ol_2(n) \cup A(n) \cup B(n) \cup D(n).$$

The "cleverness" of this proof technique comes from the careful definition of these sets.

II. **Bound the relative "growth rates" of each bad set.** That is, for suitable integers x, we want to place *upper* bounds on each of

$$\frac{|A(n)|}{|\mathcal{T}(n-x)|}, \qquad \frac{|B(n)|}{|\mathcal{T}(n-x)|}, \qquad \frac{|D(n)|}{|\mathcal{T}(n-x)|},$$

and place a *lower* bound on

$$\frac{|\mathcal{C}ol_2(n)|}{|\mathcal{C}ol_2(n-x)|}.$$

III. **Use induction** to show that

$$|A(n)| + |B(n)| + |D(n)| = o(|\mathcal{C}ol_2(n)|).$$

Heuristic Idea: For each bad set, we can expand the ratio similar to

$$\frac{|A(n)|}{|\mathcal{C}ol_2(n)|} = \frac{|A(n)|}{|\mathcal{T}(n-x)|} \cdot \frac{|\mathcal{T}(n-x)|}{|\mathcal{C}ol_2(n-x)|} \cdot \frac{|\mathcal{C}ol_2(n-x)|}{|\mathcal{C}ol_2(n)|}.$$

In this expansion, we can bound $\frac{|A(n)|}{|\mathcal{T}(n-x)|}$ and $\frac{|\mathcal{C}ol_2(n-x)|}{|\mathcal{C}ol_2(n)|}$ using II, and we know $\frac{|\mathcal{T}(n-x)|}{|\mathcal{C}ol_2(n-x)|} = 1 + o(1)$ by induction. We expect that $\frac{|A(n)|}{|\mathcal{T}(n-x)|}$ will be large, but assuming we have done steps I and II well, $\frac{|\mathcal{C}ol_2(n-x)|}{|\mathcal{C}ol_2(n)|}$ will be tiny enough that the overall quantity comes out to be $o(1)$.

Then, using I and III we get that

$$|\mathcal{T}(n)| \leq |\mathcal{C}ol_2(n)| + |A(n)| + |B(n)| + |D(n)| \leq (1 + o(1))|\mathcal{C}ol_2(n)|.$$

1.2 Part I: Definitions

Intuition: If Theorem 1.1 were true, then we expect a typical triangle-free graph to be 2-colorable. Thus, the bad sets $A(n)$, $B(n)$, and $D(n)$ should reflect properties that a typical 2-colorable graph (i.e. random bipartite graph) does *not* have.

In general, we can image a random bipartite graph has color classes roughly size $n/2$, and has each edge between these color classes inserted with probability $1/2$. Thus, we would expect that vertices in the graph should not have "too few" neighbors, which motivates our below definitions of the sets $A(n)$ and $B(n)$. In defining the set $D(n)$, we then seek to specify a set of properties which "cover the remaining cases," so that $D(n) = \mathcal{T}(n) \setminus (A(n) \cup B(n) \cup \mathcal{C}ol_2(n))$.

We define the sets as follows. For these definitions, we use $\Gamma(v)$ and $\Gamma(S)$ to denote the set of neighbors of a vertex v or set of vertices S, respectively.

- $A(n) \subseteq \mathcal{T}(n)$ is the subset of graphs containing a vertex v such that $|\Gamma(v)| \leq \log_2 n$.
- $B(n) \subseteq \mathcal{T}(n)$ is the subset of graphs containing a set Q of size $|Q| = \log_2 n$, such that

$$|\Gamma(Q)| \leq (1/2 - 1/1000)n.$$

- $D(n) \subseteq \mathcal{T}(n) \setminus (A(n) \cup B(n))$ is the subset of graphs containing an edge $\{x, y\}$ and sets $Q_x \subseteq \Gamma(x)$ and $Q_y \subseteq \Gamma(y)$, such that

$$|Q_x| = \log_2 n,$$
$$|Q_y| = \log_2 n,$$
$$|\Gamma(Q_x) \cap \Gamma(Q_y)| \geq n/100.$$

  Note: $\Gamma(Q_x) \cap \Gamma(Q_y)$ would have to be empty in a 2-colorable graph, since otherwise we get a 5-cycle. Thus, we certainly have $D(n) \subseteq \mathcal{T}(n) \setminus (A(n) \cup B(n) \cup \mathcal{C}ol_2(n))$.

It is possible via a graph theory argument (omitted here) to show that $\mathcal{T}(n) \subseteq \mathcal{C}ol_2(n) \cup A(n) \cup B(n) \cup D(n)$. The basic idea is to show that any graph $G \in \mathcal{T}(n) \setminus [A(n) \cup B(n) \cup D(n)]$ is 2-colorable.

1.3 Part II: Bounding Sets by Constructing Them

To get the size bounds for part II, we will give a method for constructing graphs in each set. For each set S that we seek to upper bound, we will formulate our construction so that *any* graph in S can be obtained from our procedure (and potentially more graphs). Thus, we can conclude that $|S|$ must be at most the number of ways to apply the procedure. In each case, we will follow the basic process outlined below for generating the graphs:

1. First choose a suitable vertex-subset of size x.
2. Then place a triangle-free graph on the remaining $n - x$ vertices.
3. Finally connect the x vertices to the rest of the graph (by placing edges).

Claim 1.3. $\frac{|A(n)|}{|\mathcal{T}(n-1)|} \leq 2^{2(\log_2 n)^2}$ *for* $n \geq n_0$.

Proof. We use the outlined strategy with $x = 1$. To generate an element of $A(n)$, first choose a vertex v. Then, place a triangle free graph on the remaining $n - 1$ vertices. Finally, choose at most $\log_2 n$ of these vertices to connect to v. This method over-counts the number of graphs in $A(n)$, so we have

$$\begin{aligned} |A(n)| &\leq n \cdot |\mathcal{T}(n-1)| \cdot \sum_{i=0}^{\log_2 n} \binom{n-1}{i} \\ &\leq n \cdot |\mathcal{T}(n-1)| \cdot (\log_2 n + 1) n^{\log_2 n} \\ &\leq 2^{2(\log_2 n)^2} \cdot |\mathcal{T}(n-1)|, \end{aligned}$$

for sufficiently large n. □

Claim 1.4. $\frac{|B(n)|}{|\mathcal{T}(n-\log_2 n)|} \leq 2^{(1/2-1/2000)n \log_2 n}$ *for* $n \geq n_0$.

Proof. We follow the same procedure as before, but with $x = \log_2 n$. To generate a graph in $B(n)$, we first pick a vertex subset Q of size $\log_2 n$. Then, we place a triangle free graph on the remaining $n - \log_2 n$ vertices, and an arbitrary graph on the first $\log_2 n$ vertices. Finally, we pick a vertex subset R of size at most $(1/2 - 1/1000)n$ to be the neighbors of Q, and add some set of edges between the vertices of Q and R. This gives us

$$\begin{aligned} |B(n)| &\leq \binom{n}{\log_2 n} \cdot |\mathcal{T}(n - \log_2 n)| \cdot 2^{\binom{\log_2 n}{2}} \cdot \sum_{\substack{R \subseteq [n] \setminus Q \\ |R| \leq (1/2-1/1000)n}} 2^{|R| \cdot |Q|} \\ &\leq n^{\log_2 n} \cdot |\mathcal{T}(n - \log_2 n)| \cdot 2^n \cdot 2^{\binom{\log_2 n}{2}} \cdot 2^{(1/2-1/1000)n \cdot \log_2 n} \\ &\leq 2^{(1/2-1/2000)n \cdot \log_2 n} \cdot |\mathcal{T}(n - \log_2 n)|, \end{aligned}$$

for sufficiently large n. □

Claim 1.5. $\frac{|D(n)|}{|\mathcal{T}(n-2)|} \leq 2^{(1-1/2000)n}$ *for* $n \geq n_0$.

For time-reasons we omit the proof for Claim 1.5, but the idea is to follow the same strategy as for Claims 1.3–1.4.

Claim 1.6. $\frac{|\mathcal{C}ol_2(n)|}{|\mathcal{C}ol_2(n-1)|} \geq 2^{(n-1)/2}$ *for all* n.

Proof. The proof of this claim is similar to the previous ones, but it differs in that we want a lower bound instead of an upper bound. Thus, we give a procedure for constructing a graph of $\mathcal{C}ol_2(n)$, such that each application of the procedure guarantees a *unique* graph of $\mathcal{C}ol_2(n)$, but does not necessarily construct all graphs in $\mathcal{C}ol_2(n)$.

Our construction proceeds as follows. To generate a graph in $\mathcal{C}ol_2(n)$, we first "remove" vertex n, and then place a 2-colorable graph on the remaining $n - 1$ vertices. Let C_1, C_2 be the two color-classes of the graph on these vertices, with $|C_1| \leq (n-1)/2 \leq |C_2|$. To ensure that the graph remains 2-colorable when we add vertex n back, we only select a subset of C_2 to connect to n.

From this construction, we see that

$$|\mathcal{C}ol_2(n)| \geq |\mathcal{C}ol_2(n-1)| \cdot 2^{(n-1)/2}$$

for all n. □

1.4 Part III: Using Induction

Inducting on n, we will show that there exist constants $c, \gamma > 1$ such that

$$|\mathcal{T}(n)| \le \big(1 + c\gamma^{-n}\big)|\mathcal{C}ol_2(n)| \qquad \text{for } n \ge 1. \tag{†}$$

The idea is to first choose $\gamma > 1$ sufficiently close to 1 so that the below induction step will work for all $n \ge n_0$, where $n_0 = n_0(\gamma)$ is chosen after γ. Only afterwards we choose $c = c(\gamma, n_0)$ sufficiently large, so that (†) trivially holds also for all $n \le n_0$ (establishing the base case). Thus, to get our desired result, it remains to complete the induction step for sufficiently large $n \geqslant n_0$ First, recall from part I that

$$\frac{|\mathcal{T}(n)|}{|\mathcal{C}ol_2(n)|} \le 1 + \frac{|A(n)|}{|\mathcal{C}ol_2(n)|} + \frac{|B(n)|}{|\mathcal{C}ol_2(n)|} + \frac{|D(n)|}{|\mathcal{C}ol_2(n)|},$$

so it suffices to show that each of $\frac{|A(n)|}{|\mathcal{C}ol_2(n)|}$, $\frac{|B(n)|}{|\mathcal{C}ol_2(n)|}$, and $\frac{|D(n)|}{|\mathcal{C}ol_2(n)|}$ is at most $\frac{c}{3}\gamma^{-n}$.

Using Claim 1.3, our inductive hypothesis, and Claim 1.6, we have

$$\begin{aligned}\frac{|A(n)|}{|\mathcal{C}ol_2(n)|} &= \frac{|A(n)|}{|\mathcal{T}(n-1)|} \cdot \frac{|\mathcal{T}(n-1)|}{|\mathcal{C}ol_2(n-1)|} \cdot \frac{|\mathcal{C}ol_2(n-1)|}{|\mathcal{C}ol_2(n)|} \\ &\le 2^{2(\log_2 n)^2} \cdot \big(1 + c\gamma^{-(n-1)}\big) \cdot 2^{-(n-1)/2} \\ &\le 2c \cdot 2^{-n/2+1/2+2(\log_2 n)^2} \le \frac{c}{3}\gamma^{-n}\end{aligned}$$

for $n \ge n_0$, by choice of γ.

Similarly, using Claim 1.4, our inductive hypothesis, and Claim 1.6, we obtain

$$\begin{aligned}\frac{|B(n)|}{|\mathcal{C}ol_2(n)|} &= \frac{|B(n)|}{|\mathcal{T}(n-\log_2 n)|} \cdot \frac{|\mathcal{T}(n-\log_2 n)|}{|\mathcal{C}ol_2(n-\log_2 n)|} \cdot \prod_{i=1}^{\log_2 n} \frac{|\mathcal{C}ol_2(n-i)|}{|\mathcal{C}ol_2(n-i+1)|} \\ &\le 2^{(1/2-1/2000)n\log_2 n} \cdot \big(1 + c\gamma^{-(n-\log_2 n)}\big) \cdot \prod_{i=1}^{\log_2 n} 2^{-(n-i)/2} \\ &\le 2c \cdot 2^{-\frac{1}{2000}n\log_2 n + (\log_2 n)^2} \le \frac{c}{3}\gamma^{-n}\end{aligned}$$

for $n \ge n_0$, by choice of γ.

Finally, applying the same strategy with Claim 1.5 gives

$$\begin{aligned}\frac{|D(n)|}{|\mathcal{C}ol_2(n)|} &= \frac{|D(n)|}{|\mathcal{T}(n-2)|} \cdot \frac{|\mathcal{T}(n-2)|}{|\mathcal{C}ol_2(n-2)|} \cdot \frac{|\mathcal{C}ol_2(n-2)|}{|\mathcal{C}ol_2(n-1)|} \cdot \frac{|\mathcal{C}ol_2(n-1)|}{|\mathcal{C}ol_2(n)|} \\ &\le 2^{(1-1/2000)n} \cdot \big(1 + c\gamma^{-(n-2)}\big) \cdot 2^{-(n-2)/2} \cdot 2^{-(n-1)/2} \\ &\le 2c \cdot 2^{-\frac{1}{2000}n + 3/2} \le \frac{c}{3}\gamma^{-n}\end{aligned}$$

for $n \ge n_0$, by choice of γ, completing the induction step.

MATH 264A: Combinatorics	Fall 2024

Lecture 4 and 5: Johansson-Molloy Theorem

October 8 and 11, 2024

Lecturer: Lutz Warnke

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Today's topics:

- Variant of Inductive Counting Approach that uses probabilistic estimates
- Illustrated via example: Chromatic Number of triangle-free graphs with large maximum degree

- *Sources:*
  arXiv:2109.15215 (Colouring locally sparse graphs with the first moment method, by Pirot-Hurley)
  arXiv:2111.06214 (Simplified proof of the Johansson-Molloy Theorem, by Martinsson)

1 Coloring Triangle-Free Graphs

Theorem 1.1 (Johansson-Molloy Theorem (1996, 2019)). *For any $\varepsilon > 0$, any triangle-free graph G with maximum degree at most Δ has chromatic number at most $\chi(G) \leqslant (1+\varepsilon)\Delta/\log\Delta$ as $\Delta \to \infty$.*

- The chromatic number bound is optimal up to a factor of two:
  a random Δ-regular graph G typically satisfies $\chi(G) = (1+o(1))\Delta/(2\log\Delta)$ as $\Delta \to \infty$.
- It is most likely difficult to improve the bound to $\chi(G) \leqslant (\beta+\varepsilon)\Delta/\log\Delta$ for $\beta \in (0,1)$, as this would yield an improvement on the upper bound of the Ramsey number $R(3,t)$, which is a heavily studied problem.
- We will give a 'simple' inductive proof of the Johannson-Molloy Theorem, by establishing that there are exponentially many valid colorings. This proof is significantly less technical than earlier proofs (which used semi-random approach to iteratively color most vertices, and then LLL to color the leftover vertices).

1.1 Inductive Proof

As usual for inductive proofs, we begin by generalizing the problem. Define

$\mathcal{C}(G)$:= set of all valid colorings of G using at most $k := (1+\varepsilon)\Delta/\log\Delta$ colors.

Gearing up towards proving $|\mathcal{C}(G)| \geqslant \ell^{|V(G)|}$, using induction we will show that

$$\frac{|\mathcal{C}(H)|}{|\mathcal{C}(H-v)|} \geqslant \ell := \Delta^{\varepsilon/2} \tag{1.1}$$

for all induced subgraphs $H \subseteq G$ and vertices $v \in V(H)$.

For any partial coloring (not all vertices are assigned colors) c of G, for any vertex $v \in V(G)$ define

$$\begin{aligned} A_c(v) &:= \text{"set of available colors for } v \in V(G) \text{ in the partial coloring } c \text{ of } G\text{"} \\ &= \text{set of all colors in } [k] \text{ that } c \text{ does \textit{not} use on any vertex in the neighborhood } \Gamma(v) \text{ of } v. \end{aligned}$$

A key insight is that we can write the left-hand side of (1.1) as an expectation:

$$\frac{|\mathcal{C}(H)|}{|\mathcal{C}(H-v)|} = \sum_{c \in \mathcal{C}(H-v)} \frac{|A_c(v)|}{|\mathcal{C}(H-v)|} = \mathbb{E}\,|A_c(v)|, \tag{1.2}$$

where in the expectation the (partial) coloring c is chosen uniformly at random from $\mathcal{C}(H-v)$.

To bound the expectation, for technical reasons we will focus our attention on neighbors of v with 'many' available colors. For any neighbor $u \in \Gamma(v)$ of v, since $c \in \mathcal{C}(H-v)$ is chosen uniformly at random we have

$$\mathbb{P}(|A_c(u)| \leqslant t) = \frac{|\{c \in \mathcal{C}(H-v) : |A_c(u)| \leqslant t\}|}{|\mathcal{C}(H-v)|} \leqslant \frac{|\mathcal{C}(H-v-u)| \cdot t}{|\mathcal{C}(H-v)|} \leqslant \frac{t}{\ell},$$

where the first inequality exploits that $|A_c(u)|$ is determined by c restricted to $H-v-u$, and the last inequality uses the induction hypothesis. The above estimate suggests that most $u \in \Gamma(v)$ will have many available colors (if $\ell \gg t$). To make this precise, using Markov's inequality and Linearity of expectation we have

$$\mathbb{P}(|\{u \in \Gamma(v) : |A_c(u)| \leqslant t)\}| \geqslant \varepsilon^2 k) \leqslant \frac{\mathbb{E}(\sum_{u \in \Gamma(v)} \mathbb{1}_{\{|A_c(u)| \leqslant t\}})}{\varepsilon^2 k} \leqslant \frac{\Delta \cdot t/\ell}{\varepsilon^2 k} = \frac{t \log \Delta}{\varepsilon^2 (1+\varepsilon)\Delta^{\varepsilon/2}} \leqslant \varepsilon, \tag{1.3}$$

where we choose t such that the last inequality holds for sufficiently large Δ (depending on ε), say, $t := \Delta^{\varepsilon/4}$. To estimate $\mathbb{E}\,|A_c(v)|$ from below, we plan to 'ignore' colors used by neighbors with few available colors (=not attempt to use them for v), and thus decompose the set of available colors as

$$\begin{aligned} A_c(v) &:= [k] - \bigcup_{u \in \Gamma(v)} c(u) \\ &= \underbrace{\Big[[k] - \bigcup_{u \in \Gamma(v) : |A_c(u)| \leqslant t} c(u) \Big]}_{=: A_c^+(v)} - \bigcup_{u \in \Gamma(v) : |A_c(u)| > t} c(u), \end{aligned}$$

where (1.3) guarantees that $|A_c^+(v)| \geqslant (1-\varepsilon^2)k$ holds with probability at least $1-\varepsilon$. Writing

$$\Gamma_A(v) := \big\{ u \in \Gamma(v) \ : \ |A_c(u)| > t \big\}$$

for the set of neighbors of v with many available colors, using linearity of expectation we obtain

$$\mathbb{E}\,|A_c(v)| = \mathbb{E}\Big(\sum_{s \in A_c^+(v)} \mathbb{1}_{\{s \text{ not used on any } u \in \Gamma_A(v)\}} \Big) =: \star.$$

We shall evaluate this expectation using careful conditioning. Since G is triangle-free, each $A_c(u)$ with $u \in \Gamma(v)$ is determined by the coloring c restricted to $G - v - \Gamma(v)$. Hence, we can condition on a partial coloring c_0 of $G - v - \Gamma_A(v)$. Since $A_c^+(v)$ is determined by $c_0 \in \mathcal{C}(G - v - \Gamma_A(v))$, it follows that

$$\star = \mathbb{E}\Big[\mathbb{E}\Big(\sum_{s \in A_c^+(v)} \mathbb{1}_{\{s \text{ not used on any } u \in \Gamma_A(v)\}} \ \Big| \ c_0 \Big) \Big] = \mathbb{E}\Big[\underbrace{\sum_{s \in A_c^+(v)} \mathbb{P}\big(s \text{ not used on any } u \in \Gamma_A(v) \mid c_0\big)}_{=:\dagger} \Big].$$

Furthermore, for any $u \in \Gamma(v)$ we have $A_c(u) = A_{c_0}(u)$ and thus

$$\Gamma_A(v) := \big\{ u \in \Gamma(v) \ : \ |A_{c_0}(u)| > t \big\}.$$

Conditional on a partial coloring c_0 of $G - v - \Gamma_A(v)$, a key observation is that the conditional distribution of c is uniform and independent. Indeed, extending c_0 by choosing a color in $A_c(u) = A_{c_0}(u)$ for each $u \in \Gamma_A(v)$ yields a valid coloring of $H - v$ (since H is triangle-free), and all colorings of $H - v$ can be obtained this way. Since the coloring $c \in \mathcal{C}(H-v)$ is chosen uniformly at random, it follows[1] that for any such coloring c' we have

$$\mathbb{P}(c = c' \mid c_0) = \prod_{u \in \Gamma_A(v)} \frac{1}{|A_{c_0}(u)|}. \tag{1.4}$$

[1] To formally see why uniform choice of c implies (1.4), first note that for any valid choice of c', i.e., which extends c_0 and satisfies $c' \in \mathcal{C}(H-v)$, the resulting probability $\mathbb{P}(c = c' \mid c_0) = \mathbb{P}(c = c')/\mathbb{P}(c_0)$ does not depend on c'. Since $\sum_{\text{valid } c'} \mathbb{P}(c = c' \mid c_0) = 1$, it thus follows that $\mathbb{P}(c = c' \mid c_0) = 1/(\#\text{ valid } c')$, which establishes equation (1.4).

