BASIC DISCRETE MATHEMATICS

DAVID GALVIN, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF NOTRE DAME

Abstract. This document includes lecture notes, homework and exams from the Spring 2017 incarnation of Math 60610 — Basic Discrete Mathematics, a graduate course offered by the Department of Mathematics at the University of Notre Dame. The notes have been written in a single pass, and as such may well contain typographical (and sometimes more substantial) errors. Comments and corrections will be happily received at dgalvin1@nd.edu.

Contents

Date: Spring 2017; version of December 13, 2017.

1. Introduction

Discrete mathematics is the study of objects that are fundamentally discrete (made up of distinct and separated parts) as opposed to continuous; think "difference equations/recurrence relations" as opposed to "differential equations", or "functions whose domain is a finite set" as opposed to "functions whose domain is a real interval". It is an area of mathematics that has been increasing in importance in recent decades, in part because of the advent of digital computers (which operate and store data discretely), and in part because of the recent ubiquity of large discrete networks. Examples include social networks (e.g., the facebook friendship network), biological networks (e.g., the phylogenetic tree of species) and ecological (e.g., the food web).

Among the basic objects that are studied in this area are graphs — sets of points, some pairs of which are joined — which can be used to model relational structures; hypergraphs

— sets of subsets of a finite set; and permutations — bijective functions from an ordered set to itself. There are numerous well-developed branches of discrete mathematics, which can be loosely categorized by the sorts of questions they ask and answer. Some examples include:

- *enumerative*: how many objects are there satisfying a certain property?
- *extremal*: what is the largest/smallest/densest/sparsest object satisfying a certain property?
- *algorithmic*: how does one efficiently construct an object satisfying a certain property?
- *probabilistic*: what does a typical (randomly selected) object look like, given that it satisfies a certain property?
- *algebraic*: what is the underlying structure of the set of objects satisfying a certain property?

This categorization is far from discrete — these branches overlap with each other in pleasing and complex ways. The tools that are used to tackle discrete problems come from all over mathematics. The method of generating function is a powerful tool in enumeration problems, and draws heavily on both real and complex analysis. Algebraic tools, particularly tools from linear algebra, are invaluable in extremal problems. Differential equations are used to track the growth rates of discretely defined families. Methods from probability and information theory are ubiquitous.

This course acts as an introduction to contemporary discrete mathematics. Roughly, the plan is to touch on the following topics:

- *enumeration*: basic counting principles (including permutations, combinations, compositions, pigeon-hole principle and inclusion-exclusion), basic counting sequences (such as binomial coefficients, Catalan numbers, Euler numbers, and Stirling and Bell numbers), and recurrence relations and generating functions;
- *structure and existence*: Graphs (including trees, connectivity, Euler trails and Hamilton cycles, matching and coloring, Turan-type problems), partially ordered sets and lattices, basic Ramsey theory, error detecting and correcting codes, combinatorial designs, and techniques from probability and linear algebra;
- *other topics*: included if time permits.

The course will have no assigned text — these notes will be developed as the semester progresses. The following books well represent the level of the course, and will prove useful as reference resources:

- Stasys Jukna, Extremal combinatorics (with applications to computer science)
- Peter Cameron, Combinatorics (topics, techniques and algorithms)
- J.H. van Lint and R.M. Wilson, A course in Combinatorics
- Miklós Bóna, Introduction to enumerative combinatorics

Remark 1.1. *This version of the document has undergone minor corrections and modifications by Darij Grinberg.*

2. Graphs and trees — basic definitions and questions

We begin with a case study — labelled trees. We'll need some fairly straightforward definitions.

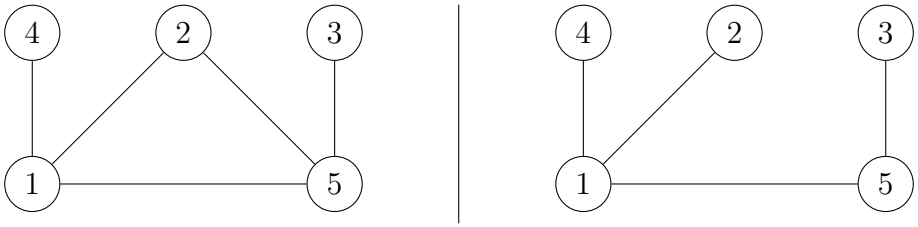
A *graph* $G = (V, E)$ is a set V of *vertices*, together with a set E of *edges*, with each element of E consisting of an unordered pair of distinct elements from V. Informally, a graph is a set of points (vertices), some pairs of which are joined by edges. We usually write an edge

$e = \{u, v\}$ as $e = uv$ and say that u and v are *adjacent* ($u \sim v$), that u an v are *endvertices* of e, and that u and v are *neighbours*. (What we have defined here is sometimes referred to as a *simple* (at most one edge between any pair of distinct vertices), *loopless* (no edges joining a vertex to itself), *undirected* (edges do not come with a built-in direction) graph).

A *walk* in G from u to v is an alternating list of vertices and edges of the form u, uw_1, w_1, w_1w_2, w_2, ..., w_k, w_kv, v. (We do allow this list to consist of one vertex and no edges when $u = v$.) Informally, a walk is a way of traveling from u to v along edges. The relation on vertices in which u is related to v iff there is a walk in G from u to v is easily seen to be an equivalence relation. The equivalence classes are called the *components* (sometimes *connected components*) of G. The graph is said to be *connected* if it has a single component; being connected means that is it possible to travel from any one vertex to any other along edges (and that the graph contains at least one vertex, because otherwise it would have zero components).

The following two figures illustrate two connected graphs:

(1)



An example of a walk in the first of these two graphs is 1, 12, 2, 25, 5, 53, 3, 35, 5, 51, 1, 12, 2. (A walk is allowed to use an edge multiple times.)

Given a graph G, we denote by $V(G)$ the set of vertices of G (and we say that G is a graph on vertex set $V(G)$); and we denote by $E(G)$ the set of edges of G.

A *tree* is a graph which is minimally connected — it is connected, but becomes disconnected on the removal of any edge. For instance, the second graph in (1) is a tree, but the first is not (since removing the edge 25 from the first graph leaves it connected). Trees are an important special class of graphs that arise naturally in a variety of applications — as, for example, decision trees, program logic trees, phylogenetic trees, and bracketology trees — and also form the backbone of the study of many properties of more general graphs.

To illustrate some of the various branches of discrete mathematics and combinatorics, I will ask, for each of the branches, a typical question from that branch that might be asked of trees.

- *enumerative*: Fix a vertex set $[n] := \{1, \ldots, n\}$. How many different trees are there on this vertex set? (What does it mean for two trees to be "different"? There are various possible answers to this; for our purposes, two trees on vertex set $[n]$ are different if they have distinct edge sets. This means, for example, that the trees T_1 and T_2 on vertex set [3] with $E(T_1) = \{12, 23\}$ and $E(T_2) = \{13, 32\}$ are considered different, even though they have the same essential "shape.")
- *extremal*: Among all trees on vertex set $[n]$, which have the most edges, and which have the fewest edges?
- *algorithmic*: Given a graph G on vertex set $[n]$, how quickly can one determine whether G is a tree? (It's open for discussion how we might measure "time" here; one possibility is to count to number of times the procedure you devise asks you to examine to list of edges [to check whether a particular pair of vertices are neighbors] in the worst case over all possible input graphs G).

- *probabilistic*: The *maximum degree* of a graph is the number of neighbours of (one of) the vertices with the largest number of neighbours. Select a tree from the set of all trees on vertex set $[n]$, with each tree equally likely to be selected. Consider the random variable Δ, that outputs the maximum degree of the chosen tree. What is the expected value of Δ?
- *algebraic*: Is there a natural underlying algebraic structure (e.g., a group structure, or a ring structure) to the set of trees on vertex set $[n]$?

Our main focus in this case study will be on the enumerative question above, but we will also answer the extremal question, and suggest implicitly some answers for some of the other questions. As we answer the extremal and enumerative questions, we will encounter "in the wild" some of the basic principles of counting that we will formalize in a little while, and will form the backbone of much of our work for the semester.

3. The extremal question for trees, and some basic properties

A little experimentation suggests that the extremal question for trees is not interesting.

Theorem 3.1. *Let $n \geq 1$. All trees on vertex set $[n]$ have $n-1$ edges.*

Proof. A *cycle* in a graph is a walk that

- starts and ends at the same vertex,
- does not repeat any other vertices,
- does not repeat any edge, and
- contains at least one edge.

A connected graph with a cycle is not minimally connected, since deleting any edge of a cycle maintains connectivity. It follows that a tree has no cycles.

Let tree T on vertex set $[n]$ by given, with edge set $\{e_1, \ldots, e_m\}$. Consider starting from the vertex set $[n]$ with no edges, and adding the edges $e_1, \ldots, e_m$ one-by-one, to form a sequence of graphs. We start (when no edges have been added) with a graph with n components. Each time a new edge is added, it must join two vertices that are in distinct components [if a new edge joined vertices in the same component, it would create a cycle], and so it must reduce the number of components by one. It follows that exactly $n-1$ edges must be added to reach a cycleless graph with a single component. □

In the above proof, we have seen that a tree is a connected graph without cycles. Conversely, a connected graph without cycles is a tree (for a proof, see problem (1) in Section 8).

Before going on to the enumerative question, it will be helpful to establish a few more very basic properties of trees and graphs. The first of these is often consider the "first theorem" of graph theory. For a vertex v of a graph G, $d(v)$ denotes the *degree* of v: the number of neighbours that v has. For example, the vertices 1, 2, 3, 4 and 5 of the first graph in (1) have degrees 3, 2, 1, 1 and 3, respectively.

Theorem 3.2. *For any graph $G = (V, E)$, the sum of the vertex degrees over all vertices equals twice the number of edges:*

$$\sum_{v \in V} d(v) = 2|E|.$$

Proof. As we sum $d(v)$ over all vertices v, each edge contributes to the sum exactly twice; specifically, the edge $e = u_1u_2$ contributes once to the sum when we consider $d(u_1)$, and contributes once when we consider $d(u_2)$. □

A more formal proof utilizes one of the simplest but most powerful methods in combinatorics, that of *double-counting*: counting a carefully chosen set of objects in two different ways, and comparing the answers. Indeed, let $\mathcal{P}$ be the set of all pairs (v, e) where v is a vertex of the graph under consideration, and e is an edge that has v as an endvertex. One way to enumerate the pairs in $\mathcal{P}$ is to consider, for each vertex v, how many edges e there are such that $(v, e) \in \mathcal{P}$. There are $d(v)$ such, and this yields

$$|\mathcal{P}| = \sum_{v \in V} d(v).$$

Another way to enumerate $\mathcal{P}$ is to consider, for each edge e, how many vertices v there are such that $(v, e) \in \mathcal{P}$. There are two such, and this yields

$$|\mathcal{P}| = \sum_{e \in E} 2 = 2|E|.$$

Equating the right-hand sides of these two expressions for $|\mathcal{P}|$ yields the result.

Corollary 3.3. *Let T be a tree on vertex set $V = [n]$.*

(a) *If $n \geq 1$, then $\sum_{v \in V} d(v) = 2n - 2$.*
(b) *If $n \geq 2$, then every vertex of T has degree at least 1, and T has at least two vertices that have degree 1.*

Proof. The claim of (a) is immediate on combining Theorem 3.1 with Theorem 3.2.

Let us prove the claim of (b). Assume that $n \geq 2$. A tree on at least two vertices clearly cannot have a vertex of degree 0, for otherwise it would have multiple components. Hence, every vertex of T has degree at least 1. Suppose that T has at most 1 vertex with degree 1. All remaining vertices have degree at least 2, and so

$$\sum_{v \in V} d(v) \geq 1 + 2(n-1) > 2n - 2,$$

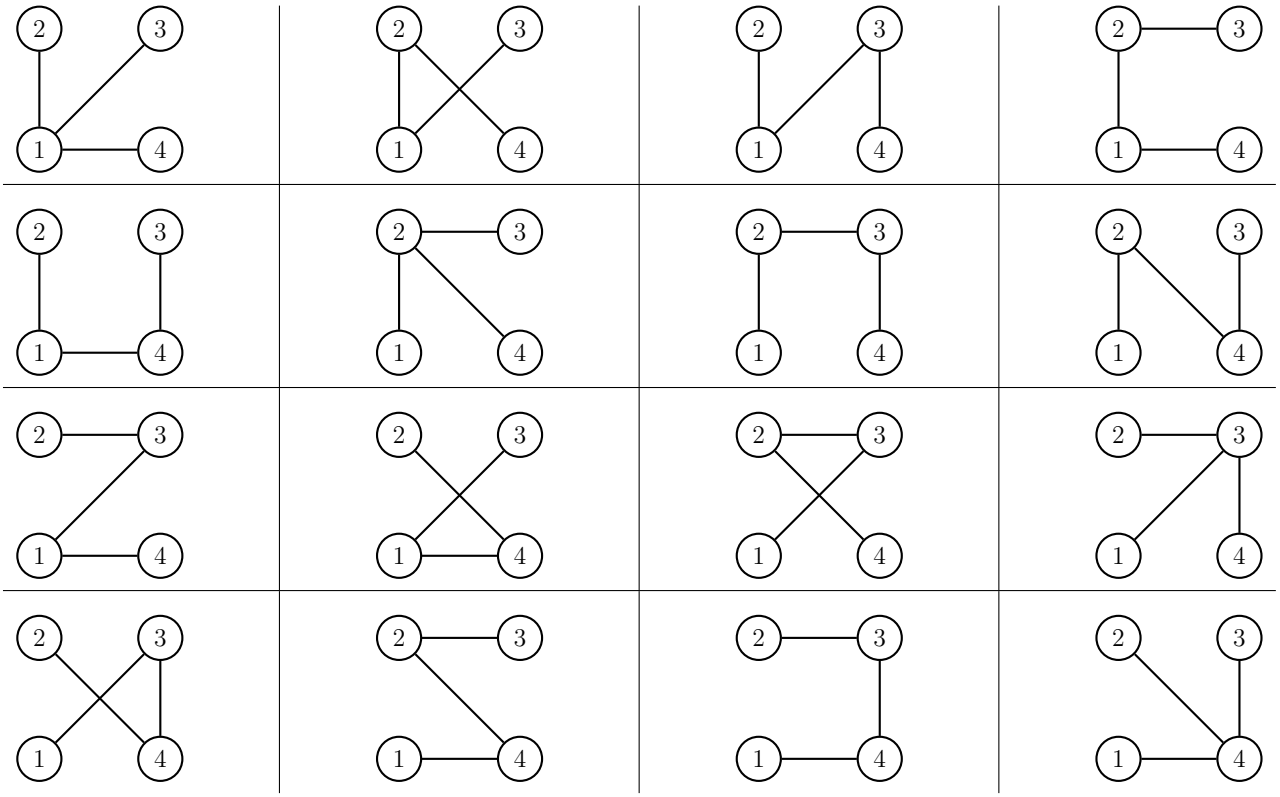
which contradicts the claim of (a). □

A vertex of degree 1 in a tree is called a *leaf*.

4. The enumerative question for trees — Cayley's formula

Here we address the question, how many different trees are there on vertex set $\{1, \ldots, n\}$, where two trees are considered different if they have distinct edge sets? Letting $t(n)$ denote the answer, we see that we are asking here not a single enumeration question, but an infinite family of them; the most satisfying answer to the question would be a simple closed formula for $t(n)$ (as a function of n).

To find $t(4)$, we need to count all trees on vertex set $[4] = \{1, 2, 3, 4\}$. There are 16 of them, all listed in the following table:



Similar doodling yields the following small values for $t(n)$:

- $t(1) = 1$,
- $t(2) = 1$,
- $t(3) = 3$,
- $t(4) = 16$, and
- $t(5) = 125$.

It's tempting to conjecture that $t(n) = n^{n-2}$ for $n \geq 1$. This is indeed true. It was first observed and proved by Borchardt in 1860[1], but mainly due to a widely-read 1889 paper of Cayley[2] it has come to be known as *Cayley's formula* [this is the first of many instances that we encounter in discrete mathematics of Stigler's law of eponymy: no scientific discovery is named after its original discoverer].

Theorem 4.1 (Cayley's formula)**.** *For $n \geq 1$, there are n^{n-2} trees on vertex set $\{1, \ldots, n\}$.*

The (short) proof will include many digressions, so rather than presenting it inside a proof environment, we move to a separate section.

5. Proof of Cayley's formula

Induction is a tempting approach to proving Theorem 4.1; removing a vertex from a graph without a cycle creates a smaller graph without a cycle. A slight problem is that the smaller

[1]C. W. Borchardt, Über eine Interpolationsformel für eine Art Symmetrischer Functionen und über Deren Anwendung, *Math. Abh. der Akademie der Wissenschaften zu Berlin* (1860), 1–20.

[2]A. Cayley, A theorem on trees, *Quart. J. Math* **23** (1889), 376–378.

graph may have many components and so not be a tree. An easy fix for this problem is to remove a leaf, so the smaller graph remains a tree. Indeed, it can be shown that

- removing a leaf from a tree always leaves a tree behind,
- and conversely, that if a new vertex is added to a tree and connected to exactly one of the other vertices, the resulting graph will be a tree.

(It may be easiest to prove these two facts using the equivalent characterizations of trees given in problem (1) in Section 8.)

A more noisome problem is that of controlling the "shape" of the trees being considered at various points in the proof. As with many simple-looking mathematical statements with few parameters, a good approach here is to first prove a rather more involved-looking statement, with many more parameters, and then deduce the original, simple-looking statement (the main point here being that the more involved-looking statement allows for a stronger induction hypothesis).

Here is what we will prove by induction on n, for $n \geq 2$:

> $P(n)$: fix a sequence $(d_1, d_2, \ldots, d_n)$ with $\sum_{i=1}^{n} d_i = 2n-2$, and with each d_i an integer that is at least 1. Let $t(d_1, \ldots, d_n)$ be the number of trees on vertex set $[n]$ for which vertex i has degree d_i. Then
>
> $$t(d_1, \ldots, d_n) = \frac{(n-2)!}{(d_1-1)! \ldots (d_n-1)!}.$$

Here $m!$ is shorthand for the m-fold product $m \cdot (m-1) \cdot (m-2) \ldots 2 \cdot 1$, with $0! = 1$. [This claim is not at all as obvious as Cayley's formula, but can be easily verified by hand for n up to, say, 6.]

Before proving this statement, we deduce Cayley's formula from it. By Corollary 3.3, we have

$$t(n) = \sum t(d_1, \ldots, d_n) = \sum \frac{(n-2)!}{(d_1-1)! \ldots (d_n-1)!}, \tag{2}$$

where both sums are over all sequences $(d_1, d_2, \ldots, d_n)$ with $\sum_{i=1}^{n} d_i = 2n-2$, and with each d_i an integer that is at least 1.

This is an example of a simple but crucial principle, the *addition principle*:

Principle 5.1 (The addition principle)**.** *If a set A is decomposed into two subsets A_1 and A_2 (that is, $A = A_1 \cup A_2$ and $A_1 \cap A_2 = \emptyset$) then*

$$|A| = |A_1| + |A_2|.$$

Put another way, if a process can be performed either *in one of m_1 ways,* or *in one of m_2 ways, with no overlap between the two sets of ways, then the total number of different ways it can be performed is $m_1 + m_2$.*

An obvious extension of this fact holds when a set is decomposed into more than two subsets.

We will use the addition principle everywhere, without much comment.

Let us take a little break to introduce some notations that will be used later on:

Definition 5.2. *Let A be a set.*

(a) *A* decomposition *of A means a list $(A_1, A_2, \ldots, A_k)$ of disjoint subsets of A such that $\bigcup_{j=1}^{k} A_j = A$. (We allow $k = 0$, but of course the empty list is a decomposition of A only if A is the empty set.) For example, the list $(\{1,5\}, \{2\}, \varnothing, \{3,4\})$ is a*

decomposition of the set $[5]$. *If* $(A_1, A_2, \ldots, A_k)$ *is a decomposition of* A, *then we say that* A is decomposed *(or* decomposes*) into the subsets* $A_1, A_2, \ldots, A_k$ *(or, more informally, into* $\bigcup_{j=1}^{k} A_j$*), and we say that* $A_1, A_2, \ldots, A_k$ *are the* parts *(or* blocks*) of this decomposition.*

(b) *An* ordered set partition *of* A *means a decomposition of* A *all of whose parts are non-empty. For instance,* $(\{1,5\}, \{2\}, \{3,4\})$ *is an ordered set partition of* $[5]$, *but* $(\{1,5\}, \{2\}, \varnothing, \{3,4\})$ *is not.*

(c) *A* set partition *(or, short,* partition*) of* A *means a set* $\{A_1, A_2, \ldots, A_k\}$ *of disjoint non-empty subsets of* A *whose union is* A *(that is, which satisfies* $\bigcup_{j=1}^{k} A_j = A$*). If* $\{A_1, A_2, \ldots, A_k\}$ *is a partition of* A, *then we say that* A is partitioned *(or* partitions*) into the subsets* $A_1, A_2, \ldots, A_k$, *and we say that* $A_1, A_2, \ldots, A_k$ *are the* parts *(or* blocks*) of this partition. Thus, a set partition differs from an ordered set partition in that it is a set, not a tuple, so its parts are not ordered (there is no "first part" and "second part" etc., but just the set of all its parts). For example,* $(\{1,2\}, \{3\})$ *and* $(\{3\}, \{1,2\})$ *are two distinct ordered set partitions of* $[3]$, *but the two partitions* $\{\{1,2\}, \{3\}\}$ *and* $\{\{3\}, \{1,2\}\}$ *of* $[3]$ *are identical.*

The difference between ordered set partitions and set partitions is often irrelevant, but becomes important when we start counting these objects; clearly, the number of set partitions of a set will usually be smaller than the number of ordered set partitions.

To get Cayley's formula from (2) we need the *multinomial formula*.

Theorem 5.3. *For each integer* $m \geq 0$, *we have*

$$(x_1 + x_2 + \ldots + x_\ell)^m = \sum \frac{m!}{a_1! \ldots a_\ell!} x_1^{a_1} \ldots x_\ell^{a_\ell},$$

where the sum is over all sequences $(a_1, a_2, \ldots, a_\ell)$ *with* $\sum_{i=1}^{\ell} a_i = m$, *and with each* a_i *an integer that is at least* 0.

Applying the multinomial formula with $m = n-2$, $\ell = n$ and each $x_i = 1$, we get

$$\begin{aligned} n^{n-2} &= \sum \frac{(n-2)!}{a_1! \ldots a_n!} \\ &= \sum \frac{(n-2)!}{(d_1-1)! \ldots (d_n-1)!} \end{aligned}$$

where the first sum is over all sequences $(a_1, a_2, \ldots, a_n)$ with $\sum_{i=1}^{\ell} a_i = n-2$, and with each a_i an integer that is at least 0, and the second sum (obtained from the first by a simple shift) is over all sequences $(d_1, d_2, \ldots, d_n)$ with $\sum_{i=1}^{\ell} d_i = 2n-2$, and with each d_i an integer that is at least 1. Combining this with (2) yields Cayley's formula.

We could prove the multinomial formula by induction, but it would be rather unpleasant. A more pleasant proof, that's much more in keeping with the spirit of the course, has a combinatorial flavour. When $(x_1 + x_2 + \ldots + x_\ell)^m$ is fully expanded out into a sum of monomials, it is easy to see that all monomials are of the form $x_1^{a_1} \ldots x_\ell^{a_\ell}$, where the sequence $(a_1, a_2, \ldots, a_\ell)$ of non-negative integers has $\sum_{i=1}^{\ell} a_i = m$, and that conversely each such sequence gives rise to a monomial in the expansion. So to prove the multinomial formula, we need only show that for each fixed sequence, the coefficient with which it occurs is $m!/(a_1! \ldots a_m!)$.

An occurrence of the monomial $x_1^{a_1} \ldots x_\ell^{a_\ell}$ corresponds exactly to a selection of a subset A_1 of a_1 elements from the set $[m] := \{1, \ldots, m\}$ (representing which a_1 of the m copies of

$(x_1 + \ldots + x_\ell)$ in the m-fold product $(x_1 + \ldots + x_\ell)(x_1 + \ldots + x_\ell)\ldots(x_1 + \ldots + x_\ell)$ from which we select the term x_1), followed by a selection of a_2 elements from the set $[m] \setminus A_1$ (representing the copies of $(x_1 + \ldots + x_\ell)$ from which we select the term x_2), and so on. So the calculation of the coefficient of $x_1^{a_1} \ldots x_\ell^{a_\ell}$ reduces to a counting problem: given a sequence $(a_1, a_2, \ldots, a_\ell)$ of non-negative integers with $\sum_{i=1}^{\ell} a_i = m$, in how many ways can we select $A_1 \subseteq [m]$, $A_2 \subseteq [m] \setminus A_1$, ..., $A_\ell \subseteq [m] \setminus (A_1 \cup A_2 \cup \ldots \cup A_{\ell-1})$, with $|A_i| = a_i$ for each i?

The most fundamental counting problem in combinatorics is contained in the one we have just encountered: how many subsets of size a does a set of size m have? We use the symbol $\binom{m}{a}$ (read "m choose a") for the answer; this *binomial coefficient* will be ubiquitous. While it will be must useful to think of $\binom{m}{a}$ as a quantity that counts something, it will also be helpful to have a way of computing it for various values of m and a.

Theorem 5.4. *For $m \geq 0$ and $0 \leq a \leq m$,*

$$\binom{m}{a} = \frac{m!}{a!(m-a)!}.$$

Proof. Let $\mathcal{S}$ be the set of all ways of ordering the elements of the set $\{1, \ldots, m\}$. We evaluate $|\mathcal{S}|$ in two different ways.

First, we make a direct count: there are m ways to decide the first element in the ordering, then $m-1$ ways to decide the second, and so on down to 1 way to decide the mth; this yields

$$|\mathcal{S}| = m!.$$

Next, we do a more indirect count: there are $\binom{m}{a}$ ways to chose the set of first a elements in the ordering, then $a!$ ways to order those elements, them $(m-a)!$ ways to order the remaining $m - a$ elements; this yields

$$|\mathcal{S}| = \binom{m}{a} a!(m-a)!.$$

The result follows from a combination of these two counts. □

The key point in the proof is an illustration of the second fundamental principle of counting, the *multiplication principle*.

Principle 5.5 (The multiplication principle)**.** *Let a set $\mathcal{A}$ be given, consisting of ordered pairs. Suppose that the set of elements that appear as first coordinates of elements of $\mathcal{A}$ has size m_1, and that, for each element x that appears as a first coordinate of an element of $\mathcal{A}$, there are exactly m_2 elements of $\mathcal{A}$ that have x as first coordinate. Then*

$$|A| = m_1 m_2$$

(with the obvious extension to a set consisting of ordered k-tuples). Put another way, if a process can be performed by first *performing one of m_1 first steps, and then, regardless of which first step was performed,* then *performing one of m_2 second steps, then the total number of different ways that the entire process can be performed is $m_1 m_2$.*

We will use the multiplication principle everywhere, without much comment. Note that it says more than just that $|A \times B| = |A||B|$, because the *set* of second coordinates/*set* of second steps is allowed to depend on the first coordinate/step; we just put a restriction on the sizes of the sets in question. For example, when ordering a set of size m, the specific set of $m - 1$ elements that are available to be put in the second positions depends very much on the choice of first element; the derivation of the formula $m!$ for the number of orderings

only requires that there always be *a* set of $m-1$ choices for the second stage of the ordering process.

Returning to the multinomial theorem, the answer to our counting problem that determines the coefficient of $x_1^{a_1}\dots x_\ell^{a_\ell}$ is

$$\binom{m}{a_1}\binom{m-a_1}{a_2}\binom{m-a_1-a_2}{a_3}\dots\binom{m-a_1-a_2-\dots-a_{\ell-1}}{a_\ell}.$$

After some algebra, this simplifies to $m!/(a_1!\dots a_\ell!)$, completing the proof of the multinomial theorem.

After that long digression into the multinomial theorem, we now turn to the (remarkably) short proof of $P(n)$ by induction on n, with the base case $n=2$ trivial. For $n\geq 3$, let sequence $(d_1,\dots,d_n)$ with each $d_i\geq 1$ and $\sum_{i=1}^n d_i = 2n-2$ be given. At least one of the d_i must be 1 (since otherwise, each d_i would be ≥ 2, and so we would have $\sum_{i=1}^n d_i \geq \sum_{i=1}^n 2 = 2n$, which would contradict $\sum_{i=1}^n d_i = 2n-2 < 2n$), and without loss of generality we assume $d_n = 1$. Thus, if T is a tree on vertex set $[n]$ with vertex i having degree d_i for each i, then the vertex n is a leaf of T, and hence is adjacent to exactly one of the vertices $1,2,\dots,n-1$. Therefore, the set $\mathcal{T}(d_1,\dots,d_n)$ of trees on $[n]$ with vertex i having degree d_i for each i decomposes into $\bigcup_{j=1}^{n-1}\mathcal{T}^j(d_1,\dots,d_{n-1})$, where $\mathcal{T}^j(d_1,\dots,d_{n-1})$ is the set of such trees with vertex n being adjacent only to vertex j. Hence,

$$\begin{aligned}\sum_{j=1}^{n-1}\left|\mathcal{T}^j(d_1,\dots,d_{n-1})\right| &= |\mathcal{T}(d_1,\dots,d_{n-1})| \\ &= t(d_1,\dots,d_{n-1}). \qquad (3)\end{aligned}$$

But for each given j, there is a one-to-one correspondence between trees in $\mathcal{T}^j(d_1,\dots,d_{n-1})$, and trees in $\mathcal{T}\left(d_1,\dots,d_{j-1},d_j-1,d_{j+1},\dots,d_{n-1}\right)$ (that is, trees on vertex set $[n-1]$ with vertex i having degree d_i for $i\neq j$, and vertex j having degree d_j-1): Indeed, the correspondence simply deletes vertex n from a tree in $\mathcal{T}^j(d_1,\dots,d_{n-1})$. It follows that

$$\begin{aligned}\left|\mathcal{T}^j(d_1,\dots,d_{n-1})\right| &= \left|\mathcal{T}\left(d_1,\dots,d_{j-1},d_j-1,d_{j+1},\dots,d_{n-1}\right)\right| \\ &= t\left(d_1,\dots,d_{j-1},d_j-1,d_{j+1},\dots,d_{n-1}\right) \\ &= \frac{(n-3)!}{(d_1-1)!\dots(d_{j-1}-1)!(d_j-2)!\,(d_{j+1}-1)!\dots(d_{n-1}-1)!}\end{aligned}$$

(by the induction hypothesis). Thus,

$$\begin{aligned}\left|\mathcal{T}^j(d_1,\dots,d_{n-1})\right| &= \frac{(n-3)!}{(d_1-1)!\dots(d_{j-1}-1)!(d_j-2)!\,(d_{j+1}-1)!\dots(d_{n-1}-1)!} \\ &= \frac{(n-3)!(d_j-1)}{(d_1-1)!\dots(d_j-1)!\dots(d_{n-1}-1)!} \\ &= \frac{(n-3)!(d_j-1)}{(d_1-1)!\dots(d_{n-1}-1)!} = \frac{(n-3)!(d_j-1)}{(d_1-1)!\dots(d_{n-1}-1)!(d_n-1)!}\end{aligned}$$

(since $d_n - 1 = 0$ and thus $(d_n - 1)! = 1$). Summing up these equalities over all $j = 1, 2, \ldots, n-1$, and comparing the result with (3), we obtain

$$\begin{aligned} t(d_1, \ldots, d_n) &= \sum_{j=1}^{n-1} \frac{(n-3)!(d_j - 1)}{(d_1 - 1)! \ldots (d_{n-1} - 1)!(d_n - 1)!} \\ &= \frac{(n-3)!}{(d_1 - 1)! \ldots (d_n - 1)!} \sum_{j=1}^{n-1} (d_j - 1) \\ &= \frac{(n-3)!}{(d_1 - 1)! \ldots (d_n - 1)!} \sum_{j=1}^{n} (d_j - 1) \\ &= \frac{(n-3)!}{(d_1 - 1)! \ldots (d_n - 1)!} \left(\sum_{i=1}^{n} d_i - n \right) \\ &= \frac{(n-3)!}{(d_1 - 1)! \ldots (d_n - 1)!} \left((2n-2) - n \right) \\ &= \frac{(n-2)!}{(d_1 - 1)! \ldots (d_n - 1)!}, \end{aligned}$$

as required, with the fifth equality using Corollary 3.3 (a). This completes the proof of Cayley's formula.

Buried in the proof of Cayley's formula is the third basic principle of counting, the *bijection principle*.

Principle 5.6 (The Bijection principle)**.** *Let A and B be two finite sets. If there is a bijection from A to B (a map which is both injective and surjective) then $|A| = |B|$.*

Bijective proofs are among the most satisfying in combinatorics, but in many cases are hard to come by, and subtle when they are found. When we use the bijection principle, it will usually be with a good deal of comment.

6. Prüfer's proof of Cayley

The answer to the question "what is $t(n)$?" is so simple — $t(n) = n^{n-2}$ — that it is tempting to look for a very simple proof that only utilized the bijection principle, and that proceeds by producing a bijection from the set of trees on vertex set $[n]$ to some set that "obviously" has size n^{n-2}. One candidate set is the set of all words[3] of length $n-2$ over alphabet $\{1, \ldots, n\}$. Prüfer[4] found such a bijection, that is remarkably simple to describe.

Start with a tree T on vertex set $[n]$, with $n \geq 2$, and associate with it a word as follows. Locate the leaf of T with the smallest name (in the usual ordering on $\{1, \ldots, n\}$), remove it from T (to form a new tree T') and record the name of the unique neighbour of the deleted vertex as the first letter of the associated word. Locate the leaf of T' with the smallest name, remove it to form T'', and record the name of the unique neighbour (in T') of the deleted vertex as the second letter of the associated word. Repeat until the tree is down to two

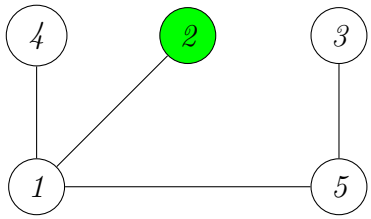
[3]A *word* over an *alphabet* A simply means a finite list of elements of A, that is, a tuple in A^k for some $k \in \mathbb{N}$. Usually, words are written as tuples without commas or parentheses – i.e., we write $a_1 a_2 \ldots a_k$ for a word $(a_1, a_2, \ldots, a_k)$ — when context allows. The *letters* of a word are its entries.