In particular, each vertex $u \in \Gamma_A(v)$ independently chooses its color uniformly from $A_{c_0}(u)$. Hence

$$\mathbb{P}\big(s \text{ not used on any } u \in \Gamma_A(v) \mid c_0\big) = \prod_{u \in \Gamma_A(v): s \in A_{c_0}(u)} \left(1 - \frac{1}{|A_{c_0}(u)|}\right),$$

which together with the AM-GM inequality[2] as well as $|A_{c_0}(u)| \geqslant t$ and $|\Gamma_A(v)| \leqslant |\Gamma(v)| \leqslant \Delta$ yields

$$\begin{aligned} \dagger &= \sum_{s \in A_c^+(v)} \prod_{u \in \Gamma_A(v): s \in A_{c_0}(u)} \left(1 - \frac{1}{|A_{c_0}(u)|}\right) \geqslant |A_c^+(v)| \left(\prod_{s \in A_c^+(v)} \prod_{u \in \Gamma_A(v): s \in A_{c_0}(u)} \left(1 - \frac{1}{|A_{c_0}(u)|}\right)\right)^{\frac{1}{|A_c^+(v)|}} \\ &= |A_c^+(v)| \Bigg(\prod_{u \in \Gamma_A(v)} \underbrace{\prod_{s \in A_c^+(v) \cap A_{c_0}(u)} \left(1 - \frac{1}{|A_{c_0}(u)|}\right)}_{\geqslant \left(1 - \frac{1}{|A_{c_0}(u)|}\right)^{|A_{c_0}(u)|} \geqslant \left(1 - \frac{1}{t}\right)^t} \Bigg)^{\frac{1}{|A_c^+(v)|}} \geqslant |A_c^+(v)| \left(1 - \frac{1}{t}\right)^{\frac{t\Delta}{|A_c^+(v)|}}, \end{aligned}$$

where the final expression is monotone increasing in $|A_c^+(v)|$. Since $|A_c^+(v)| \geqslant (1-\varepsilon^2)k$ holds with probability at least $1 - \varepsilon$, using $(1 - 1/t)^t \geqslant e^{1-O(1/t)} = e^{1-o(1)}$ it follows that

$$\begin{aligned} \mathbb{E}\,|A_c(v)| &\geqslant \mathbb{E}(\dagger) \geqslant \mathbb{E}\left(|A_c^+(v)| \left(1 - \frac{1}{t}\right)^{\frac{t\Delta}{|A_c^+(v)|}} \right) \\ &\geqslant (1 - \varepsilon) \cdot (1 - \varepsilon^2)k \underbrace{\left(1 - \frac{1}{t}\right)^{\frac{t\Delta}{(1-\varepsilon^2)k}}}_{\geqslant e^{\frac{(1-o(1))\Delta}{(1-\varepsilon^2)k}}} \geqslant \Theta\left(\frac{\Delta}{\log \Delta}\right) \cdot \exp\left(-\frac{(1 - o(1)) \log \Delta}{(1 - \varepsilon^2)(1 + \varepsilon)}\right) \gg \Delta^{\varepsilon/2} = \ell, \end{aligned}$$

where for the last inequality we assumed that $\varepsilon > 0$ is sufficiently small (as we may) and that Δ is sufficiently large (depending on ε). By (1.2) this completes the induction step and thus the proof of Theorem 1.1. $\square$

[2] The inequality of arithmetic and geometric means states that $\sum_{i \in [k]} x_i/k \geqslant (\prod_{i \in [k]} x_i)^{1/k}$ when $x_i \geqslant 0$ for all $i \in [k]$.

MATH 264A: Combinatorics	**Fall 2024**

Lecture 6: Stanley–Wilf Conjecture (Draft)

October 15, 2024

Lecturer: Lutz Warnke — *Scribes: Finn Southerland, Yunseong Jung (writing), Matthew Cho, Natalie Dodson (editing)*

Today's topics:

- Stanley–Wilf Conjecture
- Tools: Extremal Bound, Pigeonhole Principle, Recursion/Induction
- *Source:*
  Excluded permutation matrices and the Stanley–Wilf conjecture (by Adam Marcus and Gabor Tardos)

1 Stanley–Wilf Conjecture: Permutation Avoidance

The goal of this lecture is to show the proof of the Stanley–Wilf conjecture, which states that the number of pattern-avoiding permutations in S_n is $\leq C^n$. Note that this is significantly smaller than the total number of permutations by a factor of $\log n$ in the exponent:

$$C^n \ll |S_n| = n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \approx e^{(1+o(1))n\log n}.$$

To rigorously state the conjecture, we first make a definition.

Definition 1.1. We say $\sigma \in S_n$ *contains* $\pi \in S_k$ $(2 \leq k \leq n)$ if there are $x_1 < x_2 < \ldots < x_k \in [n]$ such that $\sigma(x_i) < \sigma(x_j) \iff \pi(i) < \pi(j)$ for all $1 \leq j \leq k$. Otherwise, we say σ *avoids* π.

Example 1.2. Let $\sigma = 1\underline{5}\underline{2}6\underline{3}\underline{7}\underline{4}8 \in S_8$. Then σ contains $\pi = 3142$, where the x_i are: 2, 4, 5, 7. We can see that these elements appear in the order 5274 in σ, which is the 3rd, 1st, 4th, and 2nd element of this subset. Conversely, one can check that σ avoids 2413.

Definition 1.3. We denote the set of π-avoiding permutations in S_n by $S_n(\pi)$.

Given this definition, we wonder: for a fixed π, how large is $S_n(\pi)$? This is the subject of the Stanley–Wilf conjecture:

Theorem 1.4 (Stanley–Wilf Conjecture: Marcus–Tardos 2004)**.** *For every permutation π, there exists $C = C(\pi) > 0$ such that $|S_n(\pi)| \leq C^n$.*

2 Matrix Avoidance: Generalization to $\{0,1\}$-Matrices

The first key idea in proving the Stanley–Wilf conjecture is to consider the *permutation matrix* P_π associated with the permutation π. We then generalize the idea of containment further to $(0,1)-$matrices, of which permutation matrices are the special case where each matrix has one 1 per column and row.

Definition 2.1. The (0,1)-matrix $A \in \{0,1\}^{n\times n}$ *contains* $B \in \{0,1\}^{k\times k}$ if there are k column and k row indices $x_1 < x_2 < \ldots x_k$, $y_1 < y_2 < \ldots < y_k$ such that if $B_{ij} = 1$, $A_{x_iy_j} = 1$. Otherwise we say A *avoids* B. We write $B \subseteq A$ or $B \subsetneq A$ respectively.

This is best understood by example:

Example 2.2.

$$B := \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \subseteq \begin{bmatrix} \underline{1} & 1 & \underline{1} \\ 1 & 0 & \underline{1} \\ 0 & \underline{1} & 0 \end{bmatrix} =: A.$$

These matrices have the same dimension, so this just means that A has a 1 wherever B does, and possibly more. An example with a larger matrix is

$$B \subseteq \begin{bmatrix} \underline{1} & 1 & 0 & \underline{1} \\ 1 & 0 & 0 & \underline{1} \\ 1 & 0 & 1 & 0 \\ 0 & \underline{1} & 1 & 0 \end{bmatrix} =: C.$$

Here 3 rows and 3 columns are selected, and considered as if they were a $k \times k$ matrix (with $k = 3$).

Note that if P_σ, P_π are the permutation matrices of $\sigma \in S_n$ and $\pi \in S_k$ respectively, these definitions of containment agree. That is, σ avoids π if and only if P_σ avoids P_π. Also, note that the $n \times n$ permutation matrices are a subset of all $n \times n$ $\{0,1\}$-matrices. Therefore, to prove the Stanley–Wilf conjecture, it suffices to bound the size of $M_n(\pi) := \{A \in \{0,1\}^{n\times n} : A \text{ avoids } P_\pi\}$.

By moving from permutation avoidance to the more general matrix avoidance, another conjecture from extremal combinatorics becomes relevant, and it is this conjecture that Marcus and Tardos actually proved.

3 Füredi–Hajnal Conjecture

Let $\mathrm{ex}(n, P)$ be the maximum number of 1's in a P-avoiding matrix $A \in \{0,1\}^{n\times n}$.

Theorem 3.1 (Füredi–Hajnal Conjecture: Marcus–Tardos 200)**.** *For every permutation matrix P, there exists $C = C(P) > 0$ such that $\mathrm{ex}(n, P) \le Cn$.*

It is not immediate that the Füredi–Hajnal Conjecture implies the Stanley–Wilf Conjecture. For example, a simple 'extremal' counting argument (that takes into account the possible number of 1s) only gives

$$|M_n| \le \sum_{0\leqslant i\leqslant Cn} \binom{n^2}{i} = n^{\Theta(n)}.$$

Instead we prove this implication using recursion, subdividing the matrix into blocks.

Lemma 3.2 (Klazar 2000)**.** *The Füredi–Hajnal Conjecture implies the Stanley–Wilf Conjecture.*

Proof. It suffices to show $|M_n(\pi)| = |M_n| \le D^n$ for some $D = D(P)$.

For any $2n \times 2n$ matrix A, we subdivide the matrix into 2×2 blocks as follows. We replace each block by 1 if and only if the 2×2 block contains at least one 1, and by 0 otherwise. This creates $n \times n$ matrix $\tilde{A}$.

Example:

$$\left(\begin{array}{cc|cc|cc|cc} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \hline 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{array}\right) \longrightarrow \left(\begin{array}{c|c|c|c} 1 & 1 & 0 & 1 \\ \hline 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 \\ \hline 0 & 1 & 1 & 0 \end{array}\right)$$

One can see the following properties:

- A avoids $P \implies \tilde{A}$ avoids P. This uses the fact that P is a permutation matrix.
- Each $\tilde{A}$ corresponds to at most $(2^4-1)^{\#1\text{s in } \tilde{A}} \leqslant 15^{Cn}$ many As, assuming the Füredi–Hajnal Conjecture.

Then, since there are $|M_n|$ many choices of $\tilde{A}$ and at most 15^{Cn} ways to extend $\tilde{A}$ to A,

$$|M_{2n}| \leq |M_n| \cdot 15^{Cn}.$$

Now consider $n \geqslant 1$, and suppose that n is of form $n = 2^k + r$ with $0 \leq r < 2^k$. Since $|M_n|$ is monotone increasing in n, using iteration it follows that

$$|M_n| \leq |M_{2^{k+1}}| \leq |M_1| \cdot 15^{C\sum_{1\leq j\leq k+1} 2^{j-1}} \leqslant |M_1| \cdot 15^{C2^{k+1}} \leq 2 \cdot 15^{2Cn},$$

establishing the Stanley–Wilf Conjecture with constant $D := 2 \cdot 15^{2C}$, say. $\square$

In order to prove the conjecture, we need following lemma:

Lemma 3.3. *Suppose that $k^2 \mid n$. Then*

$$\mathrm{ex}(n, P) \leq 2k^3 n\binom{k^2}{k} + (k-1)^2 \mathrm{ex}\big(n/k^2, P\big).$$

Proof. We subdivide any $n \times n$ matrix A that avoids P into $k^2 \times k^2$ blocks as follows. We replace each block by 1 if and only if the $k^2 \times k^2$ block contains at least one by 1, and by 0 otherwise. This creates a $n/k^2 \times n/k^2$ matrix $\tilde{A}$. Note that $\tilde{A}$ avoids P, since A avoids P (similar as in the earlier the 2×2 blocks argument).

Since $\tilde{A}$ avoids P, it follows that A contains at most $k^4 \cdot \mathrm{ex}(n/k^2, P)$ many ones, which is not good enough for our purposes. To improve this estimate we shall distinguish different types of blocks, so that we only need to invoke the worst case bound $\mathrm{ex}(n/k^2, P)$ for blocks with much fewer ones.

A $k^2 \times k^2$ block B is called **wide** if there are 1s in at least k columns. Similarly, a $k^2 \times k^2$ block B is called **tall** if there are 1's in at least k rows. Note that if a block B is neither wide or tall, then it contains at most $(k-1)^2$ many 1s.

Claim 3.4. *The number of wide (tall) blocks is at most $k\binom{k^2}{k} \cdot n/k^2$.*

This claim can be proved by an application of the Pigeonhole Principle, by showing that if there are more wide (tall) blocks than claimed, then the permutation matrix P is contained in A (we defer the proof to the homework problems).

Putting things together, we arrive at

$$\begin{aligned}\mathrm{ex}(n, p) \leq\ & k^4 \cdot \text{number of wide blocks} \\ & + k^4 \cdot \text{number of tall blocks} \\ & + (k-1)^2 \cdot \text{number of other blocks with at least one 1} \\ \leq\ & 2k^3 n\binom{k^2}{k} + (k-1)^2 \cdot \mathrm{ex}\big(n/k^2, P\big).\end{aligned}$$

$\square$

The above recurrence for $\mathrm{ex}(n, P)$ implies Füredi–Hajnal Conjecture with $C = 2k^4\binom{k^2}{k}$: indeed, using induction one can show that $\mathrm{ex}(n, P) \leqslant Cn$ holds for any $n \geqslant 1$ (with some minor care in the case $k^2 \nmid n$).

MATH 264A: Combinatorics	**Fall 2024**

Lecture 7: Entropy Review

October 17, 2022

Lecturer: Lutz Warnke

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Today's topics:

- Entropy Review:
  - Definition and Key Properties
  - Connection to Counting
  - Basic Counting Examples

1 Definition of Entropy (Shannon, 1948)

Consider a random variable X which takes values in some finite set $\mathcal{S}$.
The *Entropy of the random variable* X "measures the information content carried by X", and is defined as:

$$\mathrm{H}(X) := -\sum_{x \in \mathcal{S}} p_x \log_2 p_x \qquad \text{where} \quad p_x := \mathbb{P}(X = x),$$

using the usual convention $0 \log_2 0 = 0$ (as $\lim_{x \to 0^+} x \log_2 x = 0$)

1.1 Natural Connection to Stirling's formula

The Stirling's formula gives a very good approximation of factorials:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \approx \left(\frac{n}{e}\right)^n \quad \text{(ignoring the polynomial term)}$$

Now consider the following approximation

$$\begin{aligned}
\binom{n}{pn} &= \#\ pn\text{-element subsets of } [n] = \{1, 2, \ldots, n\} \\
&= \frac{n!}{((1-p)n)!(pn)!} \approx \frac{\left(\frac{n}{e}\right)^n}{\left(\frac{(1-p)n}{e}\right)^{(1-p)n}\left(\frac{pn}{e}\right)^{pn}} \\
&= \frac{1}{((1-p)^{(1-p)}p^p)^n} = 2^{(-(1-p)\log_2(1-p) - p\log_2 p)n}.
\end{aligned}$$

In the exponent, we can see that the entropy formula $\mathrm{H}(X) = -(1-p)\log_2(1-p) - p\log_2 p$ naturally appears if we consider X to be a Bernoulli random variable with $\mathbb{P}(X = 1) := p$ and $\mathbb{P}(X = 0) := 1 - p$.

1.2 Connection to Counting

The following key property shows that a bound on $\mathrm{H}(X)$ will give an estimate on the cardinality of $\mathcal{S}$.

Key Property ("Maximality of uniform distribution"). *Writing Support*$(X) = \{x \in \mathcal{S} : \mathbb{P}(X = x) > 0\}$, *we have* $\mathrm{H}(X) \le \log_2 |\mathit{Support}(X)|$, *with equality iff* X *has uniform distribution.*

Proof. Let $\mathcal{S} := \text{Support}(X)$. Then

$$\mathrm{H}(X) = \sum_{x \in \mathcal{S}} p_x \log_2 \frac{1}{p_x}.$$

Note that $\sum_{x \in \mathcal{S}} p_x = 1$. Since the log-function is concave, Jensen's inequality yields

$$\mathrm{H}(X) \leq \log_2 \left(\sum_{x \in \mathcal{S}} p_x \cdot \frac{1}{p_x} \right) = \log_2(|\text{Support}(X)|).$$

Equality in Jensen's holds when the $1/p_x$ are all same, which happens for the uniform distribution. □

<u>**Entropy as Counting Tool**</u>: If X is chosen uniformly from $\mathcal{C}$, then

$$|\mathcal{C}| = 2^{\mathrm{H}(X)}.$$

Hence knowledge about $\mathrm{H}(X)$ translates into bounds on $|\mathcal{C}|$. (Note that X can be a vector of random variables such that X is uniformly distributed, but each element need not be uniformly distributed.)

1.3 Entropy of multiple random variables

The concept of entropy naturally extends to multiple random variables.
The *Joint Entropy* of X and Y is given by

$$\mathrm{H}(X, Y) := \mathrm{H}\big((X, Y)\big) = -\sum_{x,y} p_{x,y} \log_2(p_{x,y}),$$

where $p_{x,y} = \mathbb{P}(X = x, Y = y)$. The *Conditional Entropy* of X given Y is given by

$$\mathrm{H}(X|Y) := \sum_{y} \mathbb{P}(Y = y) \underbrace{\mathrm{H}(X|Y = y)}_{= -\sum_x p_{x|y} \log_2(p_{x|y})},$$

where $p_{x|y} = \mathbb{P}(X = x | Y = y)$.

Key Properties. *For our counting purposes, the following properties will allow us to estimate the Entropy:*

(a) <u>*Sub-additivity*</u>*:* $\mathrm{H}(X_1, X_2, \ldots, X_k) \leq \sum_{i \in [k]} \mathrm{H}(X_i)$
"Information in $X_1, X_2, \ldots, X_k$ individually is $\geq$ information in $(X_1, X_2, \ldots, X_k)$ together"

(b) <u>*Chain Rule*</u>*:* $\mathrm{H}(X_1, X_2, \ldots, X_k) = \sum_{i \in [k]} \mathrm{H}(X_i | X_1, X_2, \ldots, X_{i-1})$
"Entropy can be understood by gradually revealing variables one-by-one"

(c) <u>*Dropping Conditioning*</u>*:* $\mathrm{H}(X|Y) \leq \mathrm{H}(X)$*, with equality iff X and Y are independent*
"Conditioning can only reduce entropy" (Caution: $\mathrm{H}(X|Y = y)$ can be larger than $\mathrm{H}(X)$.)

Proof. Proof of (a): follows by combining (b) and (c).
Proof of (b) for $k = 2$: Note that

$$\mathrm{H}(X|Y) = \sum_{y} \mathbb{P}(Y = y) \sum_{x} \mathbb{P}(X = x|Y = y) \log_2 \left(\frac{1}{\mathbb{P}(X = x|Y = y)} \right). \tag{1.1}$$

Using Bayes rule we have $\mathbb{P}(X = x|Y = y) = \mathbb{P}(X = x, Y = y)/\mathbb{P}(Y = y)$, so that

$$\begin{aligned}
\mathrm{H}(X|Y) &= \sum_{x,y} \mathbb{P}(X = x, Y = y) \Big(\log_2 \mathbb{P}(Y = y) - \log_2 \mathbb{P}(X = x, Y = y) \Big) \\
&= \sum_{x,y} \mathbb{P}(X = x, Y = y) \log_2 \mathbb{P}(Y = y) - \sum_{x,y} \mathbb{P}(X = x, Y = y) \log_2 \mathbb{P}(X = x, Y = y)
\end{aligned}$$

$$= \sum_y \mathbb{P}(Y=y)\log_2 \mathbb{P}(Y=y) + \mathrm{H}(X,Y) = -\mathrm{H}(Y) + \mathrm{H}(X,Y).$$

Therefore $\mathrm{H}(X,Y) = \mathrm{H}(X|Y) + \mathrm{H}(Y)$. (We can use induction to extend this proof to $k > 2$.)
Proof of (c): Using $\mathbb{P}(Y=y)\,\mathbb{P}(X=x|Y=y) = \mathbb{P}(X=x)\,\mathbb{P}(Y=y|X=x)$ we re-write equation (1.1) as

$$\mathrm{H}(X|Y) = \sum_x \mathbb{P}(X=x) \sum_y \mathbb{P}(Y=y|X=x)\log_2\left(\frac{1}{\mathbb{P}(X=x|Y=y)}\right).$$

Since $\sum_y \mathbb{P}(Y=y|X=x) = 1$, using Jensen's inequality and $\sum_y \mathbb{P}(Y=y) = 1$ we obtain that

$$\begin{aligned}
\mathrm{H}(X|Y) &\leq \sum_x \mathbb{P}(X=x)\log_2\left(\sum_y \frac{\mathbb{P}(Y=y|X=x)\,\mathbb{P}(X=x)\,\mathbb{P}(Y=y)}{\mathbb{P}(X=x|Y=y)\,\mathbb{P}(Y=y)\,\mathbb{P}(X=x)}\right)\\
&= \sum_x \mathbb{P}(X=x)\log_2\left(\sum_y \frac{\mathbb{P}(Y=y)}{\mathbb{P}(X=x)}\right) = \mathrm{H}(X),
\end{aligned}$$

Equality in Jensen's holds when the $1/\,\mathbb{P}(X=x|Y=y)$ are the same for all y, so that $\mathrm{H}(X|Y) = \mathrm{H}(X)$ holds iff X and Y are independent. □

2 Two Simple Counting Examples (using Entropy)

The "Binary entropy function" is defined as

$$\mathrm{h}(p) := -p\log_2(p) - (1-p)\log_2(1-p).$$

2.1 Example 1

Claim 2.1. $|\mathcal{C}| \leq 2^{\sum_{i\in[n]} \mathrm{h}(p_i)}$, *where*
$\mathcal{C}$ = *collection of subsets of* $[n]$,
p_i = *fraction of subsets (from* $\mathcal{C}$*) that contain element* $i \in [n]$.

Proof. Choose a set X uniformly at random from $\mathcal{C}$, and write $X = (X_1, X_2, \ldots, X_n)$ with

$$X_i = \begin{cases} 0 & \text{if } i \notin X, \\ 1 & \text{if } i \in X. \end{cases}$$

Then

$$|\mathcal{C}| \overset{\text{unif.}}{=} 2^{\mathrm{H}(X)} \overset{\text{sub-add.}}{\leqslant} 2^{\sum_{i\in[n]} \mathrm{H}(X_i)}.$$

Noting that $\mathbb{P}(X_i = 1) = p_i$ and $\mathbb{P}(X_i = 0) = 1 - p_i$ by assumption, we infer $\mathrm{H}(X_i) = \mathrm{h}(p_i)$. □

2.2 Example 2

Claim 2.2. $\sum_{0\leq i\leq k} \binom{n}{i} \leq 2^{\mathrm{h}(k/n)n}$ *for each* $0 \leq k \leq n/2$

Proof. Define $\mathcal{C} := \{\text{all subsets of } [n] \text{ with } \leq k \text{ elements}\}$, so that $\sum_{i=0}^{k} \binom{n}{k} = |\mathcal{C}|$ by construction.
Draw uniform random $X \in \mathcal{C}$, and write $X = (X_1, X_2, \ldots, X_n)$ with

$$X_i = \begin{cases} 0 & \text{if } i \notin X, \\ 1 & \text{if } i \in X. \end{cases}$$

Then

$$|\mathcal{C}| \overset{\text{unif.}}{=} 2^{\mathrm{H}(X)} \overset{\text{sub-add.}}{\leqslant} 2^{\sum_{i=1}^{n} \mathrm{H}(X_i)}.$$

How to determine $\mathbb{P}(X_i = 1)$ and thus $\mathrm{H}(X_i)$? By symmetry $\mathbb{P}(X_i = 1) = \mathbb{P}(X_j = 1)$ holds, so that

$$\mathbb{P}(X_i = 1) = \frac{1}{n}\sum_{j=1}^{n} \underbrace{\mathbb{P}(X_j = 1)}_{=\mathbb{E}[X_j]} = \frac{1}{n}\mathbb{E}\Bigg[\underbrace{\sum_{j=1}^{n} X_j}_{=|X|\leq k}\Bigg] \leq \frac{k}{n}.$$

Hence $\mathrm{H}(X_i) = \mathrm{h}\big(\mathbb{P}(X_i = 1)\big) \leq \mathrm{h}(k/n)$, since $\mathrm{h}(p)$ is monotone increasing for $0 \leq p \leq 1/2$. $\square$

3 Final Remarks and Outlook

To bound $|\mathcal{C}|$, in this lecture we

(i) Defined X as uniform choice on $\mathcal{C}$ so that $|\mathcal{C}| = 2^{\mathrm{H}(X)}$.
(ii) Wrote $X = (X_1, \ldots, X_n)$ and applied sub-additivity, so that $|\mathcal{C}| \leq 2^{\sum_{1\leq i\leq n} \mathrm{H}(X_i)}$.

In the next lecture we will refine (ii) and:

(a) Replace sub-additivity by chain rule $\mathrm{H}(X) = \sum_i \mathrm{H}(X_i|X_1, \ldots, X_{i-1})$.
(b) Reveal the X_i in *random* order.

MATH 264A: Combinatorics **Fall 2024**

Lecture 8: Entropy Method I

October 22, 2024

Lecturer: Lutz Warnke

Today's topics:

- Entropy Method I:
  - Counting via Randomized Chain-Rule (Chain-Rule and Random Order of Variables)
  - Example: Bregman's Theorem

1 Matchings in bipartite graphs

We want to find a perfect matching of G, i.e. a vertex disjoint collection of edges that contain all vertices.

Theorem 1.1 (Bregman 1973). *Let G be a bipartite graph with vertex classes $V(G) = L \cup R$. Then*

$$\#\textit{perfect matchings} = \#PM(G) \leq \prod_{v \in L} (d_v!)^{1/d_v},$$

where d_v is the degree of v in G.

We remark that this is sharp when $d \mid n$, for n being the size of one of the parts of the bipartite graph. For $K_{d,d}$, the complete bipartite graph with two parts of size d, we take n/d copies of $K_{d,d}$. Then,

$$\#\mathrm{PM}(G) = (d!)^{n/d} = \prod_{1 \leq j \leq \frac{n}{d}} \prod_{1 \leq i \leq d} (d!)^{1/d}.$$

1.1 Proof via Entropy Method (Jaikumar Radhakrishnan 1997)

1. WLOG $L = \{1, \ldots, n\}$.
2. Random experiment: choose X uniformly at random from all PM of G. Note that $\#\mathrm{PM}(G) = 2^{H(X)}$.
3. Representation of X as $X = (X_1, \ldots, X_n)$ where X_i is the (unique) neighbor of i in the matching.

Attempt 1: Subadditivity.

$$\log_2(\#\mathrm{PM}(G)) \overset{\text{unif}}{=} H(X) \overset{\text{subadd}}{\leq} \sum_{1 \leq i \leq n} H(X_i) \overset{\max}{\leq} \sum_{1 \leq i \leq n} \log_2(\#\text{poss choices for } X_i) \leq \sum_{1 \leq i \leq n} \log_2 d_i.$$

This implies that

$$\#\mathrm{PM}(G) \leq 2^{\sum_{1 \leq i \leq n} \log_2 d_i} = \prod_{i=1}^{n} d_i.$$

Note that this is larger than $(d_i!)^{1/d_i} \sim (2\pi d_i)^{1/d_i} \frac{d_i}{e}$; this is off by a factor of e. Thus, we will need a better entropy estimate.

Attempt 2: Chain rule.

$$\log_2(\#\mathrm{PM}(G)) \overset{\text{unif}}{=} H(X) \overset{\text{CR}}{=} \sum_{1 \leq i \leq n} H(X_i \mid X_1, \ldots, X_{i-1}) \leq \sum_{1 \leq i \leq n} \log_2 |\Gamma(i) - \{X_1, \ldots, X_{i-1}\}|,$$

where $|\Gamma(i) - \{X_1, \ldots, X_{i-1}\}|$ is the number of neighbors of i that are not among $X_1, \ldots, X_{i-1}$, or equivalently, the number of neighbors of i that are available as a matching-partner of i.

Problem: unclear how to improve the $\log_2 |\Gamma(i) - \{X_1, \ldots, X_{i-1}\}| \le \log_2 d_i$ estimate. We need to try to refine this entropy estimate more.

Refinement (i): Rewrite conditional entropy estimate in a more convenient/usable form:

$$\begin{aligned} H(X_i \mid X_1, \ldots, X_{i-1}) &\stackrel{\text{def}}{=} \sum_{x_1, \ldots, x_{i-1}} \mathbb{P}(X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) \cdot H(X_i \mid X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) \\ &\le \sum_{x_1, \ldots, x_{i-1}} \mathbb{P}(X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) \log_2 |\Gamma(i) - \{X_1, \ldots, X_{i-1}\}| \\ &\le \mathbb{E}_{X_1, \ldots, X_{i-1}} \log_2 |\Gamma(i) - \{X_1, \ldots, X_{i-1}\}| = \mathbb{E}_X \log_2 |\Gamma(i) - \{X_1, \ldots, X_{i-1}\}|, \end{aligned}$$

where we used that $|\Gamma(i) - \{X_1, \ldots, X_{i-1}\}|$ only depends on $X_1, \ldots, X_{i-1}$.

Refinement (ii): All of the above also works for some fixed order $\sigma : [n] \to [n]$ of $L = \{1, \ldots, n\}$. In particular,

$$\log_2(\#\mathrm{PM}(G)) = H(X) = \sum_{1 \le i \le n} H(X_i \mid X_j : j <_\sigma i) \le \sum_{1 \le i \le n} \mathbb{E}_X \log_2 \underbrace{|\Gamma(i) - \{X_j : j <_\sigma i\}|}_{=: N_i(\sigma, X)}, \tag{1.1}$$

where $j <_\sigma i$ if $\sigma(j) < \sigma(i)$.

Problem: Different orderings σ will give different estimates.

Attempt 3: Randomized chain rule.
We will average Equation 1.1 by taking a random ordering $\pi : [n] \to [n]$ of vertices in $L = \{1, \ldots, n\}$. Since the left hand side of (1.1) does not depend on the order of X, using independence of π, X it follows that

$$\log_2(\#\mathrm{PM}(G)) = H(X) \le \mathbb{E}_\pi\left(\sum_{1 \le i \le n} \mathbb{E}_X \log_2 N_i(\pi, X)\right) = \mathbb{E}_X\left(\sum_{1 \le i \le n} \mathbb{E}_\pi(\log_2 N_i(\pi, X) \mid X)\right).$$

Key point: for fixed $X = (X_1, \ldots, X_n)$, now $N_i(\pi, X) = |\Gamma(i) - \{X_j : j <_\sigma i\}|$ has a distribution (due to random π) which we can calculate and thus improve on $\log_2 N_i(\pi, X) \le \log_2 d_i$.

Note that each neighbor of i has a unique matching partner (in the same vertex class as i), and we only care about the order of these d_i vertices in determining $N_i(\pi, X)$. In particular, $N_i(\pi, X)$ equals

- If i comes first: $|\Gamma(i)| = d_i$.
- If i comes 2nd: $|\Gamma(i)| - 1 = d_i - 1$.

  $\vdots$
- If i comes last: 1 (the unique matching partner of i in X).

Note that direct enumeration gives

$$\begin{aligned} \mathbb{P}(N_i(\pi, X) = j \mid X) &= \frac{\#\text{permuations } \pi \text{ with } N_i(\pi, X) = j}{n!} \\ &= \frac{\binom{n}{d_i} \cdot 1 \cdot (d_i - 1)! \cdot (n - d_i)!}{n!} = \frac{n!}{(n - d_i)! d_i!} \cdot \frac{(d_i - 1)!(n - d_i)!}{n!} = \frac{1}{d_i}, \end{aligned}$$

where we choose the d_i locations of the matching partners, place i in the j-th position, permute these other matching partners, and place the rest. (Alternatively, we can note that since the location of the matching partners is determined by X and thus fixed, in a random permutation, each position has the same probability $\frac{1}{d_i}$ by symmetry.)

Hence

$$\mathbb{E}_\pi\left(\log_2 N_i(\pi, X) \mid X\right) = \frac{1}{d_i} \log_2 d_i + \frac{1}{d_i} \log_2(d_i - 1) + \cdots + \frac{1}{d_i} \log_2 1 = \frac{1}{d_i} \log_2(d_i!).$$

Thus

$$\log_2(\#\mathrm{PM}(G)) \le \mathbb{E}_X\left(\sum_{1\le i\le n} \frac{1}{d_i}\log_2(d_i!)\right) = \sum_{1\le i\le n} \frac{1}{d_i}\log_2(d_i!),$$

which implies $\#\mathrm{PM}(G) \le \prod_{1\le i\le n}(d_i!)^{1/d_i}$, as desired. □

2 Final Remarks and Outlook

General Pattern: today, we used

1. Conditional on $X_1, \ldots, X_{i-1}$, we could bound the number of choices for X_i by $N_i(X_1, \ldots, X_{i-1})$.
2. Randomized chain rule:

$$H(X) \le \mathbb{E}_X\left(\sum_{1\le i\le n} \mathbb{E}_\pi(\log_2 N_i(X_j : j <_\pi i) \mid X)\right),$$

   and then we calculated the distribution of $N_i(X_j : j <_\pi i)$.

Next time, we will use

1. Apply Jensen's inequality to $\mathbb{E}_\pi(\log_2 N_i(X_j : j <_\pi i) \mid X)$ to get

$$H(X) \le \mathbb{E}_X\left(\sum_{1\le i\le n} \log_2 \mathbb{E}_\pi(N_i(X_j : j <_\pi i) \mid X)\right).$$

   Writing $N_i(X_j : j <_\pi i) = \sum I_k$ for $I_k \in \{0,1\}$, it suffices to estimate $\mathbb{E}_\pi I_k = \mathbb{P}_\pi(I_k = 1)$, which is a conceptually much simpler task than determining the distribution of $N_i(X_j : j <_\pi i)$.
2. Use random ordering based on (independent) continuous time construction simplifies calculations.

MATH 264A: Combinatorics **Fall 2024**

Lecture 9: Entropy Method II

October 24, 2024

Lecturer: Lutz Warnke

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Today's topics:

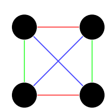
- Entropy Method II:
  - Refine entropy framework from last class
  - Two combinatorial applications

- *Sources:*
  arXiv:1108.5042 (An Upper bound on the number of Steiner triple systems, by Linial-Luria)
  arXiv:1705.05225 (New bounds on the number of n-queens configurations, by Luria)

1 1-factorization of K_n

Definition 1.1. A 1-factorization of K_n is a partition of the edges of K_n into $n-1$ perfect matchings. Equivalently, it is a proper edge-coloring of K_n using $n-1$ colors. We define a function $F(n)$ to be the number of 1-factorizations of K_n. Note this is only nonzero for n even.

As an example, the following is a 1-factorization of K_4.



Now we will prove an asymptotic upper bound on $F(n)$, namely that

$$\mathrm{F}(n) \leqslant \left((1+o(1))\frac{n}{e^2}\right)^{\binom{n}{2}}.$$

As usual, let X be a uniformly random 1-factorization of K_n. Write $X = (X_e)_{e\in E(K_n)}$, where X_e is the color of e in X. In order for the method to work, we need to be able to reconstruct X from knowing every X_e; this is certainly the case here. Using the entropy framework from last lecture, we know that

$$\log_2 \mathrm{F}(n) \leqslant \sum_{e\in E(K_n)} \mathbb{E}_X\big(\mathbb{E}_\pi\left(\log_2 N_e^\pi \mid X\right)\big),$$

where π is a random ordering of the edges, and N_e^π is the number of values X_e can take given the previous X_f as ordered in π.

Last time, we picked a uniformly random permutation, but this starts to complicate matters when we want to condition. Instead, we will generate π in a way that maximizes independence. To do this, we will have each edge e independently pick an arrival time $\pi_e \in [0,1]$ uniformly at random, and then order edges in decreasing order of $(\pi_e)_e$. In other words, if $\pi_e < \pi_f$, then the edge f arrives before e. Note that collisions have probability 0.

Now for each edge e, it will be useful to condition on π_e, and write

$$\mathbb{E}_\pi\left(\log_2 N_e^\pi \mid X\right) = \mathbb{E}_\pi\big(\mathbb{E}_\pi\left(\log_2 N_e^\pi \mid X, \pi_e\right) \mid X\big) \overset{\text{Jensen's}}{\leqslant} \mathbb{E}_\pi\big(\log_2 \mathbb{E}_\pi\left(N_e^\pi \mid X, \pi_e\right) \mid X\big).$$

Now we can write N_e^π as a sum of indicator random variables. We have

$$N_e^\pi = \sum_{c=1}^{n} \mathbb{1}_{\{c \text{ available at } e\}},$$

so that

$$\mathbb{E}(N_e^\pi \mid X, \pi_e) = \sum_{c=1}^{n-1} \mathbb{P}(c \text{ available at } e \mid X, \pi_e).$$

The color $X(e)$ will always be available - since X is a proper coloring, no other edge adjacent to e can have the same color. For any other color c, there will be exactly two edges adjacent to e with color c (since we use $n-1$ colors, each color must be adjacent to every vertex, so e will have one edge of color c sharing each of its endpoints). The color c is available if and only if both of those edges arrive after e does, which is to say that their arrival times are at most π_e. But the distribution is uniform and independent, so this happens with probability π_e^2. Hence we have

$$\mathbb{E}_\pi(N_e^\pi \mid X, \pi_e) = 1 + (n-2)\pi_e^2,$$

so that

$$\mathbb{E}_\pi\left(\log_2 \mathbb{E}_\pi\left(N_e^\pi \mid X, \pi_e\right)\right) = \mathbb{E}_\pi\left(\log_2(1+(n-2)\pi_e^2)\right) = \int_0^1 \log_2(1+(n-2)t^2)\,\mathrm{d}t.$$

By asymptotic methods (details omitted here, see the homework problems), it follows that

$$\begin{aligned}\int_0^1 \log_2(1+(n-2)t^2)\,\mathrm{d}t &= \int_0^1 \log_2((n-2)t^2)\,\mathrm{d}t + O\left(\frac{1}{\sqrt{n}}\right)\\ &= \log_2(n-2) + \int_0^1 \log_2(t^2)\,\mathrm{d}t + O\left(\frac{1}{\sqrt{n}}\right)\\ &= \log_2(n-2) + \frac{2}{\ln 2} + O\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Putting things together, this finally allows us to conclude that

$$\log_2 \mathrm{F}(n) \leqslant \sum_e \left[\log_2(n-2) + \frac{2}{\ln 2} + O\left(\frac{1}{\sqrt{n}}\right)\right] = \binom{n}{2}\left(\log_2(n) + \frac{2}{\ln 2} + O\left(\frac{1}{\sqrt{n}}\right)\right),$$

which in turn gives

$$\mathrm{F}(n) \leqslant \left(2^{\log_2 n} \cdot 2^{\frac{-2}{\ln 2}} \cdot 2^{O\left(\frac{1}{\sqrt{n}}\right)}\right)^{\binom{n}{2}} = \left(n \cdot e^{-2} \cdot e^{O\left(\frac{1}{\sqrt{n}}\right)}\right)^{\binom{n}{2}}.$$

Using that for small $x = o(1)$ we have $e^x = 1 + x + O(x^2) = 1 + x(1+o(1))$, it follows that

$$\mathrm{F}(n) = \left(\frac{n}{e^2}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right)\right)^{\binom{n}{2}},$$

which gives the claimed upper bound on $\mathrm{F}(n)$.

2 Steiner Triple Systems of K_n

Definition 2.1. A Steiner triple system on K_n is an edge-disjoint collection of triangles that cover K_n (only exists when $n \equiv 1, 3 \mod 6$). Define $\mathrm{STS}(n)$ to be the number of Steiner triples systems of K_n.

We will prove an asymptotic upper bound on STS(n), namely that

$$\mathrm{STS}(n) \leqslant \left((1+o(1))\frac{n}{e^2}\right)^{\frac{1}{3}\binom{n}{2}},$$

which in fact holds with asymptotic equality when STS(n) is nonzero (thought the $o(1)$ term in the upper and lower bounds differ).

As before, let X be a uniformly random Steiner triple system of K_n. Write $X = (X_e)_{e \in E(K_n)}$, where X_e is the unique vertex k belonging to the triangle containing e in X. We may write

$$\log_2 \mathrm{STS}(n) \leqslant \sum_{e \in E(K_n)} \mathbb{E}_X\Big(\mathbb{E}_\pi\big(\mathbb{E}_\pi\left(\log_2 N_e^\pi \mid X, \pi_e\right) \mid X\big)\Big).$$

This situation is a little more complicated than the previous one: we only need to look at $1/3\binom{n}{2}$ edges to determine the system. Once we have seen an edge e, we know the triangle it belongs to. If the other edges of this triangle are f and g, then $X(f), X(g)$ are completely determined once we've seen $X(e)$. Conversely, if we saw f or g before e, then $N_e^\pi = 1$. (Recall that this is all valid, since we are conditioning on the Steiner triple system X on the outside).

With this in mind we define the event $F_e = \{X_f, X_g \text{ come after } X_e\}$. As before, $\mathbb{P}(F_e \mid \pi_e, X) = \pi_e^2$. Since $\log_2 1 = 0$, it follows that

$$\mathbb{E}_\pi(\log_2 N_e^\pi \mid X, \pi_e) = \mathbb{P}(F_e \mid \pi_e, X) \cdot \mathbb{E}_\pi(\log_2 N_e^\pi \mid X, \pi_e, F_e) \overset{\text{Jensen's}}{\leqslant} \pi_e^2 \cdot \log_2 \mathbb{E}_\pi(N_e^\pi \mid X, \pi_e, F_e).$$

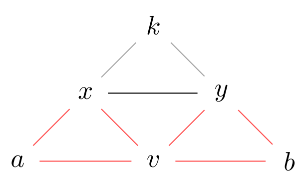
Let $X(e) = k$. Then k will always be available for e. Of the other $n-1$ vertices, $X(e)$ can never be either endpoint in e, but otherwise could be anything. Hence we can write

$$N_e^\pi = 1 + \sum_{\substack{v \in [n] \\ v \notin \{k\} \cup e}} \mathbb{1}_{\{v \text{ available for } e\}},$$

so that

$$\mathbb{E}_\pi(N_e^\pi \mid X, \pi_e, F_e) = 1 + \sum_{\substack{v \in [n] \\ v \notin \{k\} \cup e}} \mathbb{P}(v \text{ available for } e \mid X, \pi_e, F_e).$$

The problem now reduces to computing the probability that a particular vertex v is available. If $e = xy$, then the edges vx and vy each belong to a unique triangle in X.