[4]H. Prüfer, Neuer Beweis eines Satzes über Permutationen, *Arch. Math. Phys.* **27** (1918), 742–744.

vertices, and stop (note that this means that the word associated with the unique tree on vertex set [2] is the empty word). The word of length $n-2$ over alphabet $\{1,\dots,n\}$ that is thus produced is called the *Prüfer code* of T. For example, the tree with edges 12, 23, ..., $(n-1)n$ would have Prüfer code $23\dots(n-1)$.

Example 6.1. *Let us show on an example how the Prüfer code of a tree is computed. (See* (4) *for a visualization.) Let T be the second graph in* (1)*. This is a tree on vertex set $[n]$ for $n=5$. To construct the Prüfer code of this tree T, we first locate the leaf of T with the smallest name; this leaf is* 2 *(since T has three leaves:* 2, 3 *and* 4*). The unique neighbor of this leaf is* 1*, so the Prüfer code of T will begin with* 1*. We now remove the leaf* 2 *from T to form a four-vertex tree T'. The leaf of T' with the smallest name is* 3*, and its unique neighbor is* 5*, so the second letter of the Prüfer code of T will be* 5*. Removing the leaf* 3 *from T' yields a three-vertex tree T''. The leaf of T'' with the smallest name is* 4*, and its unique neighbor is* 1*, so the third letter of the Prüfer code of T will be* 1*. Removing the leaf* 4 *from T'' yields a two-vertex tree T'''. At this point, the algorithm ends, and we conclude that the Prüfer code of T is* 151*.*

(4)



smallest leaf: 2*;*
its unique neighbor: 1*;*
Prüfer code of T starts with 1

smallest leaf: 3*;*
its unique neighbor: 1*;*
Prüfer code of T starts with 15

smallest leaf: 4*;*
its unique neighbor: 1*;*
Prüfer code of T starts with 151

only two vertices of T left;
Prüfer code of T is 151*.*

Theorem 6.2. *For $n \geq 2$, the map from the set of trees on vertex set $[n]$ to the set of words of length $n-2$ over alphabet $\{1, \ldots, n\}$ given by assigning to each tree its Prüfer code, is a bijection; in particular, $t(n) = n^{n-2}$.*

Before proving this theorem, it will be helpful to give a more formal definition of the Prüfer code, and to extend it beyond vertex set $\{1, \ldots, n\}$. Given a tree T on vertex set $\{x_1, \ldots, x_n\}$ ($n \geq 2$) on which an order $x_1 < x_2 < \ldots < x_n$ has been placed, the Prüfer code $P(T)$ of the tree is defined inductively as follows:

- If $n = 2$: for the unique tree T on vertex set $\{x_1, x_2\}$, $P(T)$ is the empty word.
- If $n \geq 3$: Let x_i be the least leaf (in the order $<$) of T, let x_j be the unique neighbor of x_i in T, and let T' be the tree on vertex set $\{x_1, \ldots, x_n\} \setminus \{x_i\}$ (with order induced from $<$) obtained from T by deleting x_i. $P(T)$ is the word $x_j P(T')$.

Evidently, $P(T)$ is a word of length $n-2$ over alphabet $\{x_1, \ldots, x_n\}$. A key tool in the proof of Theorem 6.2 is the following observation.

Claim 6.3. *Fix $n \geq 2$ and a tree T on vertex set $\{x_1, \ldots, x_n\}$. For each $1 \leq i \leq n$, the number of times that x_i occurs in $P(T)$ is one less than $d(x_i)$, the degree of x_i in T.*

Proof. We proceed by induction on n, with $n = 2$ evident. For $n \geq 3$, with the notation as above we have $P(T) = x_j P(T')$. The word $x_j P(T')$ evidently contains $0 = d(x_i) - 1$ occurrences of x_i. By induction, for each $i' \neq i$, $P(T')$ contains $d'(x_{i'}) - 1$ occurrences of $x_{i'}$, where d' indicates degree in T'. But for $i' \neq i, j$ we have $d'(x_{i'}) = d(x_{i'})$, and so we have the required $d(x_{i'}) - 1$ occurrences of $x_{i'}$ in $P(T)$. Finally, we have $d'(x_j) = d(x_j) - 1$ and so considering also the leading x_j we get the required $d(x_j) - 1$ occurrences of x_j in $P(T)$. □

We use this first to show that the map $T \mapsto P(T)$ is injective, by induction on n, with the case $n = 2$ evident. Let T_1, T_2 be two different trees on $\{x_1, \ldots, x_n\}$, $n \geq 3$. If the vectors $(d_1(x_1), \ldots, d_1(x_n))$ and $(d_2(x_1), \ldots, d_2(x_n))$ are different (where d_i indicates degree in T_i) then it is immediate from Claim 6.3 that $P(T_1) \neq P(T_2)$. If they are the same, then the trees T_1 and T_2 have the same leaves. Hence, there is some i such that x_i is the least leaf of both T_1 and T_2. If the unique neighbor of x_i in T_1 is different from the unique neighbour of x_i in T_2, then $P(T_1) \neq P(T_2)$, as they start out differently. If both have the same unique neighbour, then $P(T_1)$ and $P(T_2)$ begin with the same letter, but since $T_1' \neq T_2'$ (else $T_1 = T_2$) and so (by induction) $P(T_1') \neq P(T_2')$ they end with different words, and so again $P(T_1) \neq P(T_2)$.

Next we show that the map $T \mapsto P(T)$ is surjective, again by induction on n, with the case $n = 2$ evident. For $n \geq 2$, let w be a word of length $n-2$ over alphabet $\{x_1, \ldots, x_n\}$. Let x_i be the least letter that does not appear in w (there must be one such, since $n-2 < n$). Consider the word w' over alphabet $\{x_1, \ldots, x_n\} \setminus \{x_i\}$ obtained from w by deleting the leading letter of w, x_j, say. By induction there is a tree T' on vertex set $\{x_1, \ldots, x_n\} \setminus \{x_i\}$ with $P(T') = w'$. But then, by construction on Prüfer codes, it is evident that the tree T obtained from T' be adding a new vertex labelled x_i and joining it to x_j has Prüfer code $x_j w' = w$. (Indeed, the smallest leaf of T must be x_i, because if there was a smaller leaf, then this leaf would also have to be a leaf of T', whence (by Claim 6.3) it would be a letter not appearing in $P(T') = w'$, but this would contradict the fact that the smallest such letter is x_i.)

We have shown that the map $T \mapsto P(T)$ is injective and surjective, and so a bijection; in particular $t(n) = n^{n-2}$, the number of words of length $n-2$ over alphabet $\{1, \ldots, n\}$.

We may use Prüfer codes to easily recover our refinement of Cayley's formula. Fix $n \geq 2$, and a sequence $(d_1, \ldots, d_n)$ with $d_i \geq 1$ for each i, and $\sum_{i=1}^{n} d_i = 2n - 2$. Combining Claim

6.3 with the fact that $T \mapsto P(T)$ is a bijection, we find that the number of trees on vertex set $\{1, \ldots, n\}$ in which vertex i has degree d_i is equal to the number of words of length $n-2$ over alphabet $\{1, \ldots, n\}$ in which letter i appears exactly $d_i - 1$ times. This is the same as the number of ways of decomposing $\{1, \ldots, n-2\} = X_1 \cup \ldots \cup X_n$ into (possibly empty) sets $X_1, X_2, \ldots, X_n$ with each X_i satisfying $|X_i| = d_i - 1$ (with X_i representing the location in the Prüfer code of the $d_i - 1$ occurrences of letter x_i). As we will see in the equality (9) further below (and could easily argue now, if we chose), the number of such decompositions, and hence the value of $t(d_1, \ldots, d_n)$, is

$$\frac{(n-2)!}{(d_1-1)! \ldots (d_n-1)!}.$$

7. Otter's formula

A final note on Cayley's formula: we have so far been considering two trees to be different if they have different edge sets. We might also consider two trees to be different only if they have different "shapes". Formally, trees $T_1 = (V_1, E_1)$ and $T_2 = (V_2, E_2)$ are *isomorphic* if there is a bijection $f : V_1 \to V_2$ satisfying $f(x)f(y) \in E_2$ iff $xy \in E_1$; so for example all three trees on vertex set $\{1, 2, 3\}$ are isomorphic, but on $\{1, 2, 3, 4\}$ there are two isomorphism classes of trees, with representative elements having edge sets $\{12, 23, 34\}$ and $\{12, 13, 14\}$. The sequence $(\tilde{t}(n))_{n=1}^{\infty}$ that counts the number of isomorphism classes of trees on n vertices starts $(1, 1, 1, 2, 3, 6, 11, 23, 47, \ldots)$, and does not admit a closed formula. Otter[5], who spent the bulk of his career at Notre Dame, established a lovely asymptotic formula.

Theorem 7.1. *There are constants* $\alpha \approx 2.955$ *and* $\beta \approx .5349$ *such that*

$$\tilde{t}(n) \sim \beta \alpha^n n^{-5/2}.$$

8. Some problems

(1) Let G be a graph on vertex set V, $|V| = n$. Argue that the following are equivalent.
   (a) G is connected, but is no longer connected on the deletion of any edge.
   (b) G is acyclic (has no cycles), but is no longer acyclic with the addition of any edge.
   (c) G has $n-1$ edges and is connected.
   (d) G has $n-1$ edges and is acyclic.
   (e) G is connected and acyclic.

   The first of these is our definition of a tree, so this exercise gives lots of different characterizations of a tree.

   Solution: We start by arguing that the last three items are equivalent. Let G on vertex set $V = \{1, \ldots, n\}$ and edge set $E = \{e_1, \ldots, e_m\}$ be given, and let the graphs G_0, G_1, G_2, …, G_m each have vertex set $\{1, \ldots, n\}$ and (respectively) edge sets $\emptyset$, $\{e_1\}$, $\{e_1, e_2\}$, …, $\{e_1, e_2, \ldots, e_m\}$.

   Suppose G has $n-1$ edges and is connected. In going from G_i to G_{i+1}, the number of components either stays the same (if the edge e_{i+1} joins two vertices in the same component of G_i) or drops by 1 (if the edge e_{i+1} joins two vertices in different components). Note (crucially) that in the former case, the addition of the new edge creates a cycle in the component to which the edge has been added, whereas in the

[5]R. Otter, The number of trees, *Ann. of Math.* **49** (1948), 583–599.

latter case, the addition of the new edge creates no new cycles (because any cycle that contains this edge would link two formerly different connected components, but then it would have to use this new edge twice). Since G_0 has n components and $G_{n-1} = G$ has 1, it must be that each time we go from G_i to G_{i+1} we join two vertices in different components. Hence at no point do we join two vertices in the same component, and so never create a cycle. This shows that G is acyclic, so c) implies d).

Suppose G has $n-1$ edges and is acyclic. Then again it is forced that each time we go from G_i to G_{i+1} we join two vertices in different components, so drop the number of components, so end with 1 component. This shows that G is connected, so d) implies e).

Suppose G is connected and acyclic. Then each time we go from G_i to G_{i+1} we join two vertices in different components (else we would create a cycle), so each time we drop the number of components by 1, so we must do this $n-1$ times to get down from n to 1 components. This shows that G has $n-1$ edges, so e) implies c).

Now we tie the first two conditions into the last three.

Suppose G is minimally connected, but has more than n edges. Then at some point in going from G_i to G_{i+1} we must join two vertices in the same component, creating a cycle, a contradiction. And if it has fewer than $n-1$ edges, then the process of going through the G_i does not result in a connected graph, a contradiction. So minimally connected G has $n-1$ edges, and a) implies c), d) and e). On the other hand, if G is connected, acyclic and has $n-1$ edges, then on the deletion of any edge we have a graph with $n-2$ edges, which cannot be connected (we do not add enough edges to go from n components to 1 as we build the G_i's); so G is minimally connected, and so c), d) and e) imply a).

Suppose G is maximally acyclic (that is, it is acyclic, but becomes no longer acyclic if any new edge is added). It cannot have more than $n-1$ edges (for if it had, at some point in the process of building the G_i we reach a graph with one component, and have to create a cycle at the next step), nor can it have fewer than $n-1$ edges (the final graph in the process of building the G_i would have at least two components, so it would be possible to add an edge that does not create a cycle). So G has $n-1$ edges, and b) implies c), d) and e). On the other hand, if G is connected, acyclic and has $n-1$ edges, then clearly adding an edge creates a cycle, so c), d) and e) imply b).

(2) Cayley's formula says that there are n^{n-2} trees on vertex set $[n]$. How many *graphs* are there on the same vertex set?

Solution: There are $\binom{n}{2}$ unordered pairs of distinct vertices that can be selected from a set of n vertices. A graph on the vertex set corresponds to a selection of a subset of those $\binom{n}{2}$ unordered pairs (these are the edges that are included in the graph), and this correspondence is clearly bijective. So the number of graphs is the number of subsets, or

$$2^{\binom{n}{2}}.$$

(3) How many trees are there on vertex set $[n]$ that have vertex 1 as a leaf? (Note that such a tree may have many other leaves).

Solution: The degree sequence $(d_1, \ldots, d_n)$ of a tree on vertex set $[n]$ with vertex 1 a leaf is of the form $(1, d_2, \ldots, d_n)$ with each $d_i \geq 1$ $(i \geq 2)$ and with $1+\sum_{i=2}^{n} d_i = 2n-2$, or $\sum_{i=2}^{n} d_i = 2n-3$. By Cayley's refined formula (2), the number of trees that have

such a degree sequence is

$$\begin{aligned}
&\sum_{(d_2,\ldots,d_n):d_i\geq 1,\ \sum d_i=2n-3} \frac{(n-2)!}{(d_2-1)!\ldots(d_n-1)!} \\
= &\sum_{(e_2,\ldots,e_n):e_i\geq 0,\ \sum e_i=n-2} \frac{(n-2)!}{e_2!\ldots e_n!} \\
= &\sum_{(e_2,\ldots,e_n):e_i\geq 0,\ \sum e_i=n-2} \binom{n-2}{e_2,\ldots,e_n} \\
= &\ (1+1+\ldots+1)^{n-2}\ (n-1\ 1\text{'s in the sum}) \\
= &\ (n-1)^{n-2}.
\end{aligned}$$

Here we used re-indexing and the multinomial theorem.

Easier: By Claim 6.3 the Prüfer codes of the trees we are counting are precisely the words of length $n-2$ on alphabet $\{2,\ldots,n\}$; there are evidently $(n-1)^{n-2}$ such words.

(4) (Following on from the previous question) What is the average number of leaves in a tree on vertex set $[n]$? In other words, where $\ell(T)$ counts the number of leaves of the tree T, compute

$$\text{ave}(n) := \frac{1}{n^{n-2}} \sum_T \ell(T)$$

where the sum is over trees on vertex set $[n]$. Your answer should not involve a summation. (**Hint**: write $\ell(T)$ as a sum over elements of $[n]$.)

Solution: Writing[6] $\ell(T) = \sum_{i=1}^n \mathbf{1}_{\{i \text{ is a leaf of } T\}}$, we have

$$\begin{aligned}
\frac{1}{n^{n-2}} \sum_T \ell(T) &= \frac{1}{n^{n-2}} \sum_T \sum_{i=1}^n \mathbf{1}_{\{i \text{ is a leaf of } T\}} \\
&= \frac{1}{n^{n-2}} \sum_{i=1}^n \sum_T \mathbf{1}_{\{i \text{ is a leaf of } T\}} \\
&= \frac{1}{n^{n-2}} \sum_{i=1}^n \sum_T \mathbf{1}_{\{1 \text{ is a leaf of } T\}} \\
&= \frac{1}{n^{n-2}} \sum_{i=1}^n (n-1)^{n-2} \\
&= n\left(1-\frac{1}{n}\right)^{n-2}.
\end{aligned}$$

The third equality relies on the fact that (for each fixed $i \in [n]$) the number of trees having i as a leaf equals the number of trees having 1 as a leaf (indeed, relabeling vertices in some fixed way provides a bijection, as long as the relabeling turns i into 1). In the second-to-last equality we used the result of the last exercise.

[6]Here, we are using the notation $\mathbf{1}_{\{\mathcal{S}\}}$ for the *truth value* of any statement $\mathcal{S}$; this is simply the integer 1 if $\mathcal{S}$ is true, and 0 if $\mathcal{S}$ is false.

(5) (Following on from the previous question) Denote by $\text{ave}(n)$ the average number of leaves in a tree on vertex set $[n]$. Compute

$$\lim_{n\to\infty} \frac{\text{ave}(n)}{n}.$$

In other words, compute the limiting proportion of vertices that are leaves in a large random tree.

Solution: From the last exercise, and using basic calculus, we have

$$\lim_{n\to\infty} \frac{\text{ave}(n)}{n} = \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^{n-2} = \frac{1}{e}.$$

9. Some basic counting problems

Here we use the basic principles and methods introduced in the last few sections to answer a collection of very basic counting problems, that will be the building blocks for answering more complicated questions later.

Question 9.1. *How many words of length k are there over an alphabet of size n? Equivalently, how many functions are there from a set X of size k to a set Y of size n?*

There are n choices for the first letter, then n for the second, and so on, so by the multiplication principle, the number of words is n^k.

Question 9.2. *How many ways are there to arrange n objects in a row? Equivalently, how many bijections are there from a set X of size n to a set Y of size n, and also equivalently, how many* permutations *are there of the set $\{1, \ldots, n\}$? (A permutation is a bijective function from $[n]$ to itself.)*

There are n choices for the first object, then $n-1$ for the second, and so on, so by the multiplication principle, the number of arrangements is

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1 = n!$$

The expression "$n!$" is read "n factorial". For example, $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$ and $5! = 120$. By convention, we consider empty products (that is, products that have no factors) to evaluate to 1, so we obtain $0! = 1$. The quantity $n!$ grows rapidly: $0! = 1$, $10! = 3,628,800$, $20! \approx 2 \times 10^{18}$, $30! \approx 2 \times 10^{32}$, and $58! \approx 2 \times 10^{78}$, around the number of elementary particles in the universe. Later we will (probably) prove the following very useful asymptotic estimate, *Stirling's formula*.

Theorem 9.3.

$$n! \sim n^n e^{-n} \sqrt{2\pi n},$$

meaning that

$$\lim_{n\to\infty} \frac{n!}{n^n e^{-n} \sqrt{2\pi n}} = 1.$$

More precisely, for all $n \geq 1$

$$n^n e^{-n} \sqrt{2\pi n} \exp\{1/(12n+1)\} < n! < n^n e^{-n} \sqrt{2\pi n} \exp\{1/12n\}.$$

The precise statement is due to Robbins[7]. Even more precise statements exist, but I could never imagine a situation where one might need them!

[7]H. Robbins, A Remark on Stirling's Formula, *The American Mathematical Monthly* **62** (1955), 26–29.

Question 9.4. *How many ways are there to draw k objects from a set of size n, and arrange them in order?*

By the multiplication principle, the number is

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) = (n)_k$$

The expression "$(n)_k$" is read "n to the power k falling".

Question 9.5. *How many ways are there to arrange n objects around a circle, with only the order of the elements mattering (not their specific positions), and with two configurations considered the same if one can be obtained from the other by a rotation?*

An informal answer uses the *overcount principle*, which I won't formalize, but will simply state as: if an initial count of a set yields answer x, but in that initial count every element is counted exactly m times, then the size of the set is x/m. (For example, the number of sheep in a field is the number of legs divided by 4). Think of arranging n people around a circular table with n chairs labelled "1" through "n". There are $n!$ such arrangements. But we are counting a set of arrangements in which two are considered the same if they differ only by a rotation, so each such arrangement has been counted n times in our initial count (once for each of the n distinct rotations). Hence by the overcount principle, the number of cyclic arrangements is $n!/n = (n-1)!$.

Here's a more formal argument, using double counting: let $\mathcal{S}(n)$ be the set of pairs (A, a) where A is a cycle arrangement of $\{1, \ldots, n\}$ and a is one distinguished element in the arrangement (the "head"). Let $c(n)$ be the number of cyclic arrangements. For each element counted by $c(n)$, there are n choices for the head a, so $|\mathcal{S}(n)| = nc(n)$. On the other hand, for each choice of head a, there are $(n-1)!$ choices for a cyclic arrangement that has head a (there are $n-1$ choices for the element to the right of a, $n-2$ choices for the element two to the right of a, etc.), so also $|\mathcal{S}(n)| = n(n-1)!$. Equating these two expressions for $|\mathcal{S}(n)|$ yields $c(n) = (n-1)!$.

10. Subsets of a set

Here is the must fundamental counting problem in combinatorics.

Question 10.1. *Fix $n \geq 0$ and $0 \leq k \leq n$. How many subsets of size k does a set of size n have?*

We have already dealt with this question, in Theorem 5.4. Using the symbol $\binom{n}{k}$ (read: "n choose k") for the answer, we have

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}. \tag{5}$$

We can prove this informally via the overcount principle: in the set of $(n)_k$ draws of k objects from a set of size n, arranged in order, each subset of size k appears $k!$ times, once for each way of arranging the elements of a set of size k in order. A more formal double-counting argument is given in Theorem 5.4.

11. Binomial coefficient identities

The expression $\binom{n}{k}$, which we call a *binomial coefficient*, for reasons that will become clear when we see Identity 11.4 presently, satisfies numerous identities, some of which can be proved using (5) or induction, but most of which are more easily and more satisfyingly proved combinatorially.

We will leave aside the issue of giving a formal definition of the term "combinatorial proof", and merely say that if we wish to establish the validity of an identity

$$A = B$$

for two integers A and B, there will typically be two approaches that we will take that we will consider "combinatorial". The first approach proceeds by constructing a set $\mathcal{S}$ and showing, usually via two different arguments, that both $|\mathcal{S}| = A$ and $|\mathcal{S}| = B$ holds. The second approach proceeds by constructing two sets $\mathcal{S}_A$, $\mathcal{S}_B$, showing (usually quite easily) that $|\mathcal{S}_A| = A$ and $|\mathcal{S}_B| = B$, and then showing that $|\mathcal{S}_A| = |\mathcal{S}_B|$ by constructing a bijection from $\mathcal{S}_A$ to $\mathcal{S}_B$ (this is usually the delicate part of this approach).

We will see many examples of both approaches in this section. Before we go on, it will be helpful to extend the definition of $\binom{n}{k}$ to all integers k by setting

$$\binom{n}{k} = 0 \text{ if } k \notin \{0, \ldots, n\}.$$

Notice that this definition is entirely consistent with the combinatorial definition of $\binom{n}{k}$. Notice also that the formula

$$\binom{n}{k} = \frac{(n)_k}{k!}$$

is now valid for all $n, k \geq 0$. In what follows we will tend to only verify identities in the "interesting" range $n \geq 0$, $0 \leq k \leq n$, and leave it to the interested reader to confirm that the range of validity extends to $k \notin \{0, \ldots, n\}$.

For the moment we leave $\binom{n}{k}$ undefined for other values of n, k; we may later address the question of a more general meaning.

Identity 11.1 (Symmetry)**.**

$$\binom{n}{k} = \binom{n}{n-k}. \tag{6}$$

Proof. Informally: $\binom{n}{k}$ counts the subsets of size k of a set of size n, by directly selecting the subset, while $\binom{n}{n-k}$ counts the same thing, by selecting the complement of the subset.

More formally: For $k < 0$ and $k > n$ the result is trivial as both sides equal 0. For $0 \leq k \leq n$, let X be a set of size n and consider

$$\mathcal{S}_k = \{\text{subsets of size } k \text{ of } X\}$$

and

$$\mathcal{S}_{n-k} = \{\text{subsets of size } n-k \text{ of } X\}.$$

We evidently have $|\mathcal{S}_k| = \binom{n}{k}$ and $|\mathcal{S}_{n-k}| = \binom{n}{n-k}$, so it remains to show that $\mathcal{S}_k$ and $\mathcal{S}_{n-k}$ are in bijective correspondence.

Consider the (clearly well-defined) function $f : \mathcal{S}_k \to \mathcal{S}_{n-k}$ given by $f(A) = X \setminus A$. It is injective (if $A \neq B$ then evidently $X \setminus A \neq X \setminus B$) and surjective (given $C \in \mathcal{S}_{n-k}$, we have $X \setminus C \in \mathcal{S}_k$, and $f(X \setminus C) = C$), so is a bijection. □

Identity 11.2 (Boundary values)**.** *For $n \geq 0$,*

$$\binom{n}{0} = \binom{n}{n} = 1.$$

Proof. This is directly from the definition. □

Identity 11.3 (Pascal's identity)**.** *For $(n, k) \neq (0, 0)$,*

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$

Proof. For $k < 0$ and $k > n$ both sides are evidently 0. For $n \geq 1$ and $k = 0$ both sides are evidently 1. The meat of the identity lies in the range $n \geq 1$, $1 \leq k \leq n$.

Informally, $\binom{n}{k}$ counts the subsets of $\{1, \ldots, n\}$ of size k directly, while $\binom{n-1}{k} + \binom{n-1}{k-1}$ counts the same thing indirectly by first counting those subsets that do not include the element n ($\binom{n-1}{k}$ such sets, since each is a subset of $\{1, \ldots, n-1\}$ of size k), and then counting the rest, namely those that do include n ($\binom{n-1}{k-1}$ such sets, since each is associated with a subset of $\{1, \ldots, n-1\}$ of size $k-1$ by deleting element n).

Formally, let $\mathcal{S}_k(n)$ be the set of subsets of $\{1, \ldots, n\}$ of size k, and let $\mathcal{S}_k^{\text{in}}(n)$ be those that include element n, and $\mathcal{S}_k^{\text{out}}(n)$ be those that do not. We have

$$\mathcal{S}_k(n) = \mathcal{S}_k^{\text{in}}(n) \cup \mathcal{S}_k^{\text{out}}(n)$$

and $\mathcal{S}_k^{\text{in}}(n) \cap \mathcal{S}_k^{\text{out}}(n) = \emptyset$ (i.e., the pair $(\mathcal{S}_k^{\text{in}}(n), \mathcal{S}_k^{\text{out}}(n))$ forms a decomposition of $\mathcal{S}_k(n)$), and so

$$|\mathcal{S}_k(n)| = |\mathcal{S}_k^{\text{in}}(n)| + |\mathcal{S}_k^{\text{out}}(n)|.$$

We have $|\mathcal{S}_k(n)| = \binom{n}{k}$. We have a bijection f from $\mathcal{S}_k^{\text{in}}(n)$ to $\mathcal{S}_{k-1}(n-1)$ given by $f(A) = A \setminus \{n\}$, so

$$|\mathcal{S}_k^{\text{in}}(n)| = \binom{n-1}{k-1},$$

and a bijection f' from $\mathcal{S}_k^{\text{out}}(n)$ to $\mathcal{S}_k(n-1)$ given by $f'(A) = A$, so

$$|\mathcal{S}_k^{\text{out}}(n)| = \binom{n-1}{k}.$$

The result follows. □

Pascal's identity leads to the famous "Pascal's triangle"; Google it if you haven't ever seen it.

We now come to the identity that gives the binomial coefficients their name.

Identity 11.4 (Binomial theorem)**.** *For $n \geq 0$, and variables x, y,*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Proof. A proof by induction using Pascal's identity is possible, but messy (and not too illuminating). Instead, we consider two combinatorial proofs.

The first we have seen already: the binomial theorem is a special case ($\ell = 2$) of the multinomial formula (Theorem 5.3).

Here's a second, slightly more formal, combinatorial proof, that assumes that $x, y \in \mathbb{C}$. If $x = 0$ the identity is evident, so we may assume $x \neq 0$. Dividing through by x^n, the identity becomes equivalent to

$$(1+z)^n = \sum_{k=0}^{n} \binom{n}{k} z^{n-k}$$

for $z \in \mathbb{C}$. We prove this identity in the case when z is a positive integer by observing that both sides count the number of words of length n that can be formed using the alphabet $\{0, 1, \ldots, z\}$. The left-hand side counts this by choosing the letters one at a time. The right-hand side does the count by first deciding how many times the letter "0" occurs in the word (this is the index k), then deciding the locations of the k 0's (this is the $\binom{n}{k}$), then deciding the remaining letters one-by-one (this is the z^{n-k}). To extend from positive integer z to arbitrary complex z, note that the right- and left-hand sides of the identity are both polynomials over $\mathbb{C}$ of degree (at most) n, so if they agree on $n+1$ values, they must agree at all values; and we have in fact shown that they agree on infinitely many values. □

The second proof above used the *polynomial principle*.

Principle 11.5. *If $f(x)$ and $g(x)$ are polynomials over $\mathbb{C}$, and there are infinitely many distinct x for which $f(x) = g(x)$, then $f(x)$ and $g(x)$ are identical. Equivalently, if $f(x)$ is a polynomial over $\mathbb{C}$, and there are infinitely many distinct x for which $f(x) = 0$, then $f(x)$ is identically 0.*

We may later formulate a polynomial principle for multivariable polynomials.

In the case $x = y = 1$ the binomial theorem yields

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n, \tag{7}$$

an identity that also admits a direct combinatorial proof — both sides count the number of subsets of a set of size n. The left-hand side does this by first specifying the size of the subset, and the right-hand side does it by going through the elements of the set one after another, and deciding for each element whether it is in the subset or not.

Setting $x = 1$ and $y = -1$ gives (for $n \geq 1$)

$$\sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} = 0, \tag{8}$$

or

$$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \ldots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \ldots;$$

a finite non-empty set has as many even-sized subsets as odd-sized. This admits a nice bijective proof, which is left as an exercise.

The *multinomial theorem*, which we have seen before (Theorem 5.3), is a generalization of the binomial theorem to more terms. Before re-stating and re-proving it, we introduce another basic counting problem.

Question 11.6. *Fix $n \geq 0$ and $a_1, \ldots, a_\ell$ with each $a_i \geq 0$ and $\sum_{i=1}^{\ell} a_i = n$. In how many ways can a set X of size n be decomposed as $X = X_1 \cup X_2 \cup \ldots \cup X_\ell$ with each $|X_i| = a_i$? (Recall that the sets X_i are allowed to be empty, and two decompositions differing in the order of the $X_1, X_2, \ldots, X_\ell$ count as different.)*

We denote by

$$\binom{n}{a_1, \ldots, a_\ell}$$

the answer, and refer to this as a *multinomial coefficient*. Note that the binomial coefficient $\binom{n}{k}$ is an instance of a multinomial coefficient, being equal to $\binom{n}{k,n-k}$ (because selecting a subset of size k is the same as decomposing into a (chosen) subset of size k and a (complementary, unchosen) subset of size $n-k$). To calculate the multinomial coefficient, we consider first selecting X_1 from X, then selecting X_2 from $X \setminus X_1$, and so on; this leads to the formula

$$\begin{aligned} \binom{n}{a_1, \ldots, a_\ell} &= \binom{n}{a_1}\binom{n-a_1}{a_2}\binom{n-a_1-a_2}{a_3}\ldots\binom{n-a_1-a_2-\ldots-a_{\ell-1}}{a_\ell} \\ &= \frac{n!}{a_1!\ldots a_\ell!}, \end{aligned} \tag{9}$$

the second equality employing (5) repeatedly.

Theorem 11.7. *For each integer $m \geq 0$ we have*

$$(x_1 + x_2 + \ldots + x_\ell)^m = \sum \binom{m}{a_1, \ldots, a_\ell} x_1^{a_1} \ldots x_\ell^{a_\ell},$$

where the sum is over all sequences $(a_1, a_2, \ldots, a_\ell)$ with $\sum_{i=1}^{\ell} a_i = m$, and with each a_i an integer that is at least 0.

Proof. When $(x_1 + x_2 + \ldots + x_\ell)^m$ is fully expanded out into a sum of monomials, it is easy to see that all monomials are of the form $x_1^{a_1} \ldots x_\ell^{a_\ell}$, where the sequence $(a_1, a_2, \ldots, a_\ell)$ of non-negative integers has $\sum_{i=1}^{\ell} a_i = m$, and that conversely each such sequence gives rise to a monomial in the expansion. So to prove the multinomial formula, we need only show that for each fixed sequence, the coefficient with which it occurs is $\binom{m}{a_1,\ldots,a_\ell}$.

When expanding the m-fold product $(x_1 + x_2 + \ldots + x_\ell)^m = (x_1 + \ldots + x_\ell)(x_1 + \ldots + x_\ell)\ldots(x_1 + \ldots + x_\ell)$, each monomial we get comes from selecting one addend from the first copy of $x_1 + \ldots + x_\ell$, one addend from the second copy, one addend from the third copy, and so on until the m-th copy. For the monomial to be $x_1^{a_1} \ldots x_\ell^{a_\ell}$, we must select the addend x_1 exactly a_1 times, the addend x_2 exactly a_2 times, and so on; the only freedom we have is in deciding the order of selection. An occurrence of the monomial $x_1^{a_1} \ldots x_\ell^{a_\ell}$ thus corresponds exactly to a selection of a_1 elements A_1 from the set $[m] := \{1, \ldots, m\}$ (representing which a_1 of the m copies of $(x_1 + \ldots + x_\ell)$ we select the addend x_1 from), followed by a selection of a_2 elements from the set $[m] \setminus A_1$ (representing the copies of $(x_1 + \ldots + x_\ell)$ from which we select the term x_2), and so on. By the answer to Question 11.6, the number of such selections is $\binom{m}{a_1,\ldots,a_\ell}$. □

We briefly mention two other counting problems to which the answer is a multinomial coefficient. The first one concerns the anagrams of a word. An *anagram* of a word w is a word obtained from w by permuting its letters.