If any red edge arrives before e, that will rule out v for $X(e)$: if for example ax arrives before e, we would know that xv belongs to the triangle axv, and so v can't be the third vertex for e. Conditioning on F_e ensures that neither gray edge arrives before e, so we only need that none of the six red edges arrive before e. This happens with probability π_e^6, so

$$\mathbb{E}_\pi(N_e^\pi) = 1 + (n-3)\pi_e^6.$$

We can now compute

$$
\begin{aligned}
\mathbb{E}_\pi\left(\log_2 \mathbb{E}_\pi\left(N_e^\pi \mid \pi_e, X\right) \mid X\right) &\leqslant \mathbb{E}_\pi\left(\pi_e^2 \log_2(1+(n-3)\pi_e^6) \mid X\right) \\
&= \int_0^1 t^2 \log(1+(n-3)t^6)\,\mathrm{d}t \\
&\overset{\text{asymp. meth.}}{=} \int_0^1 t^2 \log((n-3)t^6)\,\mathrm{d}t + O\left(\frac{1}{\sqrt{n}}\right) \\
&\leqslant \frac{1}{3}\left(\log(n-3) - \frac{2}{\ln 2}\right) + O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

This is independent of X, so taking the expectation with respect to X and then summing over all $\binom{n}{2}$ edges gives us that

$$
\log_2 \mathrm{STS}(n) \leqslant \binom{n}{2}\left(\frac{1}{3}\left(\log(n-3) - \frac{2}{\ln 2}\right) + O\left(\frac{1}{\sqrt{n}}\right)\right),
$$

which in turn gives

$$
\mathrm{STS}(n) \leqslant \left(\frac{n}{e^2}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right)\right)^{\frac{1}{3}\binom{n}{2}}.
$$

which gives the claimed upper bound on $\mathrm{STS}(n)$. And as a matter of fact, this is best possible up to the error term in the base.

3 Final Remarks

- The essence of the entropy method proofs was to use Jensen's to reduce from understanding what N_e^π is, to understanding if one option is available – which is conceptually a much easier task (and at the core of the method).
- The asymptotic evaluation of the resulting integral is a more mechanical/routine part of the entropy method.

MATH 264A: Combinatorics **Fall 2024**

Lecture 11: Asymptotic Methods: Estimating Sums by Integrals

October 31, 2024

Lecturer: Lutz Warnke

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Today's topics:

- Estimating sums by integrals
- Laplace Method Idea (identify range of 'main contribution', and use different estimates in different ranges).

- *Source:* Sedgewick and Flajolet (An Introduction to the Analysis of Algorithms), Sections 4.4-4.7

Today, we are interested in estimating sums by integrals and, as a general principle, identifying which parts of a sum contribute the most. We have more powerful and refined techniques for analyzing integrals and may get more information that we could obtain by looking at the sum alone.

1 Motivating Example

There are (at least) two natural approaches to analyze the following number:

$$a_n = \textit{number of tilings of a } 2 \times n \textit{ rectangles with } 2 \times 1 \textit{ rectangles/dominoes}.$$

Recursion: $a_n = a_{n-1} + a_{n-2}$
By analyzing the generating function $A(x) = \sum_{n\geqslant 0} a_n x^n$, using partial fractions one can obtain, as $n \to \infty$,

$$a_n = [x^n]A(x) = \frac{1}{\sqrt{5}}\left(\left(\tfrac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\tfrac{1-\sqrt{5}}{2}\right)^{n+1}\right) \sim \frac{1}{\sqrt{5}}\left(\tfrac{1+\sqrt{5}}{2}\right)^{n+1}.$$

Explicit Formula: $a_n = \sum_{k=0}^{\lfloor n/2\rfloor}\binom{n-k}{k}$
In this 'asymptotic methods' part of the course we will learn techniques how to obtain the asymptotic estimate

$$a_n = \sum_{k=0}^{\lfloor n/2\rfloor}\binom{n-k}{k} \sim \frac{1}{\sqrt{5}}\left(\tfrac{1+\sqrt{5}}{2}\right)^{n+1}$$

directly from this 'complicated' sum (by comparison of the sum with suitable integrals). As we shall illustrate in the homework problems, one advantage of this approach is that it is usually more robust, since small modifications of such sums can often be handled without much additional effort.

2 Basic Estimation Techniques

2.1 Regions of Contribution

The first concept we need is that different regions of a sum may contribute more to the total. As a simple yet illustrative example, we consider finite sums. There are some extreme cases that are useful to look at.

Principle 1: *Tails of rapidly decreasing sums are negligible.* Consider

$$D_N := N! \sum_{0\le k\le N} \frac{(-1)^k}{k!} = N!e^{-1} - N! \sum_{k>N} \frac{(-1)^k}{k!}.$$

We define R_N to be the tail term above. Then

$$|R_N| \le \sum_{k>N} \frac{N!}{k!} \le \frac{1}{N+1} + \frac{1}{(N+1)^2} + \frac{1}{(N+1)^3} + \cdots = \sum_{j\ge 1} \left(\frac{1}{N+1}\right)^j = \frac{\frac{1}{N+1}}{1-\frac{1}{N+1}} = \frac{1}{N},$$

so that $D_N = N!e^{-1} + O(1/N)$, i.e., here the main contribution to the sum comes from the first term alone.

Principle 2: *The last term(s) of rapidly increasing sums are dominant.* As an example,

$$\sum_{0\le k\le N} k! = N!\left(1 + \frac{(N-1)!}{N!} + \sum_{0\le k\le N-2} \frac{k!}{N!}\right) = N!\left(1 + O\left(\frac{1}{N}\right)\right),$$

because for $k \le N-2$, $\frac{k!}{N!} \le \frac{1}{(N-1)N}$, so the sum $\sum_{0\le k\le N-2} \frac{k!}{N!}$ is less than $\frac{1}{N}$. In other words, here the main contribution to the sum comes from the last term alone.

This is much more nuanced than a blanket "number of terms times largest terms" approach. Different parts of a sum have different contributions, and should be estimated differently.

2.2 Approximation by Integral

The next idea that we make note of is *estimating the sums by the integral:* $\sum_{a\le k<b} f(k) \approx \int_a^b f(x)dx$. From calculus, we approximate an integral by looking at its Riemann sum; here we are doing the reverse. Explicitly,

$$\sum_{a\le k<b} f(k) - \int_a^b f(x)dx = \sum_{a\le k<b} \left(f(k) - \int_k^{k+1} f(x)dx\right).$$

We want to estimate the error term. A simple bound gives that

$$\sum_{a\le k<b} f(k) = \int_a^b f(x)dx + \Delta, \qquad \text{where} \qquad |\Delta| \le \sum_{a\le k<b} \max_{x\in[k,k+1]} |f(k) - f(x)|.$$

The monotone case is special: when the function f is monotone increasing or decreasing, then the bound on $|\Delta|$ telescopes, and all we are left with is the difference between the endpoints, so

$$|\Delta| \le |f(b) - f(a)|.$$

In applications, we can often decompose a sum into monotone subintervals.

The Euler-Maclaurin summation formula *takes the derivatives of the function into account to get a more precise estimate.* One version of the formula gives

$$|\Delta| \le \int_a^b |f'(x)|dx,$$

although (much) more refined versions exist.

2.2.1 Examples

Example 1 Consider the Nth harmonic number $H_N := \sum_{k=1}^N \frac{1}{k}$. Exploiting monotonicity, we obtain

$$\left|\sum_{1\le k\le N} \frac{1}{k} - \int_1^{N+1} \frac{1}{x}\,dx\right| \le H_1 - H_{N+1} = 1 - \frac{1}{N+1} = O(1).$$

Since $\int_1^{N+1} \frac{1}{x}dx = \ln(N+1) = \ln(N) + \ln(\frac{N+1}{N}) = \ln(N) + O(1)$, we conclude that $H_N = \ln(N) + O(1)$.

Example 2 Consider the sum of the first N cubes $\sum_{1 \le k \le N} k^3$. The Euler-Maclaurin formula implies that

$$\left| \sum_{1 \le k \le N} k^3 - \int_1^{N+1} x^3 \, dx \right| \le \int_1^{N+1} 3x^2 \, dx = x^3 \Big|_1^{N+1} = O(N^3)$$

and the original integral is estimated by

$$\int_1^{N+1} x^3 \, dx = \frac{x^4}{4} \Big|_1^{N+1} = \frac{(N+1)^4}{4} - \frac{1}{4} = \frac{N^4}{4} \left(1 + \frac{1}{N}\right)^4 + O(1) = \frac{N^4}{4} + O\left(N^3\right).$$

Combining this with the error term, we conclude that $\sum_{1 \le k \le N} k^3 = (1 + O(1/N)) \cdot N^4/4$.

3 Laplace Method

The Laplace method is a strategy that combines the two above principles of approximating by integrals and paying attention to which regions contribute most to a sum. Many terms have negligible effect on the asymptotics and don't need to be carefully estimated by the integral.

3.1 Method Outline

The main steps of the Laplace method are as follows:

Step 1. *Neglect tails of original sum/restrict ourselves to the range that contains the main contribution.*
For example, if $f(k)$ is negligible outside of I^*, then

$$\sum_{k \in I} f(k) = \sum_{k \in I^*} f(k) + \sum_{k \in I \setminus I^*} f(k) \approx \sum_{k \in I^*} f(k).$$

Step 2. *Approximate the main summands, probably by integral.*
Continuing the example, if we can approximate f by $\hat{f}$ inside I^*, then

$$\sum_{k \in I^*} f(k) \approx \sum_{k \in I^*} \hat{f}(k) \approx \int_{I^*} \hat{f}(x) \, dx.$$

Step 3. *Extend the range to include the tails.*
Continuing the example, if the integral of $\hat{f}(x)$ over $J \setminus I^*$ is negligible, then

$$\int_{I^*} \hat{f}(x) \, dx = \int_J \hat{f}(x) \, dx - \int_{J \setminus I^*} \hat{f}(k) \, dx \approx \int_J \hat{f}(x) \, dx.$$

To sum up, we conclude that

$$\sum_{k \in I} f(k) \approx \int_J \hat{f}(x) dx,$$

where the intergral may have some nice form that we can analyze more explicitly (e.g., Gaussian integral).

Remark: When using the Laplace method, instead of just approximating a sum by integral, we first restrict the range, approximate well there, and then extend the range in a safe way. The main advantage of this method comes from the approximation of f by $\hat{f}$, which may not be possible over the entire interval I.

3.2 Example: Ramanujan Q-function

Let's look at a concrete example. We will analyze the asymptotics of the Ramanujan Q-function

$$Q(N) = \sum_{1 \le k \le N} \binom{N}{k} \frac{k!}{N^k},$$

which arises in analysis of hashing algorithms, random maps, caching, etc.

Claim 3.1. $Q(N) = (1 + o(1))\sqrt{\pi N/2}$.

Proof. In order to estimate the summands, we define

$$Q_{N,k} := \binom{N}{k}\frac{k!}{N^k} = \frac{N!}{(N-k)!N^k} = \frac{N(N-1)\cdots(N-k+1)}{N\cdot N\cdots N} = \prod_{1\le i<k}\left(1-\frac{i}{N}\right).$$

Aiming at an *upper bound (or tail bound)*, using $1 - x \le e^{-x}$ we obtain that

$$Q_{N,k} \leqslant e^{-\sum_{1\le i<k}\frac{i}{N}} = e^{-\binom{k}{2}/N}$$

for all $k \ge 1$. Aiming at an *approximation*, using $1 - x = e^{-x+O(x^2)}$ we also obtain that

$$Q_{N,k} = e^{-\sum_{1\le i<k}\frac{i}{N}+O\left(\sum_{1\le i<k}\frac{i^2}{N^2}\right)} = e^{-\binom{k}{2}/N+O(k^3/N^2)}. \tag{3.1}$$

Note that if $k = o(N^{2/3})$, then the error term in the exponent is $O(k^3/N^2) = o(1)$. Recalling $e^{o(1)} = 1+o(1)$, this shows that $e^{-\binom{k}{2}/N}$ is a very good approximation of $Q_{N,k}$ when $k = o(N^{2/3})$.
Let $k_0 = N^{1/2+\varepsilon}$ for $\varepsilon > 0$ such that $k_0 = o(N^{2/3})$. We are now positioned to use the Laplace method.

Step 1. *Neglect the tails.* Consider the contribution of the tail, $k \ge k_0 = N^{1/2+\varepsilon}$. We get that

$$\sum_{k_0\le k\le N} Q_{N,k} \le \sum_{k_0\le k\le N} e^{-\binom{k}{2}/N} \le N\cdot e^{-\Omega(k_0^2/N)} \le N\cdot e^{-\Omega(N^{2\varepsilon})} = o(1).$$

Asymptotically, this expression is exponentially small and can be safely ignored.
Step 2. *Approximate the contributing region by integral.* Using the approximation in equation (3.1) and approximating $\binom{k}{2}/N = \frac{k^2}{2N} + O(k/N)$, we see that

$$\sum_{1\le k<k_0} Q_{N,k} = \sum_{1\le k<k_0} e^{-\frac{k^2}{2N}+O(k/N+k^3/N^2)} = (1+o(1))\sum_{1\le k<k_0} e^{-\frac{k^2}{2N}},$$

where the last equality holds because $k = o(N^{2/3})$. This expression is now nicely integrable. Indeed, using monotonicity, we obtain the approximation

$$\sum_{1\le k<k_0} e^{-k^2/2N} = \int_1^{k_0} e^{-x^2/2N}\,dx + O(1).$$

Step 3. *Extend the range of the integral.* We extend the integral

$$\int_1^{k_0} e^{-x^2/2N}\,dx = \int_0^{\infty} e^{-x^2/2N}\,dx - \int_0^{1} e^{-x^2/2N}\,dx - \int_{k_0}^{\infty} e^{-x^2/2N}\,dx = \int_0^{\infty} e^{-x^2/2N}\,dx + O(1).$$

This is a scaling of the well-known Gaussian integral, so

$$\int_0^{\infty} e^{-x^2/2N}\,dx = \sqrt{2N}\int_0^{\infty} e^{-z^2}\,dz = \sqrt{2N}\cdot\sqrt{\pi}/2 = \sqrt{N\pi/2}.$$

Finally, putting together our approximations, we conclude that

$$Q(N) = \sum_{1\le k\le N} Q_{N,k} = (1+o(1))\sqrt{\pi N/2},$$

and we remark that with some more care in the approximations (summing $O(k/N + k^3/N^2)\cdot e^{-k^2/(2N)}$ separately) one can obtain $Q(N) = \sqrt{\pi N/2} + O(1)$. $\square$

This proof relied on the fact that k was not too big in the region where we approximated by integral. The term $O(k/N + k^3/N^2)$ that we were able to neglect because k was small would have prevented us from approximating by the Gaussian integral over the entire range.

4 Final Remarks

- The key takeaway is that the power of the Laplace method lies in the ability to approximate a sum by a conveniently integrable function. A direct approximation by integral, over the entire region, may not have a straightforward form.

MATH 264A: Combinatorics	**Fall 2024**

Lecture 12: Asymptotic Methods: More on Laplace's Method

November 5, 2024

Lecturer: Lutz Warnke

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Today's topics:

- More on the Laplace Method
  - How to estimate integrals.
  - Examples (Integral + Sums)
- *Sources:*
  - Flajolet-Sedgewick (Analytic Combinatorics), Appendix B.6
  - Moore-Mertens (Nature of Computation), Section 14.4.2 + Appendix A.6.1

1 Review of Laplace Method

1. Identify the source of the 'main contribution'.
2. Use different estimates in different ranges.
3. Heuristically,
$$\underbrace{\sum_{k\in I} f(k) \sim \sum_{k\in I^*} f(k)}_{\text{Restrict interval}} \sim \underbrace{\sum_{k\in I^*} \hat{f}(k)}_{\substack{\text{Approximate function}\\ \text{on restriction}}} \sim \underbrace{\int_{I^*} \hat{f}(x)dx}_{\substack{\text{Approximate sum}\\ \text{by integral}}}.$$
Some of these steps can involve further approximation or ignoring tails.

1.1 Last Lecture:

$$f(k) = \binom{N}{k}\frac{k!}{N^k} \qquad\qquad I = \{1,\dots,N\}$$
$$\hat{f}(k) = e^{-k^2/(2N)} \qquad\qquad I^* = \{N^{1/2+\delta},\dots,N\}$$

In particular, after substitution $z = k/\sqrt{2N}$ could use Gaussian Integral:
$$\int_{N^{1/2+\delta}}^{N} e^{-k^2/2N}\,dk \approx \sqrt{2N}\underbrace{\int_0^\infty e^{-z^2}\,dz}_{=\sqrt{\pi}/2} = \sqrt{N\pi/2}.$$

1.2 Motivating Question (for today):

What if the function $\hat{f}(k)$ is more complicated?
(So that we can not/no longer directly use Gaussian Integration.)

2 Laplace Method

- It turns out that we do systemic variant of this for certain integrals that frequently arise in applications:

$$I_n = \int_a^b f(x)e^{nh(x)}dx,$$

where our main focus is on the asymptotic behavior of I_n as $n \to \infty$.

- Here, we assume f and h are are sufficiently smooth and independent of n. Also, a and b should be independent of n.
- For the proof, we want $-\infty < a < b < \infty$ but it is possible to extend these ideas to infinite intervals.
- Assuming that h has a unique maximum, we intuitive expect that, when n is large, this integral is dominated by the contribution of those x near the maximum, the heuristic idea being that the contribution of x far away (from the maximum) are exponentially smaller and thus negligible.
- The following theorem formalizes all of this.

Theorem 2.1 (Laplace Method Theorem, simplified)**.** *Let f, h be sufficiently smooth functions on $[a, b]$. Assume that h has a unique maximum at a point x_0 inside (a, b), so that $h'(x_0) = 0$. We will also assume that $h''(x_0) < 0$ and $f(x_0) \neq 0$. Then*

$$I_n := \int_a^b f(x)e^{nh(x)}dx \sim \underbrace{\sqrt{\frac{2\pi}{n|h''(x_0)|}}}_{=Polynomial\ Correction} \cdot \underbrace{f(x_0)e^{nh(x_0)}}_{=Max\ Value/Peak} \qquad \textit{as } n \to \infty.$$

Proof outline. At first, we will assume that f is identically 1. We will split the integral

$$I_n = \int_a^b e^{nh(x)}dx = \int_{x_0-\epsilon}^{x_0+\epsilon} e^{nh(x)}dx + \int_{|x-x_0|\geqslant\epsilon} e^{nh(x)}dx.$$

Step 1: Taylor Approximation
The idea is to do a Taylor approximation of h at x_0, noting that $h'(x_0) = 0$ and $h''(x_0) < 0$,

$$h(x) = h(x_0) - \frac{1}{2}|h''(x_0)|(x - x_0)^2 + O(|x - x_0|^3).$$

Using this approximation, we have

$$\int_{x_0-\epsilon}^{x_0+\epsilon} e^{nh(x)}dx \approx e^{nh(x_0)} \int_{x_0-\epsilon}^{x_0+\epsilon} e^{-|h''(x_0)|(x-x_0)^2/2}dx.$$

We now do a change of variables: $y = (x - x_0)\sqrt{n|h''(x_0)|/2}$, to obtain the integral

$$e^{nh(x_0)}\sqrt{\frac{2}{n|h''(x_0)|}} \int_{-\epsilon\sqrt{n|h''(x_0)|/2}}^{\epsilon\sqrt{n|h''(x_0)|/2}} e^{-y^2}dy.$$

For large n, this is well approximated by the Gaussian integral $\int_{\mathbb{R}} e^{-y^2}dy = \sqrt{\pi}$, so we overall get

$$\int_{x_0-\epsilon}^{x_0+\epsilon} e^{nh(x)}dx \approx e^{nh(x_0)}\sqrt{\frac{2\pi}{n|h''(x_0)|}}.$$

Step 2: Contributions away from x_0 are negligible
From the hypotheses, there is $\delta = \delta(\epsilon) > 0$ such that $h(x) \leqslant h(x_0) - \delta$ for all $x \in [a, x_0 - \epsilon] \cup [x_0 + \epsilon, a]$.

(This follows from our assumptions by compactness of the set $[a, x_0 - \epsilon] \cup [x_0 + \epsilon, b]$, our assumption on the second derivative, and also the uniqueness of the maximum of h).
Because of this, when x is outside of $[x_0 - \epsilon, x_0 + \epsilon]$, we get that

$$\int_{|x-x_0|\geqslant\epsilon} e^{nh(x)} dx \leqslant e^{n(h(x_0)-\delta)}|b-a| \ll e^{nh(x_0)}/\sqrt{n}.$$

Putting things together, as $n \to \infty$, we see that the 'main' contribution to the integral I_n indeed comes from the interval $[x_0 - \epsilon, x_0 + \epsilon]$, which gives the claimed asymptotics of I_n (see Step 1).
Finally, we can extend the outlined argument to general f by Taylor expanding f around x_0, i.e., using $f(x) = f(x_0) + O(|x - x_0|)$. □

3 Examples

3.1 Stirling's Formula (we skipped this in class)

We can apply this result to the integral

$$\Gamma(N+1) = \int_0^\infty e^{-x} x^N dx,$$

where $\Gamma(N+1) = N!$ when N is an integer. By a change of variables using $x = Nz$, we obtain the integral

$$N! = \int_0^\infty e^{-Nz}(Nz)^N N dz = N^{N+1} \int_0^\infty e^{N(\ln z - z)} dz.$$

By ignoring the tail (which is negligible, say, when $x > 2$), we can apply the previous result to get

$$N! \sim \sqrt{2\pi N} \left(\frac{N}{e}\right)^N.$$

3.2 Tiling Example (concrete application)

Our goal is to use the Laplace Method approach to show, as $n \to \infty$, that

$$a_n = \sum_{0\leqslant k\leqslant\lfloor n/2\rfloor} \binom{n-k}{k} \sim \frac{1}{\sqrt{5}} \left(\tfrac{1+\sqrt{5}}{2}\right)^{n+1}.$$

To approximate the sum by an integral, the plan is approximate the binomial coefficients using Stirling's Formula (for which we need $n - 2k \to \infty$ and $k \to \infty$ to get negligible error terms).
Step 1: Neglect the tails.
For $k_0 := \sqrt{n}$, using $\binom{n-k}{k} \leqslant (n-k)^k \leqslant n^k$ we see, with room to spare, that

$$\sum_{0\leqslant k\leqslant k_0} \binom{n-k}{k} \leqslant (k_0+1)n^{k_0} \leqslant e^{o(n)} \ll \left(\tfrac{1+\sqrt{5}}{2}\right)^n.$$

Similarly, using $\binom{n-k}{k} = \binom{n-k}{n-2k} \leqslant n^{n-2k}$ we also see, with room to spare, that

$$\sum_{\lfloor n/2\rfloor - k_0\leqslant k\leqslant\lfloor n/2\rfloor} \binom{n-k}{k} \leqslant (k_0+1)n^{k_0+1} \leqslant e^{o(n)} \ll \left(\tfrac{1+\sqrt{5}}{2}\right)^n.$$

Step 2: Approximate the main terms (using Stirling).
By Stirling's Formula, for $k_0 \leqslant k \leqslant \lfloor n/2\rfloor - k_0$ we obtain (since $k \to \infty$ and $n - 2k \to \infty$) that

$$\binom{n-k}{k} = \frac{(n-k)!}{(n-2k)!k!} \sim \frac{\sqrt{2\pi(n-k)}\left(\frac{n-k}{e}\right)^{n-k}}{\sqrt{2\pi(n-2k)}\left(\frac{n-2k}{e}\right)^{n-2k}\sqrt{2\pi k}\left(\frac{k}{e}\right)^k}$$

$$= \sqrt{\frac{1-k/n}{2\pi(1-2k/n)k}} \left[\left(\frac{1-k/n}{1-2k/n}\right)^{1-k/n} \left(\frac{1-2k/n}{k/n}\right)^{k/n} \right]^n .$$

Step 3: Write $k = cn$ and optimize for c.
For $k = cn$ the previous estimate reduces to

$$\binom{n-k}{k} \sim \underbrace{\sqrt{\frac{1-c}{2\pi(1-2c)cn}}}_{=f(c)/\sqrt{n}} \underbrace{\left[\left(\frac{1-c}{1-2c}\right)^{1-c} \left(\frac{1-2c}{c}\right)^{c} \right]^n}_{=e^{nh(c)}} = \frac{1}{\sqrt{n}} \cdot f(c) \cdot e^{nh(c)}.$$

By optimizing (using a computer algebra system; please verify this yourself!):

$$h(c) \text{ has a unique maximum at } c_0 = 1/2 - 1/(2\sqrt{5}),$$

which moreover satisfies

$$e^{nh(c_0)} = \left(\tfrac{1+\sqrt{5}}{2}\right)^n \qquad \text{and} \qquad h''(c_0) < 0.$$

Step 4: Neglect contributions far away from optimum c_0.
In our upcoming integral approximations we would like to work with upper and lower bounds that do not depend on n, so we shall again neglect tails. In particular, using the previous estimates we know that $h(c) \leqslant h(c_0) - \delta$ when $|c - c_0| \geqslant \varepsilon$ (similar as in the proof of the Laplace Method Theorem), and so

$$\sum_{k_0 \leqslant k \leqslant (c_0-\varepsilon)n} \underbrace{\binom{n-k}{k}}_{\sim \frac{1}{\sqrt{n}} \cdot f(k/n) \cdot e^{nh(k/n)}} + \sum_{(c_0+\varepsilon)n \leqslant k \leqslant \lfloor n/2 \rfloor - k_0} \underbrace{\binom{n-k}{k}}_{\sim \frac{1}{\sqrt{n}} \cdot f(k/n) \cdot e^{nh(k/n)}} \leqslant n^{O(1)} \cdot e^{n(h(c_0)-\delta)} \ll \left(\tfrac{1+\sqrt{5}}{2}\right)^n .$$

Step 5: Approximate the main contributions using Laplace Method Theorem.
Using the above Stirling's Formula approximation together with integral comparison, we see that

$$\begin{aligned} \sum_{(c_0-\varepsilon)n \leqslant k \leqslant (c_0+\varepsilon)n} \binom{n-k}{k} &\sim \sum_{(c_0-\varepsilon)n \leqslant k \leqslant (c_0+\varepsilon)n} \frac{1}{\sqrt{n}} \cdot f(k/n) \cdot e^{nh(k/n)} \\ &\approx \frac{1}{\sqrt{n}} \int_{(c_0-\varepsilon)n}^{(c_0+\varepsilon)n} f(k/n) \cdot e^{nh(k/n)} dk = \sqrt{n} \int_{c_0-\varepsilon}^{c_0+\varepsilon} f(x) \cdot e^{nh(x)} dx, \end{aligned}$$

where we used the substitution $x = k/n$ and thus $dk = ndx$ in the last step (for time-reasons we did not make the negligible integral approximation error precise). Using the Laplace Method Theorem (which applies with $a = c_0 - \varepsilon$ and $b = c_0 + \varepsilon$ by the properties of f and h) to approximate the integral, it follows that

$$\sqrt{n} \int_{c_0-\varepsilon}^{c_0+\varepsilon} f(x) \cdot e^{nh(x)} dx \sim \sqrt{n} \cdot \sqrt{\frac{2\pi}{n|h''(c_0)|}} f(c_0) e^{nh(c_0)} \tag{3.1}$$

$$= \sqrt{\frac{1-c_0}{(1-2c_0)c_0|h''(c_0)|}} \cdot \left(\tfrac{1+\sqrt{5}}{2}\right)^n = \frac{1}{\sqrt{5}} \left(\tfrac{1+\sqrt{5}}{2}\right)^{n+1} , \tag{3.2}$$

where we omitted verifying the last equality (which can be done with the help of a computer algebra system). Finally, putting things together, we overall obtain the claimed asymptotic estimate

$$a_n = \sum_{0 \leqslant k \leqslant \lfloor n/2 \rfloor} \binom{n-k}{k} \sim \frac{1}{\sqrt{5}} \left(\tfrac{1+\sqrt{5}}{2}\right)^{n+1} .$$

Remark: we presented the argument in a way that mimics the discovery process, rather than trying to write it in the most compact way (for example, if one exploits that Stirling's Formula is always correct up to $O(1)$ factors, then using the Step 4 argument one sees that here the first 'neglecting tails' Step 1 can be avoided).

MATH 264A: Combinatorics	**Fall 2024**

Lecture 13: More on Asymptotic Methods: Bootstrapping

November 7, 2024

Lecturer: Lutz Warnke

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Today's topics:

- Bootstrapping - Using asymptotic info we currently have to obtain improved estimates/asymptotics.
- Multiple examples, generally of the form:
  - Inverting Implicit Equations: $z = F(z)$, solve for z
  - (Approximate) Recursions: $p_{n+1} = F(p_1, \cdots, p_n)$, estimate p_n

1 Core idea of Bootstrapping

- We start with a weak/rough estimate obtained through some simple reasoning.
- By repeatedly plugging it into the implicit equation or recursive formula, we can often iteratively obtain a stronger estimate (in a completely rigorous way).

2 Examples

2.1 Example 1 - Solution to a Polynomial Equation

Solve for z where $z = \frac{1}{2}(1 + z^{k+1})$, for each k as $k \to \infty$

- **Input (Rough Information)**: $z \in [0, 1 - \epsilon]$

We bootstrap this information to improve our estimate of z by repeatedly applying (2.1).

$$z = \frac{1}{2}(1 + z^{k+1}) \tag{2.1}$$

Apply (2.1):

$$z = \frac{1}{2}(1 + z^{k+1}) = \begin{cases} \leq \frac{1}{2}(1 + (1-\epsilon)^{k+1}) \\ \geq \frac{1}{2}. \end{cases}$$

So we can say that z is within the range $\frac{1}{2}(1 \pm (1-\epsilon)^{k+1})$.
Apply (2.1) again:

$$\begin{aligned} z &= \frac{1}{2}\Big[1 + \Big(\frac{1}{2}\Big)^{k+1}\big(1 \pm (1-\epsilon)^{k+1}\big)^{k+1}\Big] \\ &= \frac{1}{2}\Big[1 + \Big(\frac{1}{2}\Big)^{k+1}\Big(1 + O\big(k(1-\epsilon)^k\big)\Big)\Big] \\ &= \frac{1}{2} + \Big(\frac{1}{2}\Big)^{k+2} + O\Big(k\Big(\frac{1}{2}(1-\epsilon)\Big)^k\Big) \end{aligned}$$

And we observe that $O\big(k\big(\frac{1}{2}(1-\epsilon)\big)^k\big) \ll (\frac{1}{2})^k$ for large k.

Hence, if we apply (2.1) again, then we can obtain an even better bound:

$$z = \cdots = \frac{1}{2} + \left(\frac{1}{2}\right)^{k+2} + O\big(k4^{-k}\big).$$

In general, more iterations require more work, but we can get more precise bounds on z.

Remark: It is important that this argument is fully rigorous, since each line of argument uses/gives an (asymptotic) upper or lower bound on the value, not just an approximation.

2.2 Example 2 - Inversion of Binomial Coefficients

Solve for n as a function of z: $z = \binom{n}{n/2}$ as $n \to \infty$

- **Side information we use for the partition**:
  Stirling: $\binom{n}{n/2} = \frac{n!}{((n/2)!)^2} = (1+o(1)) \cdot \frac{c2^n}{\sqrt{n}}$, where $c = \sqrt{\frac{2}{\pi}}$ is a constant.

We can rewrite this as:

$$\log_2 z = \log_2(1+o(1)) + \log_2 c + n - \frac{1}{2}\log_2 n \tag{2.2}$$

Thus $\log_2 z \sim n$ as $n \to \infty$, from which we can obtain:

$$\log_2 n = \log_2\big((1+o(1))\log_2 z\big) = \log_2\big(1+o(1)\big) + \log_2\log_2 z = \log_2\log_2 z + o(1)$$

Re-inserting this into (2.2) gives:

$$\begin{aligned} n &= \log_2 z + \frac{1}{2}\log_2 n - \log_2 c + o(1) \\ &= \log_2 z + \frac{1}{2}\log_2\log_2 z - \log_2 c + o(1) \end{aligned}$$

Remark: We could get better estimates, if we had started with a better initial Stirling's approximation.

2.3 Example 3 - Solution to implicit Branching Process Equation

Suppose $\rho > 0$ behaves according to the following expression (where $\varepsilon > 0$):

$$1 - \rho = e^{-(1+\epsilon)\rho}$$

What happens to ρ as $\epsilon \to 0$?

- **Input (Side Information)**: We can argue that $\rho \in (0,1)$, which is the information we will start with.

We do a Taylor's expansion:

$$(1+\epsilon)\rho = -\log(1-\rho) = \sum_{k\geq 1}\frac{\rho^k}{k} = \rho + \frac{\rho^2}{2} + \frac{\rho^3}{3} + \frac{\rho^4}{4} + \cdots .$$

As $\rho \neq 0$, we can divide both sides by ρ and subtract 1 to obtain:

$$\epsilon = \frac{\rho}{2} + \frac{\rho^2}{3} + \frac{\rho^3}{4} + \cdots . \tag{2.3}$$

But we need to find ρ in terms of ϵ! To achieve this we shall apply our main trick:

- **Step 1:** Observe that $\epsilon \geq \frac{\rho}{2}$ (as $\rho > 0$).
  This gives us the rough estimate
  $$0 < \rho \leq 2\epsilon,$$
  which crucially implies $\rho \to 0$ as $\epsilon \to 0$.

- **Step 2:** we can upgrade the asymptotics, by inserting the rough estimate into (2.3):

$$\epsilon = \frac{\rho}{2} + O(\rho^2) = \rho\left(\frac{1}{2} + O(\epsilon)\right)$$

Which implies:

$$\rho = \frac{2\epsilon}{1+O(\epsilon)} = 2\epsilon(1+O(\epsilon)) \qquad \text{as } \frac{1}{1+x} = 1 + O(x).$$

- **Step 3:** we can further upgrade the estimate by reapplying (2.3):

$$\epsilon = \frac{\rho}{2} + \frac{\rho^2}{3} + O(\rho^3) = \rho\left(\frac{1}{2} + \frac{1}{3}(2\epsilon + O(\epsilon)) + O(\epsilon^2)\right)$$

which implies:

$$\rho = \frac{2\epsilon}{1+\frac{4}{3}\epsilon(1+O(\epsilon))} = \cdots = 2\epsilon - \frac{8}{3}\epsilon^2 + O(\epsilon^3) \qquad \text{as } \frac{1}{1+x} = 1 - x + O(x^2).$$

And we can repeat this over and over again, to get the higher order terms/series expansion of ρ.

2.4 Example 4 - n-th Prime Number $p = p_n$

We can rewrite this as $n := \pi(p)$, where $\pi(p)$ is the number of primes at most p.

- **Fact (Side Information):** $\pi(p) = \frac{p}{\log p} + O(\frac{p}{(\log p)^2})$ as $p \to \infty$.

Noting that $p \geq n$, we obtain an approximate recursion:

$$p = n\log p\left(1 + O\left(\frac{1}{\log p}\right)\right) = n\log p\left(1 + O\left(\frac{1}{\log n}\right)\right) \tag{2.4}$$

To bootstrap, we first obtain rough/crude information: by taking log on both sides we obtain

$$\log p = \log n + \log\log p + O\left(\frac{1}{\log n}\right),$$

which implies $\log p \sim \log n$ as $p \geq n \to \infty$, from which we conclude that

$$\log p = \log n + \log((1+o(1))\log n) + o(1) = \log n + \log\log n + o(1).$$

Re-inserting into (2.4) for refinement, we obtain:

$$p_n = p = n\log n + n\log\log n + O(n).$$

Remark: We could obtain a better estimate for p_n, if we had started with a better estimate for $\pi(p)$.

3 Example 5 - Asymptotics of a Quadratic Recursion

Suppose $a_0 = \frac{1}{2}$ and $a_n = a_{n-1}(1 - a_{n-1})$ for $n \geq 1$.

- **Heuristic observation:** In $a_n = a_{n-1} - a_{n-1}^2$, the main contribution to the final growth rate should come from the linear term a_{n-1} rather than the quadratic term $-a_{n-1}^2$.

With a simple inductive proof, we can obtain a crude bound:

$$\left(\frac{1}{2}\right)^{n+1} \leq a_n \leq \frac{1}{2}.$$

Inversion trick: Take inverse and apply partial fractions:

$$\frac{1}{a_n} = \frac{1}{a_{n-1}(1-a_{n-1})} = 1 + \frac{1}{a_{n-1}} + \frac{a_{n-1}}{1-a_{n-1}}. \tag{3.1}$$

Together with the crude estimate, this gives

$$\frac{1}{a_n} \geq 1 + \frac{1}{a_{n-1}} \geq \cdots \geq n + \frac{1}{a_0} = n+2.$$

We can use this lower bound to bootstrap an upper bound (by applying (3.1)):

$$\frac{1}{a_n} = 1 + \frac{1}{a_{n-1}} + \frac{1}{\frac{1}{a_{n-1}} - 1}.$$

Indeed, we can repeatedly apply $\frac{1}{a_{n-1}} - 1 \geq n$ on this estimate to obtain

$$\frac{1}{a_n} \leq n + \left(\frac{1}{n} + \frac{1}{n-1} + \cdots + 1\right) + \frac{1}{a_0} = n + \log n + O(1),$$

since the harmonic series satisfies $H_n := \sum_{1 \leqslant i \leqslant n} 1/i = \log n + O(1)$. To sum up, we have

$$\frac{1}{a_n} = n + O(\log n), \tag{3.2}$$

which in particular implies that

$$a_n = \frac{1}{n + O(\log n)} = \frac{1}{n} \cdot \frac{1}{1 + O(\frac{\log n}{n})} = \frac{1}{n}\left(1 + O\left(\frac{\log n}{n}\right)\right)$$

Remark: We can obtain better estimate by iterating this bootstrapping argument with (3.2), to obtain

$$\frac{1}{a_n} = n + \log n + o(\log n).$$

4 Example 6 - Asymptotics of implicitly defined Lambert Function

The Lambert Function $W(z)$ is the unique positive real satisfying

$$W(z)e^{W(z)} = z. \tag{4.1}$$

With a simple argument by contradiction[1], one can obtain the following rough/crude estimate:

$$1 \leq W(z) \leq \ln z \qquad \text{for } z \geq e. \tag{4.2}$$

To bootstrap, we rewrite (4.1) and re-insert the crude estimate (4.2): we obtain as $z \to \infty$,

$$W(z) = \ln\left(\frac{z}{W(z)}\right) = \ln z - \ln W(z) = \ln z - O(\ln\ln z). \tag{4.3}$$

In fact, more careful bootstrapping (using $1 - x = e^{-(1+o(1))x}$ as $x \to 0$) yields, as $z \to \infty$,

$$\begin{aligned} W(z) &= \ln z - \ln W(z) = \ln z - \ln\Big(\ln z - \ln W(z)\Big) \\ &= \ln z - \ln\ln z - \ln\left(1 - \frac{\ln W(z)}{\ln z}\right) = \ln z - \ln\ln z + \frac{O(\ln\ln z)}{\ln z}. \end{aligned}$$

5 Final Remarks

We used multiple examples to illustrate that bootstrapping is a powerful method/technique, which can be used to improve simple/trivial bounds to obtain more precise asymptotic bounds/estimates.

[1] If $W(z) < 1$, then $W(z) \cdot e^{W(z)} < 1 \cdot e^1 \leq z$, which contradicts the definition of $W(z)$. Similarly, if $W(z) > \ln z$ then $e^{W(z)} > \ln z \cdot e^{\ln z} \geq \ln e \cdot z \geq z$, which contradicts the definition of $W(z)$.

MATH 264A: Combinatorics **Fall 2024**

Lecture 14: Switching Method I

November 12, 2024

Lecturer: Lutz Warnke

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Today's topics:

- Switching Method (Combinatorial Double Counting)
- Counting Examples: exact/upper bound/asymptotic bounds

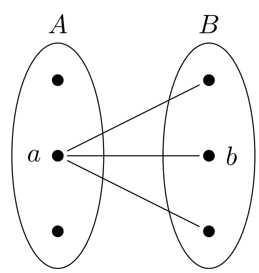
1 Switching Method

Goal: Compare the size of two sets $|A|$ and $|B|$ via local changes.

Idea:

(i) Define map $\phi : A \to 2^B$ called the "Switching Operation"

(ii) Consider the bipartite auxiliary graph, where ab is an edge iff $b \in \phi(a)$:



(iii) Double-Count edges of the auxiliary bipartite graph:

$$\sum_{a\in A} \underbrace{\deg a}_{=|\phi(a)|} = \sum_{b\in B} \underbrace{\deg b}_{=|\phi^{-1}(b)|},$$

where $|\phi(a)|$ is the "*Forward Switching Degree*," and $|\phi^{-1}(b)|$ is the "*Reverse Switching Degree*."

Kind of Results:

(i) Exact results

(ii) Bounds:

$$\left(\min_{a\in A} \deg a\right) \cdot |A| \leqslant \left(\max_{b\in B} \deg b\right) \cdot |B| \qquad \Longrightarrow \qquad \frac{|A|}{|B|} \leqslant \frac{\max_{b\in B} \deg b}{\min_{a\in A} \deg a}$$

(iii) Asymptotic Result:

$$\begin{cases} \deg a \approx x \text{ for all } a \in A, \\ \deg b \approx y \text{ for all } b \in B \end{cases} \qquad \Longrightarrow \qquad x|A| \approx |B|y \qquad \Longrightarrow \qquad \frac{|A|}{|B|} \approx \frac{y}{x}$$

1.1 Exact Results via Switching Method: Permutations

$\mathcal{S}_n :=$ all permutations π of $[n]$.
$\mathcal{S}_n^* :=$ all $\pi \in \mathcal{S}_n$ where $\pi(i) > \pi(j) > \pi(k)$, and i, j, k are fixed integers.

Want:

$$\mathbb{P}(\text{random permutation } \pi \in \mathcal{S}_n \text{ is in } \mathcal{S}_n^*) = \frac{|\mathcal{S}_n^*|}{|\mathcal{S}_n|}$$

Switching Operation: $\phi : \mathcal{S}_n^* \to 2^{\mathcal{S}_n}$

$$\begin{matrix}\pi(i)\\ \pi(j)\\ \pi(k)\end{matrix} \xrightarrow[\substack{\text{orderings of}\\ \pi(i),\pi(j),\pi(k)}]{\text{all possible}} \begin{matrix}\pi(i)\\ \pi(j)\\ \pi(k)\end{matrix}$$

and everything else stays unchanged (as usual for switching operations, which are 'local')

Forward Switching: Number of switchings that can be applied to $\pi \in \mathcal{S}_n^*$:

$$\deg \pi = |\phi(\pi)| = 3!$$

Reverse Switching: Number of ways $\pi' \in \mathcal{S}_n$ can be obtained by switching some $\pi \in \mathcal{S}_n^*$:

$$\deg \pi' = |\phi^{-1}(\pi)| = 1$$

Double-Counting:

$$3! \cdot |\mathcal{S}_n^*| = \sum_{\pi \in \mathcal{S}_n^*} \deg \pi = \sum_{\pi' \in \mathcal{S}_n} \deg \pi' = |\mathcal{S}_n| \quad \Rightarrow \quad \frac{|\mathcal{S}_n^*|}{|\mathcal{S}_n|} = \frac{1}{3!} \quad \Rightarrow \quad |\mathcal{S}_n^*| = \frac{n!}{6}$$

Remark: While one could solve this problem with direct counting, this indirect argument is more robust.