Question 11.8. *An n-letter word has a_i repetitions of letter i, $i = 1, \ldots, k$ (so each $a_i \geq 0$, and $\sum_{i=1}^{k} a_i = n$). How many distinct anagrams does it have?*

It is evident that the answer here is $\binom{n}{a_1,\ldots,a_k}$ (the locations of the a_1 occurrences of letter 1 correspond to a selection of a subset A_1 of size a_1 from $\{1, \ldots, n\}$; the locations of the a_2

occurrences of letter 2 correspond to a subsequent selection of a subset A_2 of size a_2 from $\{1,\ldots,n\}\setminus A_1$; and so on). Equivalently, $\binom{n}{a_1,\ldots,a_k}$ is the number of ways of arranging in order the elements of a *multiset* M (a set in which elements may appear with multiplicity greater than 1), where M has k elements $\{x_1,\ldots,x_k\}$, with x_i appearing with multiplicity a_i.

A *lattice path* in $\mathbb{Z}^d$ is a sequence $(p_0,p_1,\ldots,p_n)$ with each $p_i\in\mathbb{Z}^d$, $p_0=(0,\ldots,0)$, and, for each $i=1,\ldots,d$, the vector p_i-p_{i-1} being one of the standard basic vectors (with one coordinate equal to 1 and $d-1$ coordinates equal to 0). A lattice path can be visualized as a walk in $\mathbb{R}^d$ starting at the origin, with each step being a unit step parallel to one of the axes and always in the positive direction. The *length* of the lattice path is n, and the path is said to *end* at p_n. It is evident that the answer to the following question is the multinomial coefficient $\binom{n}{a_1,\ldots,a_d}$; we may find this interpretation useful later.

Question 11.9. *How many lattice paths of length n are there in $\mathbb{Z}^d$, that end at $(a_1,\ldots,a_d)$?*

There are a staggering array of identities involving binomial coefficients (some of which we have seen already); Riordan[8] has a classic book devoted to them, while Gould[9] has catalogued a vast number of them on his website. Most of these identities admit a variety of proofs — inductive, combinatorial and analytic. Below, I mention only some famous identities that admit easy combinatorial proofs; later in the semester, when we address generating functions, we'll see other ways of entering the world of identity-proving.

Identity 11.10 (The upper summation identity)**.** *For $k\geq 0$ and $n\geq k$*

$$\sum_{m=k}^{n}\binom{m}{k}=\binom{n+1}{k+1}.$$

When $k=0$ this reduces to the vacuous statement that the sum of $n+1$ 1's is $n+1$. At $k=1$ we recover the well-known and well-loved identity

$$1+2+3+\ldots+n=\frac{n(n+1)}{2}.$$

The upper summation identity is sometimes referred to as the *hockey stick identity*; the reason for this can be seen by tracing the terms of the identity out on Pascal's triangle.

Proof. (Upper summation identity) Denote by $\mathcal{S}_{n+1,k+1}$ the set of subsets of $\{1,\ldots,n+1\}$ of size $k+1$, and for each $k\leq m\leq n$ let $\mathcal{S}^m_{n+1,k+1}$ be those subsets whose largest element is $m+1$.

We have $|\mathcal{S}_{n+1,k+1}|=\binom{n+1}{k+1}$. We also have that the $\mathcal{S}^m_{n+1,k+1}$'s form a decomposition of $\mathcal{S}_{n+1,k+1}$ as m runs from k to n (each element of $\mathcal{S}_{n+1,k+1}$ has a largest element, and it is somewhere between $k+1$ and $n+1$). So if we can show $|\mathcal{S}^m_{n+1,k+1}|=\binom{m}{k}$, we are done. But this is straightforward — there is an obvious bijection $f:\mathcal{S}_{m,k}\to\mathcal{S}^m_{n+1,k+1}$, namely $f(A)=A\cup\{m+1\}$, and $|\mathcal{S}_{m,k}|=\binom{m}{k}$. □

Notice that we are sometimes presenting the combinatorial proofs of identities quite formally, as with the upper summation identity above, and sometimes more informally, as for example with the brief discussion of (7). There is great value to being comfortable with both styles of presentation. An informal description of a combinatorial argument, as long as it is written with suitable clarity, tends to convey very clearly what is essentially going on in an

[8]J. Riordan, *Combinatorial identities*, John Wiley and Sons, New York, 1968.

[9]H. Gould, West Virginia University, `http://www.math.wvu.edu/~gould/`.

argument. On the other hand, "the devil is in the details" sometimes, and it is essential to be able to turn an informal argument into a carefully notated formal one to make sure that everything is correct. Moreover, we will sometimes encounter combinatorial arguments that will be involved enough that careful (formal) notation is essential for clarity.

The presentation of the proof of the upper summation identity above represents a good prototype for a more formal presentation of a combinatorial argument. For the record, here's a good prototype for a more informal presentation of the same argument:

"The right-hand side is the number of subsets of $\{1,\dots,n+1\}$ of size $k+1$, counted directly. The left-hand side counts the same, by first specifying the largest element in the subset (if the largest element is $k+1$, the remaining k elements must be chosen from $\{1,\dots,k\}$, in $\binom{k}{k}$ ways; if the largest element is $k+2$, the remaining k must be chosen from $\{1,\dots,k+1\}$, in $\binom{k+1}{k}$ ways; etc.)."

The remaining identities are left as exercises, all appearing in the next section.

12. Some problems

In the problems in this section that merely state an identity without given a question, the question is always to exhibit a combinatorial proof of the identity.

(1) From the binomial theorem we find that for $n \geq 1$

$$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \ldots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \ldots.$$

Give a bijective combinatorial proof of this fact. That is, construct sets $\mathcal{S}_1$ and $\mathcal{S}_2$ with $|\mathcal{S}_1| = \binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \ldots$ and $|\mathcal{S}_2| = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \ldots$, and exhibit a bijection between $\mathcal{S}_1$ and $\mathcal{S}_2$.

(2)

Identity 12.1 (The parallel summation identity)**.** *For $m, n \geq 0$*

$$\sum_{k=0}^{n} \binom{m+k}{k} = \binom{n+m+1}{n}.$$

Proof. The right-hand side is the number of subsets of $\{1,\dots,n+m+1\}$ of size n, counted directly. The left-hand side counts the same, by first specifying the smallest element *not* in the subset (if the smallest missed element is 1, all n elements must be chosen from $\{2,\dots,n+m+1\}$, in $\binom{m+n}{n}$ ways, giving the $k=n$ term; if the smallest missed element is 2, then 1 is in the subset and the remaining $n-1$ elements must be chosen from $\{3,\dots,n+m+1\}$, in $\binom{m+n-1}{n-1}$ ways, giving the $k=n-1$ term; etc., down to: if the smallest missed element is $n+1$, then all of $1,\dots,n$ are in the subset and the remaining 0 elements must be chosen from $\{n+2,\dots,n+m+1\}$, in $\binom{m+0}{0}$ ways, giving the $k=0$ term). □

(3) Derive the parallel summation identity from the upper summation identity, using an early, extremely simple binomial coefficient identity.

(4)

Identity 12.2 (The cancellation, or committee-chair, identity)**.** *For $n \geq k \geq 1$*

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1} \text{ or } k\binom{n}{k} = n\binom{n-1}{k-1}.$$

(5)

Identity 12.3 (The committee-subcommittee identity)**.** *For $n \geq k \geq r \geq 0$*

$$\binom{k}{r}\binom{n}{k} = \binom{n}{r}\binom{n-r}{k-r}.$$

(6)

Identity 12.4 (Vandermonde's identity)**.** *For $m, n, r \geq 0$*

$$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k}.$$

A remark is in order here. In the particular case $m = n = r$, Vandermonde's identity becomes

$$\sum_{k=0}^{n} \binom{n}{k}\binom{n}{n-k} = \binom{2n}{n}.$$

By symmetry (i.e., the identity (6)), this rewrites as

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}.$$

There is no similarly nice expression for $\sum_{k=0}^{n} \binom{n}{k}^r$ for any $r > 2$.

(7)

Identity 12.5 (The binomial theorem for falling powers)**.** *With the notation $(x)_k = x(x-1)\ldots(x-k+1)$ for all non-negative integers k, we have*

$$(x+y)_n = \sum_{k=0}^{n} \binom{n}{k}(x)_k(y)_{n-k}.$$

Here x and y are (complex) variables, and $n \geq 0$.

The polynomials $(x)_k = x(x-1)\ldots(x-k+1)$ for all $k \geq 0$ (with x being the variable) are known as the *falling powers* or *falling factorials*; another common notation for them is $x^{\underline{k}}$. They form a basis of the vector space of all polynomials (in one variable x, over any ground field); indeed, they are linearly independent (since they have distinct degrees), and any polynomial can be represented as a linear combination of these polynomials $(x)_k$. (To verify the latter claim, we can proceed by induction over the degree: If q is a polynomial of degree d with leading term $c_d x^d$, then the polynomial $q - c_d(x)_d$ has smaller degree than q, and we can obtain a representation of q as a linear combination of the $(x)_k$ from a similar representation of $q - c_d(x)_d$.)

Proof of the binomial theorem for falling powers. Here's a combinatorial proof. Let x and y be positive integers. The number of words in alphabet $\{1, \ldots, x\} \cup \{x+1, \ldots, x+y\}$ of length n with no two repeating letters, counted by selecting letter-by-letter, is $(x+y)^{\underline{n}}$. If instead we count by first selecting k, the number of letters from $\{1, \ldots, x\}$ used, then locating the k positions in which those letters appear, then selecting the $n-k$ letters from $\{x+1, \ldots, x+y\}$ letter-by-letter in the order that they appear in the word, and finally selecting the k letters from $\{1, \ldots, x\}$ letter-by-letter in the order that they appear in the word, we get a count of $\sum_{k=0}^{n} \binom{n}{k} x^{\underline{k}} y^{\underline{n-k}}$. So the identity is true for positive integers x, y.

The left-hand side and right-hand side are polynomials in x and y of degree n, so the difference is a polynomial in x and y of degree at most n, which we want to show is identically 0. Write the difference as $P(x,y) = p_0(x) + p_1(x)y + \ldots + p_n(x)y^n$ where each $p_i(x)$ is a polynomial in x of degree at most n. Setting $x = 1$ we get a polynomial $P(1,y)$ in y of degree at most n. This is 0 for all integers $y > 0$ (by our combinatorial argument), so by the polynomial principle it is identically 0. So each $p_i(x)$ evaluates to 0 at $x = 1$. But the same argument shows that each $p_i(x)$ evaluates to 0 at any positive integer x. So again by the polynomial principle, each $p_i(x)$ is identically 0 and so $P(x,y)$ is. This proves the identity for all real x, y. □

(8) Derive the binomial theorem for falling powers from Vandermonde's identity.
(9) Using the binomial theorem for falling powers, derive the following.

Identity 12.6 (The binomial theorem for rising powers)**.** *With the notation* $x^{(k)} = x(x+1)\ldots(x+k-1)$ *for all non-negative integers* k*, we have*

$$(x+y)^{(n)} = \sum_{k=0}^{n} \binom{n}{k} x^{(k)} y^{(n-k)}.$$

Here again x *and* y *are (complex) variables, and* $n \geq 0$*.*

The polynomials $x^{(k)} = x(x+1)\ldots(x+k-1)$ for all $k \geq 0$ (with x being the variable) are known as the *rising powers* or *rising factorials*; another common notation for them is $x^{\overline{k}}$. They form a basis of the vector space of all polynomials (in one variable x, over any ground field); the proof is analogous to the corresponding proof for falling powers.

Also, falling powers and rising powers are related by the identity

$$x^{(k)} = (x+k-1)_k = (-1)^k (-x)_k = (-1)^k (-x-k+1)^{(k)}.$$

Proof of the binomial theorem for rising powers. Set $x' = -x$ and $y' = -y$; we have

$$(x+y)^{\overline{n}} = (-x'-y')^{\overline{n}} = (-1)^n (x'+y')^{\underline{n}}$$

and

$$\begin{aligned}
\sum_{k=0}^{n} \binom{n}{k} x^{\overline{k}} y^{\overline{n-k}} &= \sum_{k=0}^{n} \binom{n}{k} (-x')^{\overline{k}} (-y')^{\overline{n-k}} \\
&= \sum_{k=0}^{n} \binom{n}{k} (-1)^k (x')^{\underline{k}} (-1)^{n-k} (y')^{\underline{n-k}} \\
&= (-1)^n \sum_{k=0}^{n} \binom{n}{k} (x')^{\underline{k}} (y')^{\underline{n-k}},
\end{aligned}$$

so the identity follows from the binomial theorem for falling powers. □

(10)

Identity 12.7 (The hexagon identity)**.** *For* $n \geq 1$ *and* $m \geq 0$

$$\binom{n-1}{m-1}\binom{n}{m+1}\binom{n+1}{m} = \binom{n}{m-1}\binom{n-1}{m}\binom{n+1}{m+1}.$$

Note that an algebraic proof of this identity is very easy. You will probably find a combinatorial proof very hard! To see why this is called the hexagon identity, mark the terms involved in Pascal's triangle. [10]

(11) The identity $\sum_{k=0}^{n}(-1)^k\binom{n}{k}=0$ (for $n\geq 1$) tells us that a set of size n has as many even-sized subsets as odd-sized subsets. Because 2^n, the number of subsets of a set of size n, is only divisible by powers of 2, such a clean statement will not hold for, for example, subsets of size a multiple of 3, one greater than a multiple of 3, and 2 greater than a multiple of 3. However, there is an approximate version, which can be proven analytically/algebraically.

Show that for each fixed $\ell\geq 2$, the proportion of subsets of a set of size n that have size evenly divisible by ℓ tends, in the limit as n grow, to $1/\ell$. That is, show that

$$\lim_{n\to\infty}\frac{\binom{n}{0}+\binom{n}{\ell}+\binom{n}{2\ell}+\cdots}{2^n}=\frac{1}{\ell}.$$

Solution: Let ω be a primitive ℓth root of unity; that is, a complex number ω with the property that $\omega^\ell=1$ but $\omega^j\neq 1$ for any $1\leq j<\ell$; $\omega=e^{2\pi i/\ell}$ works nicely here.

Apply the binomial theorem to expand $(1+x)^n$ for $x=1,\omega,\omega^2,\ldots,\omega^{k-1}$, and sum to get

$$\sum_{k=0}^{\ell-1}(1+\omega^k)^n=\sum_{k=0}^{\ell-1}\sum_{j=0}^{n}\binom{n}{j}(\omega^k)^j=\sum_{j=0}^{n}\binom{n}{j}\sum_{k=0}^{\ell-1}(\omega^j)^k.$$

If j is multiple of ℓ then $(\omega^j)^k=1$ for all k so that $\sum_{k=0}^{\ell-1}(\omega^j)^k=\ell$. If j is not a multiple of ℓ then, using the usual geometric series formula,

$$\sum_{k=0}^{\ell-1}(\omega^j)^k=\frac{1-(\omega^j)^\ell}{1-\omega^j}=0$$

(note that this makes sense since in this case $\omega^j\neq 1$). It follows that

$$\sum_{j=0}^{n}\binom{n}{j}\sum_{k=0}^{\ell-1}(\omega^j)^k=\ell\left(\binom{n}{0}+\binom{n}{\ell}+\binom{n}{2\ell}+\cdots\right)$$

and that

$$\frac{\binom{n}{0}+\binom{n}{\ell}+\binom{n}{2\ell}+\cdots}{2^n}=\frac{1}{\ell}\sum_{k=0}^{\ell-1}\frac{(1+\omega^k)^n}{2^n}=\frac{1}{\ell}\left(1+\sum_{k=1}^{\ell-1}\left(\frac{1+\omega^k}{2}\right)^n\right).$$

Because $(1+\omega^k)/2$, being the average of two distinct complex numbers both on the unit circle, has absolute value less than 1, we have

$$\left(\frac{1+\omega^k}{2}\right)^n\to 0$$

for each $k\in\{1,\ldots,\ell-1\}$, so

$$\frac{\binom{n}{0}+\binom{n}{\ell}+\binom{n}{2\ell}+\cdots}{2^n}\to\frac{1}{\ell},$$

as required.

[10] See `http://math.stackexchange.com/questions/20749/the-hexagonal-property-of-pascals-triangle`, where a nice interpretation in terms of Pascal's triangle is given.

(12) (An exact version of a special case of the last exercise.) Fix $k \geq 0$ and set $n = 4k+2$. Prove that exactly one-quarter of all subsets of $\{1, \ldots, n\}$ have size divisible by 4.

Solution: Using the fact that the sum of every second binomial coefficient, starting from $\binom{4k+2}{0}$, is half the total sum, we get that

$$\binom{4k+2}{0} + \binom{4k+2}{2} + \ldots + \binom{4k+2}{4k} + \binom{4k+2}{4k+2} = 2^{4k+1}.$$

Now applying the identity $\binom{n}{\ell} = \binom{n}{n-\ell}$ to every second term in this sum (starting from the second) we get

$$\binom{4k+2}{0} + \binom{4k+2}{4k} + \ldots + \binom{4k+2}{4k} + \binom{4k+2}{0} = 2^{4k+1},$$

or, since the left-hand side above is an interleaving of $\binom{4k+2}{0} + \binom{4k+2}{4} + \ldots + \binom{4k+2}{4k}$ and $\binom{4k+2}{4k} + \ldots + \binom{4k+2}{4} + \binom{4k+2}{0}$,

$$2\left(\binom{4k+2}{0} + \binom{4k+2}{4} + \ldots + \binom{4k+2}{4k}\right) = 2^{4k+1}$$

or

$$\binom{4k+2}{0} + \binom{4k+2}{4} + \ldots + \binom{4k+2}{4k} = 2^{4k},$$

as required.