1.2 Bounds via Switching Method: Perfect Matchings (PM)

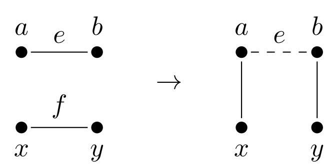
$\mathcal{M} :=$ all PM of G
$\mathcal{M}_e :=$ all PM of G that contain $e \in E(G)$

Want: lower bound on

$$\mathbb{P}(\text{random PM of } G \text{ contains } e) = \frac{|\mathcal{M}_e|}{|\mathcal{M}|} \qquad \text{when} \quad \underbrace{\min_v \deg_G(v)}_{=\delta(G)} \geqslant \left(\frac{1}{2} + \epsilon\right) n,$$

where n denotes the number of vertices (as usual)

Switching Operation: $\phi : \mathcal{M}_e \to 2^{\mathcal{M} \setminus \mathcal{M}_e}$



Remove two edges $e, f \in M \in \mathcal{M}_e$, and add ax, by as edges to obtain $M' \in \mathcal{M} \setminus \mathcal{M}_e$.
(Note that this operation does not modify the degree sequence. Furthermore, if e and f were edges in a perfect matching, their transformation still maintains a perfect matching.)

Reverse Switching: Number of ways $M' \in \mathcal{M} \setminus \mathcal{M}_e$ can be obtained by switching some $M \in \mathcal{M}_e$:

$$\begin{aligned} \deg M' &= \text{number of choices for } xy \\ &\leqslant 1 \quad (x \text{ and } y \text{ are uniquely determined by } a, b \text{ and } M') \end{aligned}$$

Forward Switching: Number of ways switching can be applied to $M \in \mathcal{M}\backslash\mathcal{M}_e$ with $a = eb$:

- Each neighbor $x \in \Gamma(a) \setminus \{b\}$ of a has a unique 'matching' partner y such that $xy \in M$. Denoting the set of all such y's by Y_a, note that the size of Y_a is at least $|Y_a| \geqslant |\Gamma(a)| - 1$. To ensure that any of the discussed $xy \in M$ is a valid choice, we require that y is also a neighbor of b, i.e., that $y \in \Gamma(b)$.
- Since that each vertex has degree at least $\delta(G) \geqslant \left(\frac{1}{2} + \epsilon\right) n$, it follows that the intersection of Y_a and $\Gamma(b)$ has size at least

$$|Y_a \cap \Gamma(b)| = |Y_a| + |\Gamma(b)| - |Y_a \cup \Gamma(b)| \geqslant (|\Gamma(a)| - 1) + |\Gamma(b)| - n \geqslant 2\varepsilon n - 1.$$

(It's not important, but by noting $b \notin Y_a \cup \Gamma(b)$ we could readily remove the -1 above.)

- Therefore, the number of choices for xy in the switching operation is at least

$$\deg M = \text{number of choices for } xy \geqslant 2\epsilon n - 1.$$

An alternative approach for calculating the forward degree $\deg M$ is as follows:

- Note that there are $2 \cdot (|M| - 1) = 2(n/2 - 1) = n - 2$ possible *ordered* pairs xy of vertices with $\{x, y\} \cap \{a, b\} = \emptyset$ such that $xy \in M$ is a matching edge. For any such pair xy, there is a switching available if (i) a and x are adjacent, and (ii) b and y are adjacent (here the order of x and y matters). The number of vertices $x \notin \Gamma(a)$ which are not adjacent to a is at most $n - \Gamma(a) \leqslant n - \delta(G) \leqslant (1/2 - \epsilon)n$. Similarly, the number of vertices $y \notin \Gamma(b)$ which are not adjacent to b is at most $n - \Gamma(b) \leqslant n - \delta(G) \leqslant (1/2 - \epsilon)n$. Hence the total number of pairs xy for which a switching is available is at least

$$2 \cdot (|M| - 1) - 2 \cdot (n - \delta(G)) \geqslant (n - 2) - 2 \cdot (1/2 - \epsilon)n \geqslant 2\varepsilon n - 2.$$

(It's not important, but by noting that $x \notin \Gamma(a) \cup a$ and $y \notin \Gamma(b) \cup b$ must hold, we could readily improve each $n - \delta(G)$ to $n - 1 - \delta(G)$ and thus remove the final -2 above.)

Double-Counting:

$$(2\epsilon n - 1) \cdot |\mathcal{M}_e| \leqslant |\mathcal{M}\backslash\mathcal{M}_e| \qquad \Longrightarrow \qquad \frac{|\mathcal{M}_e|}{|\mathcal{M}|} \leqslant \frac{|\mathcal{M}_e|}{|\mathcal{M}\backslash\mathcal{M}_e|} \leqslant \frac{1}{2\epsilon n - 1}$$

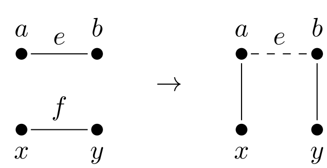
1.3 Asymptotic Estimate via Switching Method: d-regular Graphs

$\mathcal{G}_d :=$ all n-vertex graphs G_d which are d-regular (no multiple edges, and $2|dn$)
$\mathcal{G}_d^e :=$ all $G_d \in \mathcal{G}_d$ that contain edge e

Want:[1] (n denotes the number of vertices, as usual)

$$\mathbb{P}(\text{random } G_d \text{ in } \mathcal{G}_d \text{ contains } e) = \frac{|\mathcal{G}_d^e|}{|\mathcal{G}_d|} \qquad \text{for} \qquad 1 \ll d \ll n$$

Switching Operation: $\phi : \mathcal{G}_d \to 2^{\mathcal{G}_d \backslash \mathcal{G}_d^e}$



Replace edges $ab, xy \in E(G_d)$ with edges ax, by, such that no multiple edges arise in the resulting graph $G_d' \in \mathcal{G}_d^e$. (Hence a, b, x, y need to be distinct, and ax, by cannot be edges of G_d)
For convenience, in the following switching arguments we will consider ordered edges.

[1] As some of you pointed out, using symmetry one can more directly estimate the probability that an edge appears in a random d-regular graph. As we shall see in the homework problems, the point is that the switching method based approach easily extends to situations where such simple symmetry considerations no longer apply, e.g., to estimate the probability that a set of edges appears in a random d-regular graph – indicating the power and versatility of the switching method.

Forward Switching: Let $G_d \in \mathcal{G}_d$. First, $\deg G_d \leqslant dn$. Furthermore

$$\begin{aligned}\deg G_d = {} & \text{ number of choices of } f \\ & = (\text{\# all ordered edges}) - (\text{\# invalid edge choices for } f) \\ & \geqslant dn - \underbrace{(d+1)}_{\substack{\text{exclude} \\ \{a\}\cup\Gamma(a) \text{ for } x}} d - \underbrace{(d+1)}_{\substack{\text{exclude} \\ \{b\}\cup\Gamma(b) \text{ for } y}} d\end{aligned}$$

Therefore, we can write

$$\deg G_d = dn + O(d^2) = dn\,(1 + O\,(d/n)) = dn\,(1 + o\,(1))$$

Reverse Switching: Let $G'_d \in \mathcal{G}_d \backslash \mathcal{G}^e_d$. We have

$$\begin{aligned}\deg G'_d = {} & \text{ number of choices of } f \\ & \leqslant d^2 \qquad (\text{as } x \in \Gamma(a) \text{ and } y \in \Gamma(b))\end{aligned}$$

Note that using our 2-edge switching operation we can not get a lower bound better than $\deg G'_d \geqslant 0$, since we have no control over the number of choices of $x \in \Gamma(a)$ and $y \in \Gamma(b)$ that are neighbors.

Double-Counting:

$$(dn - 2(d+1)d) \cdot |\mathcal{G}^e_d| \leqslant |\mathcal{G}_d \setminus \mathcal{G}^e_d| \cdot d^2 \qquad \Longrightarrow \qquad \frac{|\mathcal{G}^e_d|}{|\mathcal{G}_d|} \leqslant \frac{|\mathcal{G}^e_d|}{|\mathcal{G}_d \setminus \mathcal{G}^e_d|} \leqslant \frac{d^2}{dn - 2(d+1)d} = \frac{d}{n - 2(d+1)}.$$

Next time we will discuss a more involved 3-edge switching operation, with which we can obtain an asymptotic bound (and not just an upper bound).

2 Final Remarks

- For some examples that we discuss in the switching method part of class, there might alternative ways to obtain the same result. Conceptually the main point is that the switching method is more robust and versatile, since it does not require an estimate an the total of objects, but only works with ratios (for example, until recently good estimates on the total number of d-regular graphs were unavailable).
- The main freedom of the method is the choice of the switching operation, which can be adapted to the problem at hand (and as we saw in the d-regular graphs example, for asymptotic results sometimes the 'natural' choice of the switching operation is not enough, but a more careful choice is needed); furthermore, experience shows that more involved switching operations can often lead to more precise estimates.

MATH 264A: Combinatorics **Fall 2024**

Lecture 15: Switching Method II (draft)

November 14, 2024

Lecturer: Lutz Warnke

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Today's topics:

- Applications: d-regular graphs (probability)
- General Switching Framework, and a variant.

1 Application in Random Graph Theory

We will examine the following problem: Given a random d-regular graph on n vertices, the probability that an edge $e = (u, w)$ appears in the graph. We use the following notations: $\mathcal{G}_{n,d}$ denotes the set of all n-vertex d-regular graphs, and $G_{n,d}$ denotes a graph chosen uniformly at random from $\mathcal{G}_{n,d}$.

Theorem 1.1. *For $1 \ll d \ll n$, $\mathbb{P}(e \in E(G_{n,d})) = (1 + o(1))\frac{d}{n}$*

Let $\mathcal{C}_1 := \{G \in \mathcal{G}_{n,d} | e \in E(G)\}$ and $\mathcal{C}_0 := \{G \in \mathcal{G}_{n,d} | e \notin E(G)\}$. Then clearly

$$\mathbb{P}(e \in E(G_{n,d})) = \frac{|\mathcal{C}_1|}{|\mathcal{G}_{n,d}|} = \frac{|\mathcal{C}_1|}{|\mathcal{C}_1| + |\mathcal{C}_0|} = \frac{1}{1 + \frac{|\mathcal{C}_0|}{|\mathcal{C}_1|}}.$$

So to prove Theorem 1.1 it suffices to show $\frac{|\mathcal{C}_1|}{|\mathcal{C}_0|} = (1 + o(1))\frac{d}{n}$.

In the last lecture, we had used '2-edge switching', wherein for distinct vertices u, w, x, y, and edges $e = (u, w), (x, y) \in E$, we produced a new graph with these edges being replaced by the edges $(u, x), (w, y)$.

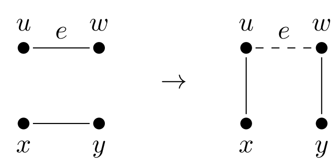


Figure 1: Illustration of 2-edge switching

However, as seen in previous lecture, we were unable to bound 'bad' cases in backward switching due to which this method only gives $\frac{|C_1|}{|C_0|} \leq (1 + o(1))\frac{d}{n}$.

Thus, we will modify this method to '3-edge switching'. In forward switching, given distinct vertices x_1, y_1, x_2, y_2, u, w and edges $e = (u, w), (x_1, y_1), (x_2, y_2)$, the transformation ϕ applied to $G_{n,d}$ will replace these with the edges $(x_1, w), (y_1, x_2), (u, y_2)$. Similarly, in backward switching, edges of the latter type will be transformed (applying ϕ^{-1}) to the former edges.

Additionally, we have the following caveats:

i. We only consider those combinations of 3 edges in $G_{n,d}$ where the 'new' edges (the edges formed after forward transformation) are not already present in $G_{n,d}$.

ii. All 6 vertices considered are distinct from each other.

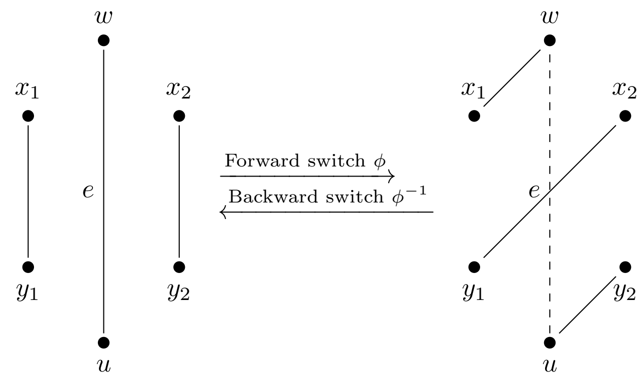


Figure 2: Illustration of 3-edge switching

For forward switching, we need to calculate the number of ways to apply the transformation ϕ to $G \in \mathcal{C}_1$. Let us define $\deg(G)$ as the number of choices for (x_1, y_1) and (x_2, y_2).

- For (x_1, y_1), note that the vertex x_1 can be any vertex which is not u, w or a neighbor of w. So there are $n - 2 - d = n - \mathcal{O}(d)$ choices for x_1 (recall that there are exactly d neighbors of w). Then y_1 needs to be a neighbor of x_1, which is distinct from u. So there are $d - \mathcal{O}(1)$ choices for y_1. Thus the total choices for (x_1, y_1) are

$$(n - \mathcal{O}(d))(d - \mathcal{O}(1)) \ = nd - \mathcal{O}(d^2) - \mathcal{O}(n) \ = nd\left(1 - \mathcal{O}\left(\frac{d}{n}\right) - \mathcal{O}\left(\frac{1}{d}\right)\right)$$

- We can argue using similar method that there are $nd\left(1 - \mathcal{O}(\frac{d}{n}) - \mathcal{O}(\frac{1}{d})\right)$ choices for (x_2, y_2). (Note that there will be some more restrictions since we have fixed x_1 and y_1 before, i.e. x_2, y_2 should be distinct from x_1, y_1; x_2 should not be a neighbor of y_1 but that will not change this asymptotic expression.)

Hence

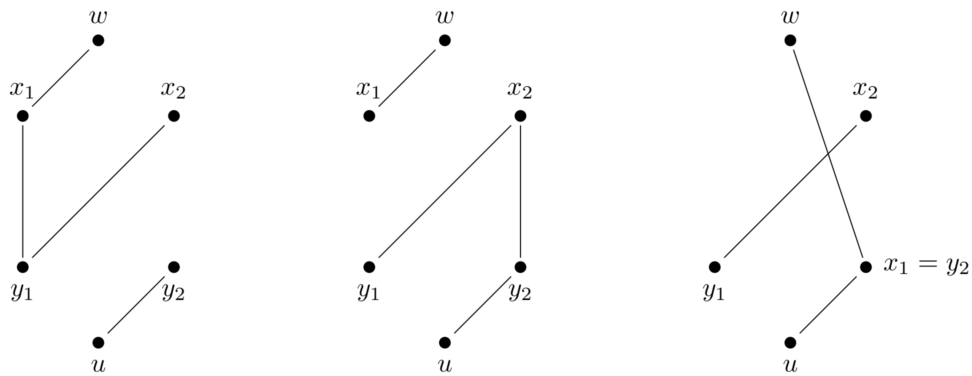
$$\deg(G) = \left(nd\left(1 - \mathcal{O}\left(\frac{d}{n}\right) - \mathcal{O}\left(\frac{1}{d}\right)\right)\right)^2 \ = \left(nd\left(1 - o(1)\right)\right)^2 \ = (nd)^2(1 + o(1))$$

where the second last inequality is obtained using the premise $1 \ll d \ll n$.

Now for the backward switching, we need to calculate the number of ways $G' \in \mathcal{C}_0$ arises by switching some $G \in \mathcal{C}_1$. Again $\deg(G')$ is the number of choices for (x_1, y_1) and (x_2, y_2).

We have the trivial upper bound for $\deg(G')$ as $\deg(G') \leq d^2 \cdot nd$ (d^2 choices for x_1, y_2 since they are neighbors of w, u respectively; n choices for x_2, and d choices for y_1 as neighbor of x_2). From this we can subtract the no. of 'bad configurations', i.e. configurations which violate the caveats we have placed. It can be seen that those configurations will be corresponding to the cases (illustrated below):

$(x_1, y_1) \in E(G')$, $(x_2, y_2) \in E(G')$, $x_1 = y_2$



For the first two cases, note that there is a path of length 3 from w (resp. u) and an edge from u (resp. w). So there are $\mathcal{O}(d^3 \cdot d) = \mathcal{O}(d^4)$ configurations for each of these cases. For the third case, $x_1 = y_2$ is a common neighbor of w and u (having d choices) and there is the edge (x_2, y_1) (having nd choices). So there are $\mathcal{O}(d \cdot nd) = \mathcal{O}(nd^2)$ configurations for this case.

Thus

$$\deg(G') = nd^3 - \mathcal{O}(d^4) - \mathcal{O}(nd^2) \ = nd^3\left(1 - \mathcal{O}\left(\frac{d}{n}\right) - \mathcal{O}\left(\frac{1}{d}\right)\right) \ = nd^3(1 + o(1))$$

An alternative approach to calculate $\deg(G')$ is as follows:

- For x_1, y_2, there are $d^2 - d$ choices (d each for x_1, y_2 being neighbors of w, u respectively, and subtracting d for maintaining $x_1 \neq y_2$). So there are $d^2 - \mathcal{O}(d) \ = d^2\left(1 - \mathcal{O}\left(\frac{1}{d}\right)\right)$ valid pairs of x_1, y_2.
- For (x_2, y_1), we can select x_2 out of n vertices and y_1 out of its d neighbors. However we need to be careful so that neither of the edges (x_1, y_1) and (x_2, y_2) are present, i.e. y_1 is not a neighbor of x_1 and x_2 is not a neighbor of y_2. There are at most d^2 such possible 'bad' (x_2, y_1) pairs. So there are $nd - \mathcal{O}(d^2) \ = nd\left(1 - \mathcal{O}\left(\frac{d}{n}\right)\right)$ valid pairs of (x_2, y_1).

Again,

$$\deg(G') = d^2\left(1 - \mathcal{O}\left(\frac{1}{d}\right)\right) \cdot nd\left(1 - \mathcal{O}\left(\frac{d}{n}\right)\right) \ = nd^3(1 - o(1)).$$

Finally the double counting argument gives

$$|\mathcal{C}_1|(n^2d^2(1 + o(1))) = |\mathcal{C}_0|(nd^3(1 + o(1))) \qquad \Longrightarrow \qquad \frac{|\mathcal{C}_1|}{|\mathcal{C}_0|} = (1 + o(1))\frac{d}{n},$$

and this proves Theorem 1.1.

2 General Switching Framework

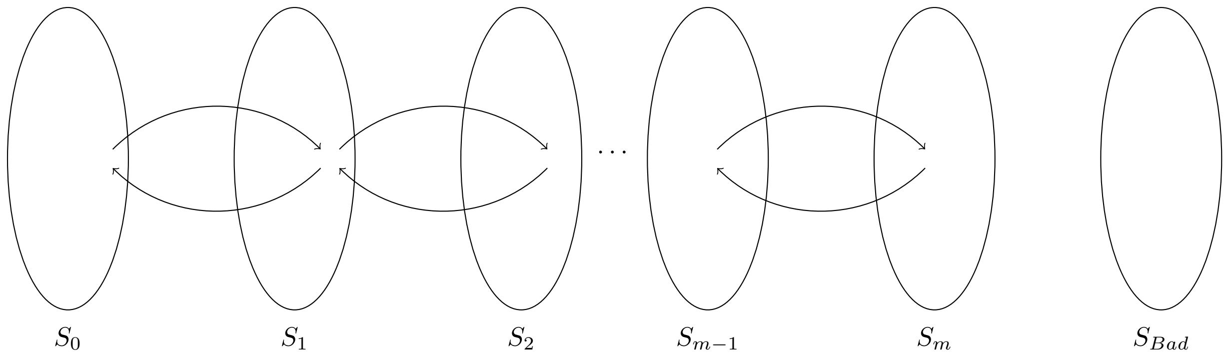
Given the set S of all possible elements of interest (eg. graphs, permutations, etc) we do the following:

i. Partition S into disjoint unions:

$$S = S_0 \sqcup S_1 \sqcup \ldots \sqcup S_m \sqcup S_{Bad}$$

ii. Show that the 'bad' set S_{Bad} consisting of undesirable elements is negligible: $\frac{|S_{Bad}|}{|S|} = o(1)$

iii. Know the relative ratios $\frac{|S_{i+1}|}{|S_i|}$ via switching.



Then to obtain estimate on $\frac{|S_0|}{|S|}$ we have the following:

$$\frac{|S|(1-o(1))}{|S_0|} \overset{\text{ii}}{=} \frac{|S \setminus S_{Bad}|}{|S_0|} \overset{\text{i}}{=} \sum_{j=0}^{m} \frac{|S_j|}{|S_0|} = \sum_{j=0}^{m} \prod_{i=0}^{j-1} \frac{|S_{i+1}|}{|S_i|}$$

In many applications we then aim at the switching estimate

$$\frac{|S_{i+1}|}{|S_i|} = \frac{\lambda}{i+1}\big(1 + o(1/m)\big)$$

for a suitable parameter $\lambda > 0$, since this readily implies (assuming $m \to \infty$) that

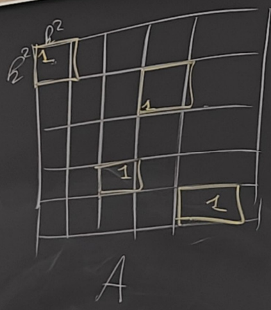
$$\sum_{j=0}^{m} \prod_{i=0}^{j-1} \frac{|S_{i+1}|}{|S_i|} = \underbrace{\sum_{j=0}^{m} \frac{\lambda^i}{i!}}_{\to e^{\lambda} \text{ as } m \to \infty} \underbrace{\big(1 + o(1/m)\big)^j}_{=1+o(1)} \to e^{\lambda},$$

which in turn gives

$$\frac{|S_0|}{|S|} \to e^{-\lambda},$$

which asymptotically equals the probability that a Poisson random variable with mean λ is zero.