(13) In how many ways can one choose a pair of subsets $S, T \subseteq \{1, \ldots, n\}$, subject to the condition that S is a subset of T?

Solution: For element i of $\{1, \ldots, n\}$, we have to decide whether it is in S, in $T \setminus S$ or not in S, so there are 3 options for each i, leading to a total of 3^n.

An alternate approach is to select k, the size of T, then select T, then select $S \subseteq T$, leading to a count of $\sum_{k=0}^{n} \binom{n}{k} 2^k$. The identity

$$\sum_{k=0}^{n} \binom{n}{k} 2^k = 3^n$$

is of course a special case of the binomial theorem.

(14) Let n, p and q be fixed non-negative integers with $p \leq n$ and $q \leq n$. Prove that

$$\sum_{k=0}^{n} \binom{n}{k}\binom{n-k}{p-k}\binom{n-p}{q-k} = \binom{n}{p}\binom{n}{q}.$$

Solution: Let

$$\mathcal{S} = \{(S, T) : S, T \subseteq \{1, \ldots, n\},\ |S| = p,\ |T| = q\}.$$

Evidently

$$|\mathcal{S}| = \binom{n}{p}\binom{n}{q}.$$

But also $\mathcal{S} = \bigcup_{k=0}^{n} \mathcal{S}_k$ where

$$\mathcal{S}_k = \{(S, T) : S, T \subseteq \{1, \ldots, n\},\ |S| = p,\ |T| = q,\ |S \cap T| = k\},$$

and the union is disjoint. Evidently

$$|\mathcal{S}_k| = \binom{n}{k}\binom{n-k}{p-k}\binom{n-p}{q-k}$$

(choose $S \cap T$ first, then $T \setminus S$, then $S \setminus T$), and summing over k gives the identity.

13. Multisets, weak compositions, compositions

Informally, a multiset is a set in which elements are allowed to appear with multiplicity greater than 1. Formally, a *multiset* is a pair $M = (X, a)$ where X is a set and $a = (a_i : i \in X)$ is a sequence of strictly positive integers indexed by elements of X (with a_i representing "how many times" element i occurs in M). Having given the formal definition, we will almost never again use it, instead working with the informal notion of set-with-multiplicities; everything we assert about multisets can easily be proved in the formal setting, but the notation tends to obscure what's going on.

The *groundset* of a multiset $M = (X, a)$ is defined to be the set X.

We have already addressed the question, "in how many different ways can the elements of a multiset M (over groundset $\{1, \ldots, \ell\}$, say, with multiplicities $(a_1, \ldots, a_\ell)$) be arranged in order", the answer being the multinomial coefficient $\binom{n}{a_1,\ldots,a_\ell}$, where $n = \sum_{i=1}^{\ell} a_i$ is the *order* of the multiset. Note that this formula remains true if we extend the definition of a multiset to allow elements of the groundset to appear with multiplicity 0. Since this extension will prove notationally convenient, we make it now and for good. (However, two multisets that differ only in multiplicity-0 elements of their groundsets are not counted as distinct.)

Here's a more subtle question.

Question 13.1. *Let M be a multiset over groundset $\{1, \ldots, \ell\}$ with multiplicities $(a_1, \ldots, a_\ell)$. For $0 \leq k \leq n := \sum_{i=1}^{\ell} a_i$, how many sub-multisets of order k does M have?*

We obtain a sub-multiset of size k by selecting integers $(b_1, \ldots, b_\ell)$ with $0 \leq b_i \leq a_i$ and with $\sum_{i=1}^{\ell} b_i = k$; each such choice of vector $(b_1, \ldots, b_\ell)$ corresponds uniquely to a sub-multiset of order k, and each sub-multiset of order k gives rise to a unique such vector. So the count of sub-multisets of order k is

$$\text{number of solutions to } \textstyle\sum_{i=1}^{\ell} b_i = k, \text{ with } 0 \leq b_i \leq a_i \text{ for each } i.$$

The restrictions on the b_i's imposed by the a_i's make this (in general) a hard sum to evaluate in closed form. In the special case where all the multiplicities are infinite, there is a simple solution in terms of binomial coefficients.

Definition 13.2. *A* weak composition *of the non-negative integer k into ℓ parts is a solution to the equation $b_1 + \ldots + b_\ell = k$ with each b_i being a non-negative integer.*

If M is a multiset over groundset $\{1, \ldots, \ell\}$ with all multiplicities infinite, then for $0 \leq k$ the number of sub-multisets of M of order k is exactly the number of weak compositions of k into ℓ parts.

Proposition 13.3. *There are $\binom{k+\ell-1}{\ell-1} = \binom{k+\ell-1}{k}$ weak compositions of k into ℓ parts when $k > 0$.*

Proof. Consider a row of $k + \ell - 1$ empty boxes. Choose $\ell - 1$ of them to fill. To that choice, we can associate a weak composition of k into ℓ parts: b_1 is the number of empty boxes in the row up to the point where the first filled box is encountered; b_2 is the number of empty boxes

between the first two filled boxes, an so on, up to b_ℓ, which is the number of empty boxes after the last filled box. It's easily verified that the association just described is a bijection from the set of selections of $\ell - 1$ boxes out of $k + \ell - 1$, to the set of weak compositions of k into ℓ parts, and so the number of such weak compositions is indeed $\binom{k+\ell-1}{\ell-1}$ (or equivalently $\binom{k+\ell-1}{k}$). □

This is a re-wording of the well-known "stars-and-bars" argument.

Definition 13.4. *A* composition *of the positive integer k (i.e., $k \geq 1$) into $1 \leq \ell \leq k$ parts is a solution to the equation $b_1 + \ldots + b_\ell = k$ with each b_i being a positive integer.*

It might seem that the empty/filled boxes argument cannot be easily modified to handle the extra restriction that consecutive boxes are not allowed to be filled; but a very simple shifting argument does the trick: the compositions of k into ℓ parts are in bijection with the weak compositions of $k - \ell$ into ℓ parts via the bijection that sends composition $(b_1, \ldots, b_\ell)$ to weak composition $(b_1 - 1, \ldots, b_\ell - 1)$.

Proposition 13.5. *There are $\binom{k-1}{\ell-1}$ compositions of k into ℓ parts when $k > 0$.*

Combining Proposition 13.5 with the binomial theorem we find that any positive integer k has

$$\sum_{\ell=1}^{k} \binom{k-1}{\ell-1} = 2^{k-1}$$

compositions in total. This formula admits a nice inductive proof that (in the inductive step) is combinatorial. The inductive step involves showing (for any integer $k > 1$) that there are twice as many compositions of k as there are of $k - 1$, which could be achieved by showing that the set of compositions of k decomposes into two sets, each of which are in bijection with the set of compositions of $k - 1$. The natural decomposition works: the compositions of k in which $b_1 = 1$ biject to the compositions of $k - 1$ via

$$(b_1, b_2, \ldots, b_\ell) \mapsto (b_2, \ldots, b_\ell),$$

while the compositions of k in which $b_1 \geq 2$ biject to the compositions of $k - 1$ via

$$(b_1, b_2, \ldots, b_\ell) \mapsto (b_1 - 1, b_2, \ldots, b_\ell).$$

It may be interesting to note that this shows that half of all compositions of an integer $k > 1$ begin with a 1.

One final note on compositions.

Question 13.6. *For $k \geq 0$, how many solutions are there to $b_1 + \ldots + b_\ell \leq k$, with each $b_i \geq 0$ and integral?*

An obvious way to answer this (assuming that $\ell > 0$) is to sum, over all $0 \leq m \leq k$, the number of weak compositions of m into ℓ parts; this gives a count of

$$\sum_{m=0}^{k} \binom{m+\ell-1}{\ell-1} = \sum_{m=0}^{k} \binom{m+\ell-1}{m}.$$

Alternately, we may introduce a "dummy" variable $b_{\ell+1}$ to take up the slack between $b_1 + \ldots + b_\ell$ and k; solutions to $b_1 + \ldots + b_\ell \leq k$, with each $b_i \geq 0$ and integral, are in bijection

with solutions to $b_1 + \ldots + b_\ell + b_{\ell+1} = k$, with each $b_i \geq 0$ and integral, that is, to weak compositions of k into $\ell + 1$ parts, giving a count of $\binom{k+\ell}{\ell} = \binom{k+\ell}{k}$. This leads to the identity

$$\sum_{m=0}^{k} \binom{m+\ell-1}{m} = \binom{k+\ell}{k}$$

(for $\ell > 0$), an instance of the parallel summation identity.

14. Set partitions

We have addressed the question of decomposing a set into parts of given sizes, with the order of the parts mattering. Here we consider the more subtle question of what happens when the order of the parts does not matter and (more crucially for the problem to become more subtle) we do not specify the sizes of the parts. In other words, we will count set partitions.

Question 14.1. *Fix $n \geq 0$ and $k \in \mathbb{Z}$. In how many ways can a set of size n be partitioned in k non-empty parts, if we care neither about the order in which the parts are presented, nor about the order of the elements within each part? In other words, how many set partitions with k parts does an n-element set have? Equivalently, how many different equivalence relations are there on a set of size n in which there are exactly k equivalence classes?*

(To remind, an equivalence class is always understood to be non-empty.)

Some notation is in order. We will refer to the parts of the partition as the *blocks*. We will visually represent a partition into blocks by first listing the elements in one of the blocks (without commas or spaces between the named elements[11]), then putting a vertical bar, then listing the elements of another of the blocks, and so on (with no vertical bar at the very end). If there is a total order on the underlying set, then we adopt the convention of listing the elements of each block in increasing order, and listing the blocks in order of increasing least element; so, for example, we will typically present the partition of $\{1, 2, 3, 4, 5\}$ into the non-empty parts $\{1, 3, 4\}$ and $\{2, 5\}$ as $134|25$, though we could just as easily have decided to present this as $413|25$ or $52|431$ (or in many other ways).

Here's a complete list of the partitions of $\{1, 2, 3, 4\}$. The one partition into one part is 1234; the seven partitions into two parts are

$$123|4,\ 124|3,\ 134|2,\ 1|234,\ 12|34,\ 13|24,\ 14|23;$$

the six partitions into three parts are

$$12|3|4,\ 13|2|4,\ 14|2|3,\ 1|23|4,\ 1|24|3,\ 1|2|34;$$

and the one partition into four blocks is $1|2|3|4$.

The number of partitions of a set of size n into k non-empty blocks is denoted $S(n, k)$ or $\left\{ {n \atop k} \right\}$, and is called a *Stirling number of the second kind*, after James Stirling, who introduced the numbers in 1730.[12] We present some small values of $\left\{ {n \atop k} \right\}$ in a table below (blank entries

[11] That is, the block $\{1, 4, 6\}$ will be simply written as 146. This works well as long as the elements belong to $\{0, 1, \ldots, 9\}$. Of course, with greater numbers, we need a less laconic notation.

[12] J. Stirling, *Methodus differentialis, sive tractatus de summatione et interpolatione serierum infinitarum*, London, 1730.

are 0).

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
$n=0$	1						
$n=1$		1					
$n=2$		1	1				
$n=3$		1	3	1			
$n=4$		1	7	6	1		
$n=5$		1	15	25	10	1	
$n=6$		1	31	90	65	15	1

The boundary values are reasoned as follows: for any $n \geq 0$ there are no equivalence relations on a set of size n with more than n equivalence classes (and none with a negative number of equivalence classes); for $n \geq 1$ there are none with exactly zero equivalence classes; and there is one equivalence relation on the empty set (that has exactly zero equivalence classes), namely the empty relation.

There is no compact expression for $\left\{ {n \atop k} \right\}$ as there is for $\binom{n}{k}$, but there is a recurrence relation similar to Pascal's identity that allows the easy calculation of $\left\{ {n \atop k} \right\}$ for large values of n and k.

Proposition 14.2. *With* $\left\{ {n \atop 0} \right\} = 0$ *for* $n \geq 1$, $\left\{ {0 \atop k} \right\} = 0$ *for* $k \geq 1$, *and* $\left\{ {0 \atop 0} \right\} = 1$, *for* $n, k \geq 1$ *we have*

$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k \left\{ {n-1 \atop k} \right\}.$$

Proof. We just give an informal proof, and let the reader provide a formal (bijective) argument if she wishes. We construct a partition of $\{1, \ldots, n\}$ into k non-empty blocks by either partitioning $\{1, \ldots, n-1\}$ into $k-1$ non-empty blocks (in $\left\{ {n-1 \atop k-1} \right\}$ ways) and then adding n as a singleton kth block, or by partitioning $\{1, \ldots, n-1\}$ into k non-empty blocks (in $\left\{ {n-1 \atop k} \right\}$ ways) and then adding element n to one of these k blocks (giving $k \left\{ {n-1 \atop k} \right\}$ options in this second case). □

The triangle of integers

$$\begin{array}{c} 1 \\ 1\ 1 \\ 1\ 3\ 1 \\ 1\ 7\ 6\ 1 \\ 1\ 15\ 25\ 10\ 1 \\ \cdots \end{array}$$

crops up a lot in applications that seem to have nothing to do with partitioning, and it is always tempting when this happens to believe that one is seeing the Stirling numbers of the second kind. The benefit of a recurrence relation of the kind given by Proposition 14.2 is that it gives an easy tool for establishing such a conjecture: if an unknown double sequence $(a(n,k))_{n,k \geq 0}$ satisfies $a(n,0) = 0$ for $n \geq 1$, $a(0,k) = 0$ for $k \geq 1$, $a(0,0) = 1$, and, for $n, k \geq 1$, $a(n,k) = a(n-1,k-1) + ka(n-1,k)$, then it is an easy induction to conclude that $a(n,k) = \left\{ {n \atop k} \right\}$ for all $n, k \geq 0$.

Let us have an example, from linear algebra.

The standard basis for the space of real polynomials on a single variable is $\{1, x, x^2, x^3, \ldots\}$. Another basis for this space is the collection of falling powers: $\{1, x, x(x-1), x(x-1)(x-$

$2), \ldots\} = \{(x)_0, (x)_1, (x)_2, \ldots\}$. For each $n \geq 0$, the polynomial x^n can be represented as a linear combination of polynomials of the form $(x)_k$. For example, we have

$$\begin{aligned} 1 &= 1 = (x)_0, \\ x &= x = (x)_1, \\ x^2 &= x(x-1) + x = (x)_2 + (x)_1, \\ x^3 &= x(x-1)(x-2) + 3x(x-1) + x = (x)_3 + 3(x)_2 + (x)_1, \\ x^4 &= (x)_4 + 6(x)_3 + 7(x)_2 + (x)_1. \end{aligned}$$

In general, we have

$$x^n = \sum_{k=0}^{\infty} a(n,k)(x)_k \tag{10}$$

for some double sequence $(a(n,k))_{n,k\geq 0}$, and it very much looks like $a(n,k) = \left\{ {n \atop k} \right\}$ for all $n, k \geq 0$. (If you are concerned about the left-hand side being a polynomial while the right-hand side is a power series, remember that the linear algebra tells us that for each n all but finitely many of the $a(n,k)$ will be 0).

Claim 14.3. *For all $n \geq 0$ we have*

$$x^n = \sum_{k=0}^{\infty} \left\{ {n \atop k} \right\} (x)_k. \tag{11}$$

Proof. With $a(n,k)$ as defined in (10) we certainly have $a(n,0) = 0$ for $n \geq 1$ (x^n has no constant term, and so the coefficient of 1 in the expansion on the right-hand side must be 0), $a(0,k) = 0$ for $k \geq 1$, and $a(0,0) = 1$ (these last two following immediately from the identity $1 = 1$). It remains to show that for $n, k \geq 1$, we have $a(n,k) = a(n-1,k-1) + ka(n-1,k)$. We prove that this is true for all $k \geq 1$ for each fixed $n \geq 1$, with the case $n = 1$ evident.

For $n \geq 2$

$$\begin{aligned}
x^n &= x \times x^{n-1} \\
&= x \sum_{k=0}^{\infty} a(n-1,k)(x)_k \\
&= \sum_{k=0}^{\infty} a(n-1,k)x(x)_k \\
&= \sum_{k=0}^{\infty} a(n-1,k)((x-k)+k)(x)_k \\
&= \sum_{k=0}^{\infty} a(n-1,k)(x-k)(x)_k + \sum_{k=0}^{\infty} ka(n-1,k)(x)_k \\
&= \sum_{k=0}^{\infty} a(n-1,k)(x)_{k+1} + \sum_{k=0}^{\infty} ka(n-1,k)(x)_k \\
&= \sum_{k=1}^{\infty} a(n-1,k-1)(x)_k + \sum_{k=1}^{\infty} ka(n-1,k)(x)_k \\
&= \sum_{k=1}^{\infty} \left(a(n-1,k-1) + ka(n-1,k)\right)(x)_k.
\end{aligned}$$

But also $x^n = \sum_{k=0}^{\infty} a(n,k)(x)_k = \sum_{k=1}^{\infty} a(n,k)(x)_k$. By uniqueness of representation of elements of a vector space as linear combinations of basis vectors, we conclude that $a(n,k) = a(n-1,k-1) + ka(n-1,k)$ for each $k \geq 1$. □

This was an easy, but somewhat awkward, proof. Here's a pleasingly short combinatorial proof of the identity

$$x^n = \sum_{k=1}^{\infty} \left\{ n \atop k \right\} (x)_k$$

for $n \geq 1$, that bypasses all of the awkwardness. Fix an integer $x \geq 1$ (by the polynomial principle, this is not a restriction). The left-hand side above counts the number of words of length n over alphabet $\{1, \ldots, x\}$, by choosing the word one letter at a time. The right-hand side counts the same thing. We first decide how many different letters appear in the word; this is our index k. We then decide on a partition of the n spaces for letters into k non-empty blocks; each block will be exactly the collection of spaces filled by a particular letter. There are $\left\{ n \atop k \right\}$ ways to do this (note that if we had decided to use more than n different letters, this term becomes 0). Finally we go through the blocks one-by-one, in some canonical order, and decide which letters appear in the spaces of that block; there are $x(x-1)\ldots(x-k+1)$ ways to decide this (note that if we had decided to use more than x different letters, this term becomes 0, since it has 0 as a term in the repeated product).