6(a)



Replace each Block with at least one 1 by 1 (0 else) $\longrightarrow$

<u>Claim</u> A avoids perm matrix P $\Longrightarrow$ $\widehat{A}$ avoids P.

<u>Proof</u>. Assume $\widehat{A}$ contains P.

Consider the corresp. 1's in P

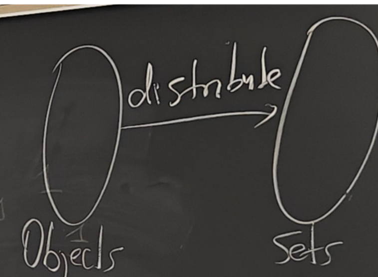
$\rightsquigarrow$ det. Blocks in A with at least one 1.

(diff Blocks on disj. sets of rows/col. $\Leftarrow$ P perm Matrix)

$\rightsquigarrow$ can select one 1 from each such Block to find P. $\square$

(b) #Blocks in A with at least one 1 $=$ #1's in $\widetilde{A} \overset{(a)}{\leq} ex\left(\frac{n}{k^2}, P\right)$

(C) Pigeonhole-Principle (Averaging)

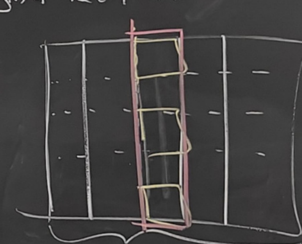


If $|O| > r|S|$ many objects are distributed into $|S|$ sets, then one set contains $\geq \frac{|O|}{|S|} > r$ objects

Proof: If every set has $< \frac{|O|}{|S|}$, the total # objects is $< |S| \cdot \frac{|O|}{|S|} = |O|$ $\lightning$

wide block B: 1's in $\geq k$ columns

Claim $\underbrace{\#\text{ wide Blocks}}_{=M} \leq \underbrace{k\binom{k^2}{k}}_{=N} \cdot \frac{n}{k^2}$

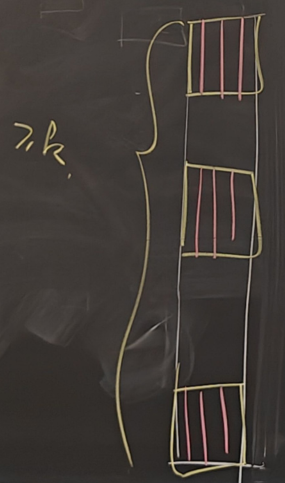


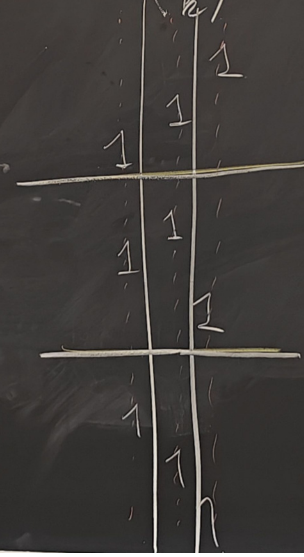
Proof: Assume $M > N$.

PHP $\Longrightarrow$ $\exists$ Colum-Bock s.t.

$$\underbrace{\begin{array}{c}\#\text{ wide Blocks}\\ \text{in that Col-Block}\end{array}}_{=\tilde{M}} \geq \frac{M}{n/k^2} > \frac{N}{n/k^2} = k\binom{k^2}{k}$$

PHP $\Longrightarrow$ in this column-Block,
#wide Blocks that agree on first k cols $\geq \dfrac{\tilde{M}}{\binom{k^2}{k}} > k$.





$\leftarrow$ can go from Top to bottom to find any perm matr P

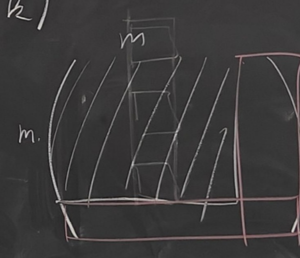
(d) $k^2 \mid n \Rightarrow ex(n,P) \leq 2 \cdot k^4 \cdot k\binom{k^2}{k}\frac{n}{k^2} + (k-1)^2 ex\left(\frac{n}{k^2},P\right)$

Ind: $ex(n,P) \leq Cn$ $\quad C := 2k^4\binom{k^2}{k}$

BC ✓

IS $\quad m =$ largest multiple of k^2 that is $\leq n$.

$ex(n,P) \leq ex(m,P) +$ max # 1's



$\boxed{4}$ can choose $\varepsilon = C \frac{\log\log\Delta}{\log\Delta}$ in Joh-Molloy-Result.

→ Need to check all "asymp." estimates where fixed ε ($\& \Delta \to \infty$) used

(i) $1 \leq \ell = \Delta^{\varepsilon/2} = e^{\frac{\varepsilon}{2}\log\Delta}$

(ii) $\varepsilon \geq \frac{t \log\Delta}{\varepsilon^4(1+\varepsilon)\Delta^{\varepsilon/2}} \overset{t=\Delta^{\frac{\varepsilon}{4}}}{=} \frac{\log\Delta}{\varepsilon^2(1+\varepsilon)\Delta^{\varepsilon/4}} \iff 1 \overset{!}{\geq} \frac{\overbrace{\log\Delta}^{e^{\log\log\Delta}}}{\underbrace{\varepsilon^3}_{=e^{-3\log(\frac{1}{\varepsilon})}}\underbrace{(1+\varepsilon)}_{=\Theta(1)}\underbrace{\Delta^{\varepsilon/4}}_{=e^{\frac{\varepsilon}{4}\log\Delta}}}$

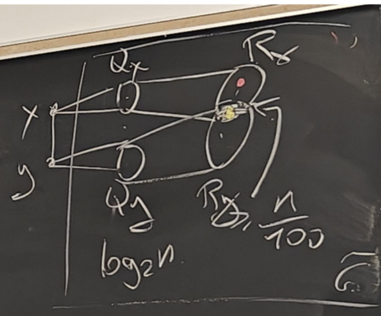
$= \frac{1}{\Theta(1)} \cdot e^{\log\log\Delta + 3\log(\frac{1}{\varepsilon}) - \boxed{\frac{\varepsilon}{4}\log\Delta}}$

$\varepsilon \geq \frac{\tilde{C}\left(\log\boxed{\log\Delta} + \log\boxed{\frac{1}{\varepsilon}}\right)}{\log\Delta}$

Can be made to work for $\varepsilon = \frac{C\log\log\Delta}{\log\Delta}$ as the $\log(\frac{1}{\varepsilon}) \overset{!}{\underset{\Delta\to\infty}{\sim}} \log\log\Delta$

More estimates

$\boxed{3}$ $|D(n)| \leq \underbrace{\binom{n}{2}}_{x,y} \cdot \underbrace{|\mathcal{T}(n-2)|}_{\Delta\text{-graph } \tilde{G} \text{ on rem. vx's}} \cdot$ #ways to connect x,y to Rest.



#choices $Q_x, Q_y \longrightarrow$ Det. R_x, R_y (by $\tilde{G}$)

$$\leq \binom{n}{\log_2 n}\binom{n}{\log_2 n} \leq n^{2\log_2 n}$$

#ways to connect x,y to Rest. $[n] \setminus (\{x,y\} \cup \underbrace{Q_x \cup Q_y}_{=O(1)})$

$=$ #ways to connect $[n] \setminus (\{x,y\} \cup Q_x \cup Q_y)$ to x,y.

$$\leq 1^{|R_x \cap R_y|} \cdot 2^{|R_x \setminus R_y|} \cdot 2^{|R_y \setminus R_x|} \cdot 3^{n - |R_x \cup R_y|}$$

$|R_x \cup R_y| = |R_x| + |R_y| - |R_x \cap R_y|$.

Observations

- no edges $x \longrightarrow R_x$
- no edges $y \longrightarrow R_y$
- no vx connected to both x, y

$$\frac{|D(n)|}{|\mathcal{T}(n-2)|} \leq n^{2+2\log_2 n} \, 2^{(|R_x| + |R_y| - 2|R_x \cap R_y|)} \cdot 2^{\boxed{\log_2 3}\,(n - |R_x| - |R_y| + |R_x \cap R_y|)}$$

$\lambda \approx 1.58\ldots \in (1,2)$

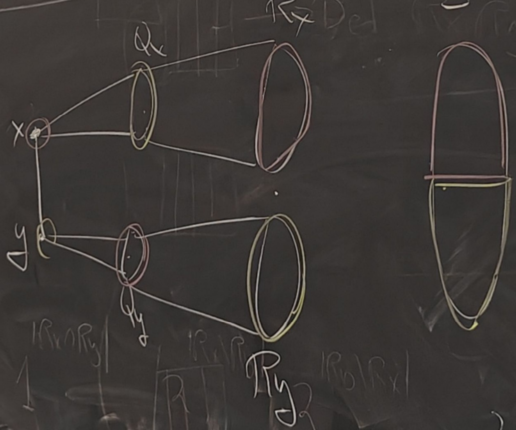
$$\leq \underbrace{n^{4\log_2 n}}_{\leq 2^{o(n)}} 2^{n} \cdot 2^{-\left(\overbrace{(2-\lambda)}^{>0}\underbrace{|R_x \cap R_y|}_{\geq \frac{n}{100}} + \overbrace{(\lambda-1)}^{>0}\underbrace{\boxed{|R_x| + |R_y|}}_{\geq 2\left(\frac{1}{2} - \frac{1}{1000}\right)n}\right)}$$

$$\leq 2^{n\left(1 - \frac{1}{2000}\right)} \quad \text{for } n \geq n_0.$$

(C) $\underbrace{T(n) \setminus [A(n) \cup B(n) \cup D(n)]}_{\mathcal{G}} \subseteq Col_2(n)$

Deferred Claims: $\boxed{G \text{ has no } C_5, C_7, C_9}$

Fix edge xy



$S = \text{Rest}$ (Remain. Vx's)

$S_x = \{v \in S : R_v \cap R_x = \emptyset\}$

$S_y = \{v \in S : R_v \cap R_y = \emptyset\}$

Properties

(i) $R_x \cap R_y = \emptyset$ as oth. find C_5.

(ii) $S = S_x \cup S_y$: note: $R_v \cap (R_x \cup R_y) \neq \emptyset$

follows from $|R_x \cup R_y| \overset{(i)}{=} |R_x| + |R_y| \geq 2 \cdot \left(\frac{1}{2} - \frac{1}{1000}\right) n = \left(1 - \frac{1}{500}\right) n$

$\Rightarrow$ fewer than $\frac{n}{500}$ vx's not in $R_x \cup R_y$

but $|R_v| \geq \left(\frac{1}{2} - \frac{1}{1000}\right) n > \frac{n}{500}$

(iii) $S_x \cap S_y = \emptyset$

(iv) $R_x \cup S_x$ & $R_y \cup S_y$ are ISETs

$\Rightarrow$ get valid 2-col.

MATH 264A: Combinatorics **Fall 2024**

Lecture 17: Switching Method III: Counting d-regular graphs (draft)

November 21, 2024

Lecturer: Lutz Warnke

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Today's topics

- Advanced Switching Example: Counting d-regular graphs.
- *Source:*
  Chapter 11 in 'Introduction to Random Graphs' by A. Frieze & M. Karonski.

1 Number of d-regular n-vertex graphs

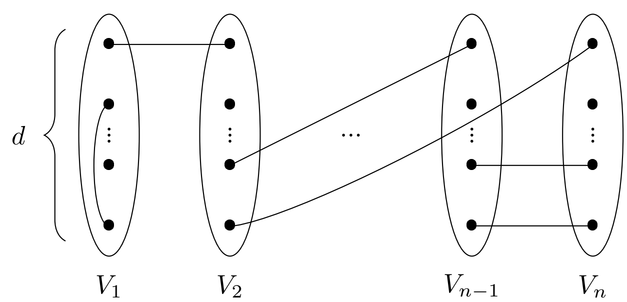
We define:

$$\mathcal{G}_{n,d} := \{\text{all simple } d\text{-regular graphs with vertex set } [n]\},$$
$$\mathcal{G}^*_{n,d} := \{\text{all } d\text{-regular graphs with vertex set } [n] \text{ where we allow loops/multiple edges}\}.$$

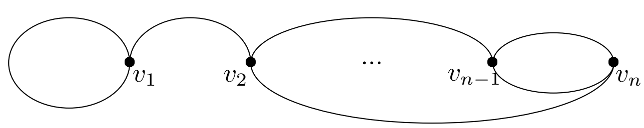
Note that a loop contributes 2 to the degree count. For technical reasons, we will look at the so-called *Pairing/Configuration Model* (more general than $\mathcal{G}^*_{n,d}$) which allows for loops and multiple edges:

$$\mathcal{P} = \mathcal{P}(n,d) := \text{All perfect matchings of } [dn].$$

We call a perfect matching of $[dn]$ a "pairing".



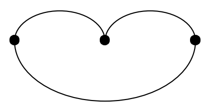
Idea: We create a grid of $d \times n$ points by replacing each vertex v_i with a "bin" V_i containing d points. Given a perfect matching on these points, we then define a projection[1] $\psi : \mathcal{P} \to \mathcal{G}^*_{n,d}$ which collapses the points within each bin V_i to vertex v_i (to obtain a multigraph). Note that edges between vertices in the same group now become loops, and any set of edges between vertices in V_i and V_j for $i \neq j$ become multiple edges.

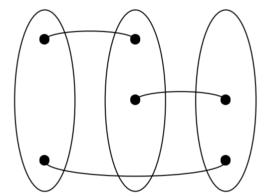


[1] The reason for working with perfect matchings is that they are easier to work with than $\mathcal{G}^*_{n,d}$.

Key point:

(a) *Each $G \in \mathcal{G}_{n,d} \subsetneq \mathcal{G}^*_{n,d}$ corresponds to $(d!)^n$ pairings.*
To see this, we need to understand how many pairings give some graph (without loops and multiple edges) after projection map: any permutation of the d points in each 'bin' gives the same graph.





Since there are n bins, this gives

$$|\psi^{-1}(\mathcal{G}_{n,d})| = |\mathcal{P}_0| = (d!)^n \cdot |\mathcal{G}_{n,d}|.$$

(b) *Cardinality of $\mathcal{P}$:*
We can generate all pairings by iterating the following: pick a point that has not yet been paired (say, the first in some lexicographic order), and then pair it with a distinct point that has not yet been paired ($dn-1$ choices). Iterating this until all points are paired up, we overall arrive at

$$|\mathcal{P}| = (dn-1)\cdot(dn-3)\cdots 1 = (dn-1)!!,$$

where we are implicitly assuming $2|dn$. Hence

$$|\mathcal{P}| = \frac{(dn)!}{(dn)(dn-2)\cdots} = \frac{(dn)!}{(\frac{dn}{2})!\cdot 2^{dn/2}}.$$

Combining (a) and (b), we have

$$|\mathcal{G}_{n,d}| \overset{(a)}{=} \frac{|\mathcal{P}_0|}{(d!)^n} = \frac{|\mathcal{P}|}{(d!)^n}\cdot\frac{|\mathcal{P}_0|}{|\mathcal{P}|} \overset{(b)}{=} \frac{(dn)!}{(\frac{dn}{2})!\cdot(d!)^n\cdot 2^{dn/2}}\cdot\frac{|\mathcal{P}_0|}{|\mathcal{P}|}.$$

Thus to estimate $|\mathcal{G}_{n,d}|$, it remains to estimate $|\mathcal{P}_0|/|\mathcal{P}|$ by the switching method.

Theorem 1.1. *There is some $\alpha > 0$ such that, for $d = o(n^\alpha)$, we have*

$$\frac{|\mathcal{P}_0|}{|\mathcal{P}|} = (1+o(1))e^{-(d^2-1)/4}.$$

Corollary 1.2. *Using Stirling's formula, we infer that the number of d-regular n-vertex graphs is*

$$|\mathcal{G}_{n,d}| = (1+o(1))\sqrt{2}e^{-(d^2-1)/4}\left(\frac{(dne^{-1})^{d/2}}{d!}\right)^n.$$

1.1 Decompositions of pairings $\mathcal{P} = \mathcal{P}(n,d)$

Define $\mathcal{P}_{i,j}$ to be the set of all pairings $p \in \mathcal{P}$ where projection $\psi(P)$ has i loops, j double edges, and also satisfies a number of technical conditions, including: no double loops, no triple edges, no vertex incident with two double edges, no vertex incident with loop and double edges, no edge joining two distinct loops (we omit some other technical conditions here to avoid clutter).

Note: $\mathcal{P}_0 = \mathcal{P}_{0,0} = \{$all $p \in \mathcal{P}$ without loops and double edges$\}$

Lemma 1.3 (Probabilistic Lemma: proof omitted here)**.** *These sets give almost decomposition, i.e.,*

$$|\mathcal{P}| = (1+o(1))\sum_{\substack{0\le i\le d^2\log n\\ 0\le j\le d^2\log n}}|\mathcal{P}_{i,j}|.$$

(For a proof, see Chapter 11 in 'Introduction to Random Graphs' by A. Frieze & M. Karonski.)

1.2 Switchings

Lemma 1.4 (Switching Lemma)**.** *For all $0 \le i, j, \le 3d^2 \log n$:*

(i) $\frac{|\mathcal{P}_{i+2,j-1}|}{|\mathcal{P}_{i,j}|} = \frac{j}{(i+2)(j+1)}$

(ii) $\frac{|\mathcal{P}_{i-1,0}|}{|\mathcal{P}_{i,0}|} = \frac{2i}{d-1}\Big(1 + O\Big(\frac{d^2 \log n}{n}\Big)\Big)$.

We shall prove (i) of Lemma 1.4 after proving the following corollary (and omit the proof of (ii) for time-reasons), which in turn proves our main counting result Theorem 1.1 for d-regular n-vertex graphs.

Corollary 1.5. *As desired, this implies $|\mathcal{P}_0|/|\mathcal{P}| = (1 + o(1))e^{-(d^2-1)/4}$.*

Proof. First we remove double edges by increasing the number of loops:

$$|\mathcal{P}_{i,j}| \overset{(i)}{=} \frac{(i+2)(i+1)}{j} \cdot |\mathcal{P}_{i+2,j-1}| \overset{(i)}{=} \cdots \overset{(i)}{=} \frac{(i+2j)!}{i!j!} \cdot |\mathcal{P}_{i+2j,0}|.$$

Next we remove loops:

$$\begin{aligned}|\mathcal{P}_{i+2j,0}| &\overset{(ii)}{=} \frac{d-1}{2(i+2j)}\Big(1 + O\Big(\frac{d^2 \log n}{n}\Big)\Big) \cdot |\mathcal{P}_{i+2j-1,0}| \\ &\overset{(ii)}{=} \cdots \overset{(ii)}{=} \Big(\frac{d-1}{2}\Big)^{i+2j} \frac{1}{(i+2j)!} \cdot |\mathcal{P}_{0,0}| \cdot \underbrace{\Big(1 + O\Big(\frac{d^2 \log n}{n}\Big)\Big)^{i+2j}}_{=1+O\left(\frac{d^4(\log n)^2}{n}\right)=1+o(1)},\end{aligned}$$

where we used that by assumption $i + 2j = O(d^2 \log n)$ and $d = o(n^\alpha)$ for small enough $\alpha > 0$. Hence

$$\frac{|\mathcal{P}_{i,j}|}{|\mathcal{P}_{0,0}|} = (1 + o(1))\lambda^{i+2j} \frac{1}{i!} \frac{1}{j!} = (1 + o(1)) \frac{\lambda^i}{i!} \frac{(\lambda^2)^j}{j!} \qquad \text{with } \lambda := (d-1)/2.$$

Now the above yields

$$\frac{|\mathcal{P}|(1 + o(1))}{|\mathcal{P}_{0,0}|} \overset{\text{Prob. Lem.}}{=} \sum_{\substack{0 \le i \le d^2 \log n \\ 0 \le j \le d^2 \log n}} \frac{|\mathcal{P}_{i,j}|}{|\mathcal{P}_{0,0}|} = (1 + o(1)) \sum_{\substack{0 \le i \le d^2 \log n \\ 0 \le j \le d^2 \log n}} \frac{\lambda^i}{i!} \frac{(\lambda^2)^j}{j!} = (1 + o(1))e^{\lambda}e^{\lambda^2},$$

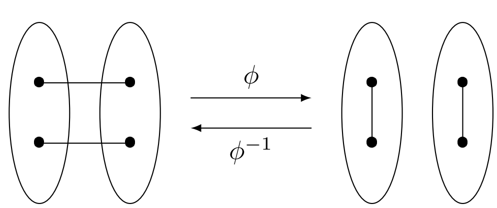
from which we infer that

$$\frac{|\mathcal{P}_0|}{|\mathcal{P}|} = \frac{|\mathcal{P}_{0,0}|}{|\mathcal{P}|} = (1 + o(1))e^{-\lambda(\lambda+1)} = (1 + o(1))e^{-(d^2-1)/4}.$$

□

1.3 Proof of Switching Lemma:

We now prove the 'double-edge removal' part (i) of Lemma 1.4:
Switching Operation: $\phi : \mathcal{P}_{i,j} \to 2^{\mathcal{P}_{i+2,j-1}}$



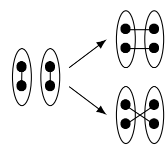
Caution: Need to carefully check that no technical conditions are violated (e.g., no double loops: as in $P_{i,j}$ no vertex incident to loops and double edges)

Forward switching ϕ: # of ways switching can be applied to $P \in \mathcal{P}_{i,j}$

$$\deg P = \# \text{ of multiple edges of } P \text{ (as there are no triple edges)} = j$$

Reverse switching ϕ^{-1}: # of ways $P' \in \mathcal{P}_{i+2,j-1}$ can be obtained from some $P \in \mathcal{P}_{i,j}$

$$\deg P' = \binom{i+2}{2} \cdot 2 = (i+2)(i+1) \text{ because there are two loops and we can resolve them in two ways:}$$



Double-Counting:

$$j \cdot |\mathcal{P}_{i,j}| = |\mathcal{P}_{i+2,j+1}| \cdot (i+2)(i+1) \qquad \Longrightarrow \qquad \frac{|\mathcal{P}_{i+2,j+1}|}{|\mathcal{P}_{i,j}|} = \frac{j}{(i+2)(i+1)}.$$

$\boxed{4}$ Random $\pi \in S_n$. $\mathbb{P}(\pi \text{ has no fixed points}) \longrightarrow \underbrace{e^{-1}}_{\mathbb{P}(\mathrm{Po}(1)=0)}$ as $n \to \infty$

$N_i = \# \pi \in S_n$ with exactly i fixed points [$\pi(x) = x$]

(a) to show:

$$\boxed{\frac{|N_{i+1}|}{|N_i|} = \frac{(1+o(L^{-1}))}{i+1} \quad \text{for all } 0 \le i \le L := \log\log\log n}$$

(b)–(d) $N_{>L} = \bigcup_{L < i \le n} N_i$ $|N_i| \le \underbrace{\binom{n}{i}}_{\text{choice of } i \text{ fixed points}} \cdot \underbrace{(n-i)!}_{\text{Perm. of rest}} = \frac{n!}{i!}$

(might create add. fixed points!)

Course Estimate:

$$\frac{|N_{>L}|}{|S_n|} = \sum_{L<i\le n} \underbrace{\frac{|N_i|}{n!}}_{\le \frac{1}{i!}} \le \sum_{i > L} \frac{1}{i!} \longrightarrow 0 \quad \text{as tail of } e = \sum_{i \ge 0} \frac{1}{i!}$$

as $n \to \infty$ ($\Rightarrow L \to \infty$)

$(\Rightarrow |N_{>L}| = o(|S_n|))$

CE

$$\frac{(1-o(1))|S_n|}{|N_0|} = \frac{|S_n| - |N_{>L}|}{|N_0|} = \sum_{0 \le i \le L} \frac{|N_i|}{|N_0|} = 1 + \sum_{1 \le i \le L} \prod_{0 \le j \le i-1} \underbrace{\boxed{\frac{|N_{j+1}|}{|N_j|}}}_{\overset{(a)}{=} \frac{1+o(L^{-1})}{j+1}}$$

Telescope

e.g. $\frac{|N_2|}{|N_0|} = \frac{|N_2|}{|N_1|} \cdot \frac{|N_1|}{|N_0|}$

$$= \frac{(1+o(L^{-1}))^i}{i!} = \frac{1+o(1)}{i!}$$

as $\underbrace{i}_{\le L} \cdot o(L^{-1}) = o(1)$

$$= (1+o(1)) \cdot \sum_{0 \le i \le L} \frac{1}{i!} \xrightarrow[\text{as } n \to \infty]{L \to \infty} e$$

$\Rightarrow \mathbb{P}(\pi \text{ no fixed point}) = \frac{|N_0|}{|S_n|} \longrightarrow e^{-1}$.

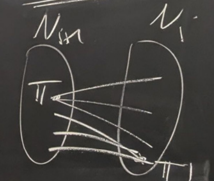
Double-Counting: $(i+1)\left(n^2-O((i+1)n)\right)\cdot|N_{i+1}| = |N_i|\cdot\left(n^2-O((i+1)n)\right)$

$\frac{1}{1-x} = 1+x+O(x^2)$

$0 \le \frac{i+1}{n} \le \frac{L+1}{n} \ll \frac{1}{L}$

$$\Rightarrow \frac{|N_{i+1}|}{|N_i|} = \frac{n^2-O((i+1)n)}{\left(n^2-O((i+1)n)\right)(i+1)} = \frac{1-O\left(\frac{i+1}{n}\right)}{\left(1-O\left(\frac{i+1}{n}\right)\right)\cdot(i+1)} = \frac{1+o(L^{-1})}{i+1}$$

$$\boxed{\frac{|N_{i+1}|}{|N_i|} = \frac{(1+o(L^{-1}))}{i+1} \quad \text{for all } 0\le i\le L,\ L = \log\log\log n}$$



Switching $\phi: N_{i+1}\to 2^{N_i}$

fixed point $\pi(a)=a$. $\longrightarrow$ $\pi(a):=B$.

$b\ne a$. $B:=\pi(b)\ne a,b$ $\qquad$ $\pi(b)=C$ $\longleftarrow$ exactly one fixed point less.

$c\ne a,b$. $C:=\pi(c)\ne b,c$. $\qquad$ $\pi(c)=a$.

Forward Switching $\longrightarrow$

switchings that can be applied to $\pi\in N_{i+1}$

$$\deg\pi = \underbrace{(i+1)}_{\text{choice of } a}\ \underbrace{\left(n-(i+1)-O(1)\right)^2}_{\text{choice of } b\ \&\ c} = (i+1)\left(n^2-O((i+1)n)\right)$$

Backward Switching $\longleftarrow$

given $\pi'\in N_i$, in how many ways can it be obtained (by switching some $\pi\in N_{i+1}$)

$$\deg\pi' = \underbrace{(n-i)}_{\substack{\text{choose } a:\ \pi'(a)=B \\ \text{(non fixed point in } \pi')}}\ \underbrace{(n-i-O(1))}_{\substack{\text{choice of non-fixed point} \\ b\ne a,\ c:\ \pi'(c)=C}} = n^2-O((i+1)n)$$

$\mapsto$ uniquely det. $c\ne a$ s.t. $a=\pi'(c)$

14 $C_2(n)$ = # 2-col. graphs

(a) Upper Bound (Simple)

$$C_2(n) \le \underbrace{\sum_{0\le k\le n}\binom{n}{k}}_{=2^n} 2^{\overset{\text{max at } k=n/2,\ \le \frac{n^2}{4}}{k(n-k)}} \le 2^{\frac{n^2}{4}+n}.$$

k | n-k

$$2^n = \sum_{0\le k\le n}\binom{n}{k} \le \binom{n}{\lfloor \frac{n}{2}\rfloor} \le (n+1)\binom{n}{\lfloor \frac{n}{2}\rfloor}$$

(b) Derive Lower Bound:

Idea: only look at graphs with color classes of size $\lfloor \frac{n}{2}\rfloor, \lceil \frac{n}{2}\rceil$.

$$C_2(n) \ge \frac{1}{2}\underbrace{\binom{n}{\lfloor \frac{n}{2}\rfloor}}_{\frac{n!}{\lfloor \frac{n}{2}\rfloor!\,\lceil \frac{n}{2}\rceil!} \overset{\text{Stirling}}{=} \frac{2^n}{\Theta(\sqrt{n})}} 2^{\lfloor \frac{n}{2}\rfloor\lceil \frac{n}{2}\rceil = \frac{n^2}{4}+O(1)} - \underbrace{\text{Overcounting that are not connected}}$$

$$\le \sum_{1\le k\le \lfloor \frac{n}{2}\rfloor}\underbrace{\binom{n}{k}}_{\le n^k}\underbrace{C_2(k)\cdot C_2(n-k)}$$

$$\le 2^{\frac{k^2}{4}+k+\frac{(n-k)^2}{4}+(n-k)}$$

$$= 2^{\frac{n^2}{4}+n-\frac{k(n-k)}{4}}$$

$$\le 2^{\frac{n^2}{4}+n}\underbrace{\sum_{1\le k\le \lfloor \frac{n}{2}\rfloor}\Big(\underbrace{n\cdot 2^{-\frac{\overset{\ge n/2}{(n-k)}}{4}}}_{n\cdot 2^{-n/8}}\Big)^k}_{\le n\cdot n\cdot 2^{-n/8}}$$

k   n-k

no edge

$$\ge 2^{\frac{n^2}{4}+n}\Big[\underbrace{\frac{1}{\Theta(\sqrt{n})} - O(n^2\cdot 2^{-n/8})}_{=\Theta(\frac{1}{\sqrt{n}})}\Big]$$

(C) Improved upper bound

$$C_2(n) \le \sum_{0\le k\le n} \binom{n}{k} 2^{k(n-k)}$$

$\Delta = \sqrt{D \log n}$

$$= \sum_{\substack{k,\\ |k-\frac n2|<\Delta}} \boxed{\binom{n}{k}} 2^{k(n-k)} + \sum_{\substack{k,\\ |k-\frac n2|\ge\Delta}} \binom{n}{k} 2^{k(n-k)}$$

$\binom{n}{k} \le (1+o(1))\sqrt{\frac{2}{\pi n}}\, 2^n$

$k = \frac n2 - j \longrightarrow (\frac n2 - j)(\frac n2 + j) = \frac{n^2}{4} - j^2$

Second sum: $\binom{n}{k} \le 2^n$, $2^{k(n-k)} \le 2^{\frac{n^2}{4} - D\log n} \ll \frac1n 2^{\frac{n^2}{4}}$

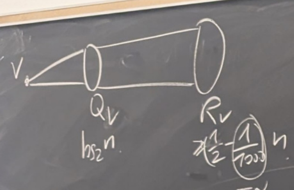
$\ll \frac{2^{n+n^2/4}}{n} < \underline{C_2(n)}$ ← can ignore!

$$= \Theta\left(\tfrac{1}{\sqrt n}\right) 2^n \sum_{0\le j\le\Delta} 2^{\frac{n^2}{4}-j^2} = \Theta\left(\tfrac{1}{\sqrt n}\right) 2^{\frac{n^2}{4}+n} \sum_{0\le j\le\Delta} 2^{-j^2}$$

$\sum_{j\ge 0} 2^{-j^2}$ converges as $\Delta \to \infty$

$$= \Theta\left(\tfrac{1}{\sqrt n}\right) 2^{n^2/4+n}$$

$\underline{C_5, C_7, C_9}$-free.



Claim: C_{2k+1}-free for $3 \le k \le 249$ (assume C_5-free)

Proof: $v_1 - \dots - v_{2k+1}$ $\qquad Q_i = Q_{v_i}$, $R_i = R_{v_i} = \Gamma(Q_{v_i})$

Double-Counting:

$K = \max \# R_i$ that any vx contained in

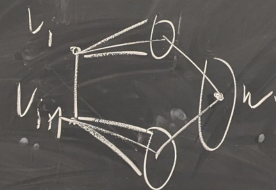
$$\sum_{\substack{w\in(n)\\ 1\le i\le 2k+1}} \mathbb{1}\{w\in R_i\} = \sum_{w\in(n)} \underbrace{\# R_i \text{ that contain } w}_{\le K} \le n\cdot K$$

$$= \sum_{1\le i\le 2k+1} \underbrace{\#\text{ vx's } w \text{ in } R_i}_{\ge(\frac12-\alpha)n} \ge (2k+1)\left(\tfrac12-\alpha\right)n.$$

$\Rightarrow K \geq k + \frac{1}{2} - \alpha(2k+1) > k$ for $k \leq 249$

$\Rightarrow K \geq k+1 \Rightarrow \exists\, w$ that is contained in $\geq k+1$ many R_i

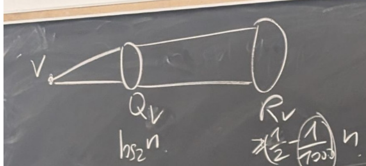
$\Rightarrow \exists\, i$ s.t. $w \in R_i \cap R_{i+1}$ (as only $2k+1$ vxs)

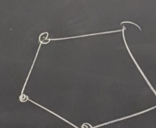


$\Rightarrow C_5$ ↯

C_5, C_7, C_9-free

$|R_v \cap R_w| \leq \frac{n}{100}$ if $vw \in E$





Claim: C_5-free

Proof: $v_1, \dots, v_5$

$Q_i = Q_{v_i}, \quad R_i = R_{v_i} = \Gamma(Q_{v_i})$

Double-Counting:

$N = \#$ vxs in at least three diff R_i

$$\sum_{\substack{w \in [n] \\ 1 \leq i \leq 5}} \mathbb{1}_{\{w \in R_i\}} \leq 5 \cdot N + 2(n - N) = 3N + 2n$$

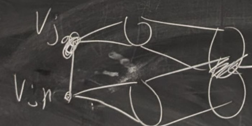
$$\geq 5\left(\frac{1}{2} - \alpha\right) n = \frac{5}{2} n - 5\alpha n$$

$\Rightarrow \quad 3N \geq \frac{n}{2} - 5\alpha n$

$\Rightarrow \quad N \geq \left(\frac{1}{6} - \frac{5\alpha}{3}\right) n > \frac{5}{100} n$

↳ for each such $u x w$ there is j s.t. $w \in R_j \cap R_{j+1}$

PHP $\Rightarrow$ $\exists$ set of $\frac{N}{5} > \frac{n}{100}$ vertices (for which there is <u>one</u> j s.t. $R_j \cap R_{j+1}$



$> \frac{n}{100}$ in common ↯ $(G \notin D(n))$

Follow-Up Courses

next quarter:

- Math 262B: Algebraic Combinatorics (Gen. Functs, Representation-Theory ...)
- Reading Seminar: discusses Research Papers (as Math 299) — follow-up on this course.

next year: Math 264A-C Probabilistic Combinatorics

$$\underbrace{\mathbb{P}_{\pi \text{ random}}\left(\pi(v)<\pi(w) \text{ for all } w\in\Gamma(v)\right)}_{\text{Prob.}} = \underbrace{\frac{\#\pi:\ \text{s.t. } \pi(v)<\pi(w)\ldots}{\text{all } \pi}}_{\text{Counting}}$$

① Inductive Counting (Change Problem Size: add vx or edge)

Basic Idea: Telescoping Product. & Control of $\frac{|S_{i+1}|}{|S_i|}$

$$|S_n| = \frac{|S_n|}{|S_{n-1}|} \cdot \frac{|S_{n-1}|}{|S_{n-2}|} \cdots \frac{|S_2|}{|S_1|} \cdot |S_1|$$

Examples: Word-Problem (existence), ISET-Problem, k-SAT ...

→ can go beyond 'vanilla' LLL. (Lovasz-Local-Lemma)

Advanced Ideas

For Δ-free with max-degree Δ:

$\chi(G) \leq (1+\varepsilon) \frac{\Delta}{\log \Delta}$ as $\Delta \to \infty$

(i) Identify $\frac{|S_{i+1}|}{|S_i|}$ as an expectation

(→ can use tools/ideas from $\mathbb{P}$-Theory)

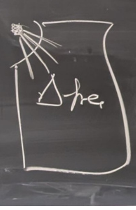
(ii) Sandw.-Idea:

$$L_n \subseteq S_n \subseteq L_n \cup B_n$$

Show: $|S_n| \leq (1+\gamma^{-n})|L_n|$

Modified Telescoping:

$$\frac{|B_n|}{|L_n|} = \underbrace{\frac{|B_n|}{|S_{n-x}|}}_{\leq \text{Enumeration (KEY)}} \cdot \underbrace{\frac{|S_{n-x}|}{|L_{n-x}|}}_{\overset{\text{I.H.}}{\leq} 1+\gamma^{-(n-x)}} \cdot \underbrace{\frac{|L_{n-x}|}{|L_n|}}_{\leq \text{Enum.}} \leq \text{small}$$



e.g. almost all Δ-free graphs are bipartite

(iii) "Clever Recursion": Stanley-Wilf (Marcus-Tardos)

Combine 'extremal number' with "Blockers"-Idea.

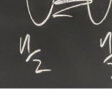
$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & \begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix} & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Outlook: $T_{n,m}$ be random $\triangle$-free graph with n vx's & m edges

$$\mathbb{P}(\chi(T_{n,m}) \leq 2) = \begin{cases} 1 & m \ll n \\ 0 & n \ll m \ll n^{3/2}\sqrt{\log n} \\ 1 & m \gg n^{3/2}\sqrt{\log n} \end{cases}$$

"Very Sparse" $\approx$ Random Graph

"Very Dense"

$\approx$ 

$n/2$ $n/2$

② Entropy-Method : "Sequentially revealing information in random order"

Need Representation $X=(X_1,\dots,X_N)$ of all objects in P.

unif distr. on P. $\longrightarrow$ random X

Π: random ord. of $1,\dots,N$

$$\log_2|P| = H(X_1,\dots,X_N)$$

$$= \sum_{i\in[N]} H(X_i \mid X_1,\dots,X_{i-1})$$

$$= \sum_{i\in[N]} \mathbb{E}_\Pi \underbrace{H(X_i \mid X_j : j <_\Pi i)}_{\le \mathbb{E}_X(\log_2 R(X_i) \mid X_j : j <_\Pi i)}$$

Range

Jensen, ..

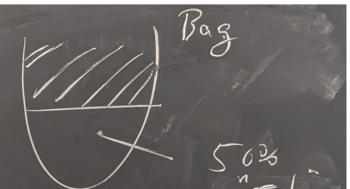
$$\le \mathbb{E}_X\Big(\sum_{i\in[N]} \log_2 \mathbb{E}_\Pi\big(R(X_i) \mid X_j : j <_\Pi i\big)\Big)$$

Idea ① although particular order can be forced into W.C.,
this is not case for random order
($\rightarrow$ typical behavior)

Idea ②: $R(X_i) = \sum_w I_w$ $\leftarrow$ write as sum of Indicators

$$\Longrightarrow \mathbb{E}_\Pi R(X_i) = \sum_w \mathbb{P}_\Pi(I_w = 1 \mid \dots)$$

this just focuses on ONE event



$n_1, n_2, n_3, \dots$

$\mathbb{P}$(random coin false) $\le \frac{1}{2}$

$\mathbb{P}$(no real coin in k steps) $\le \left(\frac{1}{2}\right)^k$

$\Longrightarrow$ Randomness hard to force into Worst Case

③ Asymp. Methods

- Estimating Sums by Integrals: $\sum_{k \in I} f(k) \approx \int_I f(x)\,dx$
- Laplace-Method: $\sum_{k \in I} f(k) \approx \sum_{k \in I^*} f(k) \overset{\text{Asymp. Approx}}{\approx} \sum_{k \in I^*} \hat{f}(k) \approx \int_{I^*} \hat{f}(x)dx \approx \int_I \hat{f}(x)\,dx$

"neglect tails"

often Gaussian-Int. $\int_0^\infty e^{-z^2} dz = \sqrt{\frac{\pi}{2}}$

e.g. $\sum_{0 \le k \le \lfloor \frac{n}{2} \rfloor} \binom{n-k}{k} \sim \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^{n+1}$

for fixed r: $\sum_{0 \le k \le n} \binom{n}{k}^r \sim \ldots$

- Bootstrapping: take estimate & improve it

e.g. • $w(z)e^{w(z)} = z$ as $z \to \infty$

can show: • $1 \le w(z) \le \log z$ for $z \ge e$ (by ↯ argument)

"Coarse Bound"

• $w(z) \ge \log z - \log\log z \longrightarrow w(z) \sim \log z$

• $w(z) = \log z - \log\log z + o(1)$

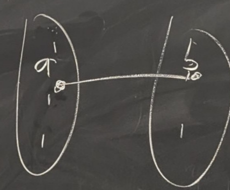
• $C_2(n)$

(4) Switching / Double Counting

"Perturbation Object: e.g. change of edge-locations"

Switching-Operator: $\phi: A \to 2^B$

aux. graph



ab edge if b can be reached by switching of a.

Double-Countin

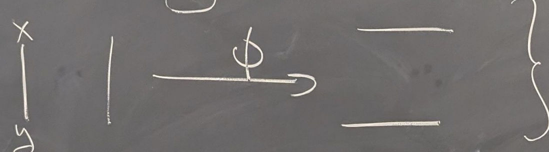
(# edges of aux gr.)

$$\sum_{a\in A} \deg a = \sum_{b\in B} \deg b$$

$$\Rightarrow \frac{|A|}{|B|} = \frac{\frac{1}{|B|}\sum_{b\in B}\deg b}{\frac{1}{|A|}\sum_{a\in A}\deg a} \approx \frac{d_B}{d_A}$$

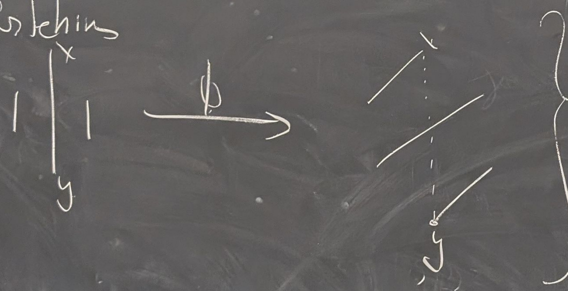
if approx regular $\deg a \approx d_A$, $\deg b \approx d_B$

e.g. 2-edge-Switching



typically gives bounds

3-edge-Switching



typically better/give asymptotics