There's a more colorful version of this argument, involving beer. In how many ways can n people sit in groups at a bar, each group with a pitcher of a different brand of beer, if the bar has x brands of beer available? One way this can be achieved is for each person to choose a beer they like; the groupings are then determined by which people want the same beer. Thought of this way, the counting problem has x^n solutions. Another strategy is for the n

people together to decide how many groups they will form (k), then form k groups ($\left\{ {n \atop k} \right\}$), then one after the other each group chooses an as-yet-unchosen beer brand ($(x)_k = x(x-1)\dots(x-k+1)$). Thought of this way, the counting problem has $\sum_{k \geq 1} \left\{ {n \atop k} \right\} (x)_k$ solutions.

The combinatorial proof gives information that is not easily extracted from the inductive proof. For example, it is hardly obvious that in the expansion $x^n = \sum_{k=0}^{\infty} a(n,k)(x)_k$, the coefficients $a(n,k)$ are non-negative. From the combinatorial proof, this is immediate.

The proof of Claim 14.3 suggests the following.

Claim 14.4. *Fix $n, k \geq 0$. The number of surjective functions $f : \{1, \dots, n\} \to \{1, \dots, k\}$ is $k! \left\{ {n \atop k} \right\}$.*

Proof. We construct a surjective function by first partitioning the domain $\{1, \dots, n\}$ into k non-empty blocks; these will be the pre-images of each of the elements $1, \dots, k$. There are $\left\{ {n \atop k} \right\}$ such partitions. We finish the construction of the surjection by going through the blocks one-by-one, in some canonical order, and deciding which element that block is the preimage of. There are $k!$ ways to make this designation. □

Here are two other situations in which the Stirling numbers of the second kind arise unexpectedly. We leave the proofs as exercises.

Claim 14.5. *For $n \geq 1$, the nth derivative with respect to x of the function $f(x) = e^{e^x}$ is*

$$f^{(n)}(x) = f(x) \sum_{k=0}^{\infty} \left\{ {n \atop k} \right\} e^{kx}.$$

Claim 14.6. *Let $f(x)$ be an infinitely differentiable function. Define a sequence of functions $f_n(x)$ recursively by*

$$f_n(x) = \begin{cases} f(x) & \text{if } n = 0 \\ x \frac{d}{dx} f_{n-1}(x) & \text{if } n \geq 1 \end{cases}$$

(so $f_0(x) = f(x)$, $f_1(x) = xf'(x)$, $f_2(x) = x^2 f''(x) + xf'(x)$ and $f_3(x) = x^3 f'''(x) + 3x^2 f''(x) + xf'(x)$). For $n \geq 1$, we have

$$f_n(x) = \sum_{k=1}^{n} \left\{ {n \atop k} \right\} x^k f^{(k)}(x)$$

where $f^{(k)}(x)$ indicates kth derivative with respect to x.

A comment is in order here. The *Weyl algebra* is the algebra on two symbols, x and D, satisfying the single relation $Dx = xD + 1$. One can realize this algebra on the space of single-variable polynomials by thinking of x as "multiply by x" and D as "differentiate with respect to x"; the relation $Dx = xD + 1$ is a manifestation of the identity $(xf(x))' = xf'(x) + f(x)$. The Weyl algebra has an application in the theory of the quantum harmonic oscillator, where x is a creation operator (increasing the degree of a polynomial by 1) and D is an annihilation operator (decreasing the degree by 1). Claim 14.6 can be rephrased as the following identity between words in the Weyl algebra:

$$(xD)^n = \sum_{k=1}^{n} \left\{ {n \atop k} \right\} x^k D^k \qquad \text{for all } n > 0.$$

This identity was explored first by Scherk[13]. The expression on the right-hand side above is referred to as to *normal order* of the word $(xD)^n$ (the expansion in terms of the words $xD, x^2D^2, \ldots$). From a computational viewpoint, it is easier to understand the effect of operating by a word on a function if the word is presented in normal order.

The recurrence given in Proposition 14.2 may be thought of as "horizontal": it expresses values of $\left\{ {n \atop k} \right\}$ in terms of values along an earlier row of the two-dimensional Stirling table. There is also a "vertical" recurrence, that expresses values of $\left\{ {n \atop k} \right\}$ in terms of values along an earlier column of the table. The proof is left as an exercise.

Proposition 14.7. *For all integers $n \geq 0$ and $k \geq 1$, we have*

$$\left\{ {n+1 \atop k} \right\} = \sum_{i=0}^{n} \binom{n}{i} \left\{ {n-i \atop k-1} \right\} = \sum_{i=0}^{n} \binom{n}{i} \left\{ {i \atop k-1} \right\}.$$

A Stirling number of the second kind is the number of partitions of a set into a fixed number of non-empty blocks. What happens if we don't care about the number of blocks? We define the nth *Bell number* $B(n)$ to be the number of partitions of a set of size n into some number of non-empty blocks. Evidently

$$B(n) = \sum_{k=0}^{n} \left\{ {n \atop k} \right\}.$$

Combining this with Proposition 14.7, we get the following recurrence relation for $B(n)$: $B(0) = 1$, and for $n \geq 0$,

$$B(n+1) = \sum_{k=0}^{n} \binom{n}{k} B(k).$$

(The natural combinatorial proof of Proposition 14.7 could also be adapted to give a direct combinatorial proof of this identity.) The sequence $(B(n))_{n \geq 0}$ of Bell numbers begins $(1, 2, 5, 15, 52, 203, 877, \ldots)$. Here are two occurrences of the Bell numbers. The first is nothing more than the definition; the second might be quite surprising, and we will leave a proof of it to later.

(1) Let N be a square-free number with n prime factors (i.e., a number of the form $N = p_1 p_2 \ldots p_n$, where the p_i's are distinct prime numbers). A *factorization* of N is an expression of the form $N = f_1 f_2 \ldots f_k$ with $f_1 \geq f_2 \geq \ldots \geq f_k > 1$ and each f_k an integer; notice that by unique factorization, we in fact have $f_1 > f_2 > \ldots > f_k$ (since N is square-free). The number of different factorizations of N is $B(n)$.

(2) Let X be a Poisson random variable with parameter 1; that is, X takes value ℓ with probability $e^{-1}/\ell!$ for $\ell = 0, 1, \ldots$, and with probability 0 for all other ℓ. The nth *moment* of X is $E(X^n)$, or $\sum_{\ell=0}^{\infty} \ell^n \Pr(X = \ell)$. For all $n \geq 0$, the nth moment of X is $B(n)$.

15. Some problems

(1) Prove Claim 14.5.

[13]H. Scherk, *De evolvenda functione $(yd.yd.yd \ldots ydX)/dx^n$ disquisitiones nonnullae analyticae*, Ph.D. thesis, University of Berlin, 1823.

(2) Prove Claim 14.6.

Solution: We begin with a prosaic, inductive proof. It is an easy induction argument that for $n \geq 1$ there are some constants $a(n,k)$, $k \geq 1$ such that for all infinitely differentiable functions $f(x)$

$$(xD)^n f(x) = \sum_{k \geq 1} a(n,k) x^k D^k f(x).$$

(and in particular $a(n,k) = 0$ for $n > k$, though we do not need that at the moment). It is also not hard to establish that these constants are uniquely determined. Using $(xD)^0 f(x) = x^0 D^0 f(x)$, we find that if we extend the definition of the $a(n,k)$ to all $n, k \geq 0$ via

$$(xD)^n f(x) = \sum_{k \geq 0} a(n,k) x^k D^k f(x)$$

for $n \geq 0$, we have that $a(0,0) = 1$ and $a(n,0) = a(0,k) = 0$ for $n, k > 0$. Also, for any $n \geq 1$, we have

$$\begin{aligned}
(xD)^n f(x) &= xD(xD)^{n-1} f(x) \\
&= xD \left(\sum_{k \geq 0} a(n-1,k) x^k D^k f(x) \right) \\
&= \sum_{k \geq 0} a(n-1,k) xDx^k D^k f(x) \\
&= \sum_{k \geq 0} a(n-1,k) x^{k+1} D^{k+1} f(x) + \sum_{k \geq 0} k a(n-1,k) x^k D^k f(x) \\
&= \sum_{k \geq 1} a(n-1,k-1) x^k D^k f(x) + \sum_{k \geq 1} k a(n-1,k) x^k D^k f(x) \\
&= \sum_{k \geq 1} \left(a(n-1,k-1) + k a(n-1,k) \right) x^k D^k f(x)
\end{aligned}$$

(here, the fourth equality sign relied on the simple fact that $Dx^k = x^k D + k x^{k-1}$). By uniqueness of the $a(n,k)$ we obtain $a(n,k) = a(n-1,k-1) + k a(n-1,k)$ for $n, k \geq 1$, so $a(n,k) = \left\{ {n \atop k} \right\}$.

There is also a slicker argument, that again starts with the fact that for $n \geq 1$ there are some constants $a(n,k)$, $1 \leq k \leq n$, such that for all infinitely differentiable functions $f(x)$

$$(xD)^n f(x) = \sum_{k=1} a(n,k) x^k D^k f(x).$$

Plugging in the function $f(x) = x^m$, where m is an arbitrary real, yields the identity

$$m^n x^m = a(n,1) m x^m + a(n,2) m(m-1) x^m + \ldots + a(n,n) (m)_n x^m,$$

and dividing through by x^m yields

$$m^n = a(n,1) m + a(n,2) m(m-1) + \ldots + a(n,n) (m)_n.$$

This identity is true for all real m; in other words, the $a(n,k)$ are exactly the coefficients that occur in the expansion of the polynomial m^n (in real variable m) in terms of the basis $\{1, m, m(m-1), \ldots, (m)_n\}$ (of the vector space of all polynomials in m

of degree $\leq n$). We have already established (both inductively and combinatorially) that these coefficients are the Stirling numbers of the second kind.

(3) Prove Proposition 14.7.

(4) For $n, k \geq 0$, let $\mathrm{Surj}(n,k)$ denote the number of surjective functions from $\{1,\ldots,n\}$ to $\{1,\ldots,k\}$. Give, **and justify combinatorially**, a recurrence relation for $\mathrm{Surj}(n,k)$ of the form

$$\mathrm{Surj}(n,k) = f(n,k)\,\mathrm{Surj}(n-1,k-1) + g(n,k)\,\mathrm{Surj}(n-1,k)$$

valid for $n,k \geq 1$ (analogous to the relation $\left\{{n \atop k}\right\} = \left\{{n-1 \atop k-1}\right\} + k\left\{{n-1 \atop k}\right\}$), and give the initial conditions (the correct values of $\mathrm{Surj}(n,k)$ when at least one of n,k is 0). Then use your recurrence to draw up the "Pascal triangle" for surjections, that is, the table $[\mathrm{Surj}(n,k)]_{n,k=1}^{6}$. Using these values, locate the surjection numbers on the Online Encyclopedia of Integer Sequences, oeis.org. (Each sequence there is identified by an "A" followed by a short string of numbers; tell me the string!)

Solution: The required identity is

$$\mathrm{Surj}(n,k) = k\,\mathrm{Surj}(n-1,k-1) + k\,\mathrm{Surj}(n-1,k).$$

Indeed, the surjections from $[n]$ to $[k]$ decompose into two classes:

- those in which the image of n only has n in its preimage, and
- those in which the preimage of the image of n includes more than just n.

The surjections in which the image of n only has n in its preimage decompose according to the choice of image of n. For each (of the k possible) choice of image of n, there are $\mathrm{Surj}(n-1,k-1)$ such surjections. The surjections in which the preimage of the image of n includes more than just n decompose again according to the choice of image of n. For each (of the k possible) choice of image of n, there are $\mathrm{Surj}(n-1,k)$ such surjections. The claimed recurrence follows.

The boundary conditions are evidently $\mathrm{Surj}(0,0) = 1$, $\mathrm{Surj}(n,0) = 0 = \mathrm{Surj}(0,k)$ for $n,k > 0$.

The table of values required is

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
$n=0$	1	0	0	0	0	0	0
$n=1$	0	1	0	0	0	0	0
$n=2$	0	1	2	0	0	0	0
$n=3$	0	1	6	6	0	0	0
$n=4$	0	1	14	36	24	0	0
$n=5$	0	1	30	150	240	120	0
$n=6$	0	1	62	540	1560	1800	720

This is either A131689 (with the 0 column) or A019538 (without).

16. Inclusion-exclusion

The Stirling numbers of the second kind (and therefore the Bell numbers) admit a fairly simple summation formula, but to understand it we need to take a digression into the principle of inclusion-exclusion.

The addition principle says that if A_1 and A_2 are disjoint, then $|A_1 \cup A_2| = |A_1| + |A_2|$. This formula fails if A_1 and A_2 have elements in common, having to be replaced by familiar

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|,$$

the point being that each element in $A_1 \cap A_2$ gets counted twice in the sum $|A_1| + |A_2|$, and so needs to get subtracted off once to correct the count. An examination of a three-set Venn diagram easily yields the following extension

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|.$$

This suggests a general formula for $|A_1 \cup A_2 \cup \ldots \cup A_n|$ in terms of the sizes of intersections of subsets of the A_i, a formula we refer to as the *principle of inclusion-exclusion*.

Theorem 16.1 (Inclusion-Exclusion)**.** *For finite sets $A_1, \ldots, A_n$, we have*

$$\begin{aligned} |A_1 \cup \ldots \cup A_n| \;=\; & |A_1| + \ldots + |A_n| \\ & -|A_1 \cap A_2| - |A_1 \cap A_3| - \ldots - |A_{n-1} \cap A_n| \\ & +|A_1 \cap A_2 \cap A_3| + \ldots \\ & \ldots \\ & +(-1)^{n-1}|A_1 \cap \ldots \cap A_n|. \end{aligned}$$

More succinctly,

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{k=1}^{n}(-1)^{k-1} \sum_{I \subseteq \{1,\ldots,n\},\ |I|=k} \left|\bigcap_{i \in I} A_i\right| = \sum_{I \subseteq \{1,\ldots,n\},\ I \neq \emptyset} (-1)^{|I|-1} \left|\bigcap_{i \in I} A_i\right|.$$

The value of the principle of inclusion-exclusion lies in the fact that it is often easier to compute sizes of intersections than sizes of unions; rather than discuss this further here, we will let several examples speak for themselves.

If all the A_i's are subsets of some ground-set U, and complementation of sets is taken to be inside U (that is, for every subset X of U, we define the complement of X to be the set $U \subset X$; we denote this complement by X^c), then an equivalent form of inclusion-exclusion is

$$\left|\left(\bigcup_{i=1}^{n} A_i\right)^c\right| = \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|} \left|\bigcap_{i \in I} A_i\right|, \tag{12}$$

where the empty intersection (i.e., the intersection $\bigcap_{i \in \varnothing} A_i$) is taken to be U. Here's one way to view this formula: think of A_i as the set of elements in U that possess some property, called property i. The left-hand side of (12) is the number of elements of U that possess *none* of the properties 1 through n; this is re-expressed on the right-hand side as an alternating sum of terms which count the number of elements that possess *at least* a certain collection of the properties (but may possess more; the key to the utility of (12) is that we do not care how many more, if even any). In the following, we shall often refer to U as the *universe* in this situation.

In many applications, we will see that the size of $\bigcap_{i \in I} A_i$ depends only $|I|$, in which case (12) becomes

$$\left|\left(\bigcup_{i=1}^{n} A_i\right)^c\right| = \sum_{k=0}^{n}(-1)^k \binom{n}{k} f(k),$$

where $f(k)$ is the common size of $\bigcap_{i \in I} A_i$ over all I of size k.

We shall prove Theorem 16.1 in two different ways.

Proof. (First proof of Theorem 16.1, inclusion-exclusion) The set $\bigcup_{i=1}^{n} A_i$ decomposes as $\bigcup_{J \subseteq \{1,\ldots,n\},\ J \neq \emptyset} A_J$, with A_J being the set of elements that are in each of A_j for $j \in J$, but

not in any of $A_{j'}$ for $j' \notin J$. We show that for each J and each $x \in A_J$, the element x is counted the same number of times on the left- and right-hand sides of the inclusion-exclusion identity. Evidently x is counted once on the left-hand side. On the right-hand side, it is counted $(-1)^{|I|}$ times for each non-empty subset I of J, so is counted in all

$$\sum_{k=1}^{|J|}(-1)^{k-1}\binom{|J|}{k} = -\left(-1+\sum_{k=0}^{|J|}(-1)^{k}\binom{|J|}{k}\right) = 1$$

times, the second equality using (8). □

Before we give the second proof of Theorem 16.1, let us state a simple lemma:

Lemma 16.2. *Let $a_1, a_2, \ldots, a_n$ be n numbers. Then,*

$$\prod_{i\in[n]}(1+a_i) = \sum_{I\subseteq[n]}\prod_{i\in I}a_i \tag{13}$$

and

$$\prod_{i\in[n]}(1-a_i) = \sum_{I\subseteq[n]}(-1)^{|I|}\prod_{i\in I}a_i. \tag{14}$$

Proof. (Proof of Lemma 16.2) Multiplying out the product

$$\prod_{i\in[n]}(1+a_i) = (1+a_1)(1+a_2)\cdots(1+a_n),$$

we obtain a sum of 2^n addends. Each addend is a product of some of the a_i; each possible product appears exactly once as an addend. Thus, the addends are in 1-to-1 correspondence with the subsets I of $[n]$. More precisely, to each subset I of $[n]$ corresponds the addend $\prod_{i\in I}a_i$. Hence, the sum is $\sum_{I\subseteq[n]}\prod_{i\in I}a_i$. This proves (13). (Alternatively, (13) can also be proven by induction on n.)

The equality (14) follows by applying (13) to $-a_i$ instead of a_i. □

Proof. (Second proof of Theorem 16.1, inclusion-exclusion) We introduce the notation $[\mathcal{S}]$ for the *truth value* of any statement $\mathcal{S}$; this is simply the integer 1 if $\mathcal{S}$ is true, and 0 if $\mathcal{S}$ is false. These truth values follow some laws; most importantly, we have

$$[\text{not } \mathcal{S}] = 1 - [\mathcal{S}] \qquad \text{for any statement } \mathcal{S},$$

and we have

$$[\mathcal{S}_1 \wedge \mathcal{S}_2 \wedge \cdots \wedge \mathcal{S}_h] = [\mathcal{S}_1][\mathcal{S}_2]\cdots[\mathcal{S}_h] \qquad \text{for any } h \text{ statements } \mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_h.$$

More generally, if I is a finite set, and $\mathcal{S}_i$ is a statement for each $i \in I$, then

$$[\mathcal{S}_i \text{ for each } i \in I] = \prod_{i\in I}[\mathcal{S}_i]. \tag{15}$$

Now, if P is any subset of U, then

$$\sum_{x\in U}[x \in P] = |P| \tag{16}$$

(because the sum $\sum_{x\in U}[x\in P]$ has exactly $|P|$ addends equal to 1 (one for each $x\in P$), and all its remaining addends are 0). Applying this to $P=\left(\bigcup_{i=1}^{n}A_i\right)^c$, we obtain

$$\sum_{x\in U}\left[x\in\left(\bigcup_{i=1}^{n}A_i\right)^c\right]=\left|\left(\bigcup_{i=1}^{n}A_i\right)^c\right|. \tag{17}$$

Now, for any $x\in U$ and $i\in I$, we have

$$\left[x\in\bigcap_{i\in I}A_i\right]=[x\in A_i \text{ for each } i\in I]=\prod_{i\in I}[x\in A_i] \tag{18}$$

(by (15)). Every $x\in U$ satisfies

$$\left[x\in\underbrace{\left(\bigcup_{i=1}^{n}A_i\right)^c}_{=\bigcap_{i=1}^{n}(A_i)^c}\right]=\left[x\in\bigcap_{i=1}^{n}(A_i)^c\right]=[x\in(A_i)^c \text{ for each } i\in[n]]=\prod_{i\in[n]}\underbrace{[x\in(A_i)^c]}_{\substack{=[\text{not } x\in A_i]\\=1-[x\in A_i]}}$$

$$\overset{\text{(by (15))}}{=}\prod_{i\in[n]}(1-[x\in A_i])=\sum_{I\subseteq[n]}(-1)^{|I|}\underbrace{\prod_{i\in I}[x\in A_i]}_{\substack{=\left[x\in\bigcap_{i\in I}A_i\right]\\ \text{(by (18))}}}\qquad\text{(by (14))}$$

$$=\sum_{I\subseteq[n]}(-1)^{|I|}\left[x\in\bigcap_{i\in I}A_i\right].$$

Summing this equality over all $x\in U$, we find

$$\sum_{x\in U}\left[x\in\left(\bigcup_{i=1}^{n}A_i\right)^c\right]=\sum_{x\in U}\sum_{I\subseteq[n]}(-1)^{|I|}\left[x\in\bigcap_{i\in I}A_i\right]=\sum_{I\subseteq[n]}(-1)^{|I|}\underbrace{\sum_{x\in U}\left[x\in\bigcap_{i\in I}A_i\right]}_{\substack{=\left|\bigcap_{i\in I}A_i\right|\\ \text{(by (16))}}}$$

$$=\sum_{I\subseteq[n]}(-1)^{|I|}\left|\bigcap_{i\in I}A_i\right|.$$

Comparing this with (17), we find $\left|\left(\bigcup_{i=1}^{n}A_i\right)^c\right|=\sum_{I\subseteq[n]}(-1)^{|I|}\left|\bigcap_{i\in I}A_i\right|$, which is precisely (12). Thus Theorem 16.1 is proven. □

We give four quick, classic applications of inclusion-exclusion.

Derangements. A *derangement* of $[n]$ is a bijection $f:[n]\to[n]$ with no fixed points, that is with no i such that $f(i)=i$. Equivalently, given n objects in a row, a derangement is a rearrangement of the n objects in which no object gets returned to its original position. Writing D_n for the number of derangements of $[n]$, we have $(D_n)_{n\geq 1}=(0,1,2,9,44,265,\ldots)$.

It's much easier to count bijections that fix (at least) a certain set of fixed points, than those that don't fix some points; this suggests an inclusion-exclusion approach to counting derangements. For $1\leq i\leq n$, let A_i be the set of bijections from $[n]$ to $[n]$ which fix element

i (perhaps among others). We seek the complement of $A_1 \cup \ldots \cup A_n$, in a universe of size $n!$ (the universe being all bijections from $[n]$ to $[n]$). For each $I \subseteq [n]$ we have

$$\left| \bigcap_{i \in I} A_i \right| = (n - |I|)!$$

(we have no choice for the images of $i \in I$ under f, but no restriction on what happens off I other than that f restricted to $[n] \setminus I$ map bijectively to $[n] \setminus I$). It follows from (12) that

$$\begin{aligned} D_n &= \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|}(n - |I|)! \\ &= \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)! \\ &= n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}. \end{aligned}$$

This says that $D_n/n!$, the probability that a uniformly chosen bijection is a derangement, is equal to the sum of the first $n+1$ terms of the Taylor series expansion of e^x (about 0) at $x = -1$, and so in particular that in the limit as n goes to infinity, the probability that a uniformly chosen bijection is a derangement tends to $1/e$.

Euler's φ function. For any integer $n \geq 1$, let $\varphi(n)$ denote the number of numbers in the range $1, \ldots, n$ that are relatively prime to n (that is, have no factors in common with n other than 1). Inclusion-exclusion gives a nice formula for calculating $\varphi(n)$ in terms of the prime factors of n.

Let the prime factorization of n be $n = \prod_{j=1}^{k} p_j^{\alpha_j}$ with the p_j's distinct primes, and the α_j's all at least 1. (If $n = 1$, then this is an empty product, so $k = 0$.) With universe $[n] = \{1, \ldots, n\}$, let A_i be all the numbers that are divisible by p_i; evidently $\varphi(n) = |(A_1 \cup \ldots \cup A_k)^c|$. For each $I \subseteq [k]$ we have

$$\left| \bigcap_{i \in I} A_i \right| = \frac{n}{\prod_{i \in I} p_i}$$

(indeed, the elements of $\bigcap_{i \in I} A_i$ are precisely the elements of $[n]$ divisible by the p_i for all $i \in I$; in other words, they are the multiples of $\prod_{i \in I} p_i$, and clearly the number of these multiples is $\frac{n}{\prod_{i \in I} p_i}$). It follows from (12) that

$$\begin{aligned} \varphi(n) &= \sum_{I \subseteq \{1,\ldots,k\}} (-1)^{|I|} \frac{n}{\prod_{i \in I} p_i} \\ &= n \prod_{j=1}^{k} \left(1 - \frac{1}{p_j}\right). \end{aligned}$$

From this it is easy to deduce a fact that is far from obvious from the definition, namely that φ is a multiplicative function: if m and n are relatively prime, then $\varphi(mn) = \varphi(m)\varphi(n)$.

A summation formula for the Stirling numbers of the second kind. Let $\mathrm{Surj}(n,k)$ be the number of surjective functions from $[n]$ to $[k]$. For each $1 \leq j \leq k$, let A_i be the set of functions from $[n]$ to $[k]$ that do not have element j in their range. With the universe being all k^n functions from $[n]$ to $[k]$, we have $\mathrm{Surj}(n,k) = |(A_1 \cup \ldots \cup A_k)^c|$. For $I \subseteq [k]$, we have

$$\left| \bigcap_{i \in I} A_i \right| = (k - |I|)^n$$

(each of the elements of $[n]$ can go to any of $k - |I|$ images; notice that we only care that *at least* the elements of I are missed, not that *exactly* the elements of I are missed; this is what makes estimating the size of $\bigcap_{i \in I} A_i$ easy). By inclusion-exclusion we have

$$\begin{aligned} \mathrm{Surj}(n,k) &= \sum_{I \subseteq \{1,\ldots,k\}} (-1)^{|I|} (k - |I|)^n \\ &= \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^n. \end{aligned}$$

Since we already know (from Claim 14.4) that the number of surjections from $[n]$ to $[k]$ is $k!\left\{ {n \atop k} \right\}$, we get the following summation formula for the Stirling numbers of the second kind:

$$\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^n = \frac{1}{k!} \sum_{r=0}^{k} (-1)^{k-r} \binom{k}{r} r^n. \tag{19}$$

(Here, in the last equality sign, we substituted $k - r$ for j and used the symmetry relation $\binom{k}{k-r} = \binom{k}{r}$.)

Partitions of a set into blocks with at least two elements each. Let $F_k(n)$ be the number of partitions of $[n]$ into k blocks, each containing at least two elements. Then, we have the following formula for $F_k(n)$ in terms of Stirling numbers of the second kind:

$$F_k(n) = \sum_{\ell=0}^{n} (-1)^\ell \binom{n}{\ell} \left\{ {n-\ell \atop k-\ell} \right\}.$$

In order to prove this, we proceed by inclusion-exclusion, and we work in the universe U of partitions of $[n]$ into k non-empty blocks. Let $A_i \subseteq U$ be the set of partitions in which element i is in a block of size 1. We seek $|(\bigcup_{i=1}^{n} A_i)^c|$. By inclusion-exclusion,

$$|(\bigcup_{i=1}^{n} A_i)^c| = \sum_{I \subseteq [n]} (-1)^{|I|} |\bigcap_{i \in I} A_i|.$$

Now, since partitions in $\bigcap_{i \in I} A_i$ fix each of the $|I|$ elements of I in singleton blocks, they must partition the remaining $n - |I|$ elements into $k - |I|$ non-empty blocks, so

$$|\bigcap_{i \in I} A_i| = \left\{ {n - |I| \atop k - |I|} \right\}.$$

It follows that

$$F_k(n) = \sum_{\ell=0}^{n} (-1)^\ell \binom{n}{\ell} \left\{ {n-\ell \atop k-\ell} \right\}.$$

17. Some problems

(1) Let D_n denote the number of derangements of $\{1,\ldots,n\}$. Prove that D_n can be calculated using either of the following recurrence relations:
   (a) $D_0 = 1$ and for $n \geq 1$, $n! = \sum_{k=0}^{n} \binom{n}{k} D_{n-k}$.
   (b) $D_0 = 1$, $D_1 = 0$ and for $n \geq 2$, $D_n = (n-1)\,(D_{n-1} + D_{n-2})$.

(2) The *Bonferroni inequalities* say that if one truncates the summation in the inclusion-exclusion formula, one alternately over-counts and under-counts the size of a union. Specifically, for each $n \geq 1$ and $1 \leq k \leq n$, if k is odd then

$$\left|\bigcup_{i=1}^{n} A_i\right| \leq \sum_{j=1}^{k} (-1)^{j-1} \sum_{I\subseteq\{1,\ldots,n\},\ |I|=j} \left|\bigcap_{i\in I} A_i\right|$$

while if k is even

$$|\bigcup_{i=1}^{n} A_i| \geq \sum_{j=1}^{k} (-1)^{j-1} \sum_{I\subseteq\{1,\ldots,n\},\ |I|=j} \left|\bigcap_{i\in I} A_i\right|.$$

In particular when $k = 1$ we get the easy and useful *union bound*:

$$|A_1 \cup \ldots \cup A_n| \leq |A_1| + \ldots + |A_n|.$$

Give a combinatorial proof of the Bonferroni inequalities.

Solution: Anything which is not in $\bigcup_{i=1}^{n} A_i$ is counted 0 times in both $|\bigcup_{i=1}^{n} A_i|$ and

$$\sum_{j=1}^{k} (-1)^{j-1} \sum_{I\subseteq\{1,\ldots,n\},\ |I|=j} \left|\bigcap_{i\in I} A_i\right|.$$

Consider an $x \in \left(\bigcap_{j\in J} A_j\right) \cap \left(\bigcap_{j\notin J} A_j^c\right)$ for some $J \subseteq [n]$, $J \neq \emptyset$, with $|J| = t$. Such an x is counted once in $|\bigcup_{i=1}^{n} A_i|$, while in

$$\sum_{j=1}^{k} (-1)^{j-1} \sum_{I\subseteq\{1,\ldots,n\},\ |I|=j} \left|\bigcap_{i\in I} A_i\right|$$

it is counted

$$\sum_{j=1}^{\min\{t,k\}} (-1)^{j-1} \binom{t}{j}$$

times. If $t \leq k$ then (as we argued in the proof of the inclusion-exclusion formula) the above is 1. So it remains to consider the case $t > k$, in which case x is being counted

$$\sum_{j=1}^{k} (-1)^{j-1} \binom{t}{j}$$

times. We need to show that if k is odd then the above is at least 1 (so that the left-hand side of the Bonferroni inequality is at most as large as the right-hand side), and that if k is even then the above is at most 1 (so that the right-hand side of the Bonferroni inequality is at most as large as the right-hand side).

Multiplying by -1 and then adding 1, this is the same as showing that

$$\sum_{j=0}^{k}(-1)^j\binom{t}{j}$$

is non-positive for k odd, and non-negative for k even.

A little doodling reveals a stronger conjecture, namely that for $t \geq 1$ and $k < t$ we have

$$\sum_{j=0}^{k}(-1)^j\binom{t}{j} = (-1)^k\binom{t-1}{k}.$$

This could easily be proven by induction, but there is also a combinatorial argument. Consider k even. We want to show that

$$\left[\binom{t}{0} + \binom{t}{2} + \ldots + \binom{t}{k}\right] - \left[\binom{t}{1} + \binom{t}{3} + \ldots + \binom{t}{k-1}\right] = \binom{t-1}{k}.$$

We could do this combinatorially by exhibiting an injective map from the subsets of a set of size t of sizes $1, 3, \ldots, k-1$, to the subsets of size $0, 2, \ldots, k$, that has the property that the cardinality of the subsets not in the image of the map is $\binom{t-1}{k}$.

Consider the map that sends a set A to $A \cup \{t\}$ if $t \notin A$, and that sends it to $A \setminus \{t\}$ if $t \in A$. This is easily seen to be an injection from the subsets of a set of size t of sizes $1, 3, \ldots, k-1$, to the subsets of size $0, 2, \ldots, k$. The subsets not caught in the image are easily seen to be those of size k that don't include element t; there are $\binom{t-1}{k}$ of these.

A similar argument works when k is odd.

(3)

Identity 17.1. *Let N, j and n be nonnegative integers with $N \geq j$ and $N \geq n$. Prove that*

$$\sum_{i=0}^{n}(-1)^i\binom{n}{i}\binom{N-i}{j} = \binom{N-n}{N-j}.$$

Proof of Identity 17.1. Let U be the set of all j-element subsets of $[N]$. For each $i \in [n]$, we let A_i the set of all $S \in U$ that don't contain i. Then, $(\bigcup_{i=1}^{n} A_i)^c$ (the complement being taken inside U) is the set of all $S \in U$ that contain each $i \in [n]$ (that is, contain all the n numbers $1, 2, \ldots, n$). Clearly, there are $\binom{N-n}{j-n}$ such S's (because each such S is a j-element subset of $[N]$ that is required to contain the n numbers $1, 2, \ldots, n$, and its remaining $j-n$ elements can be chosen arbitrarily from the remaining $N-n$ elements of $[N]$). Thus,

$$\left|(\bigcup_{i=1}^{n} A_i)^c\right| = \binom{N-n}{j-n} = \binom{N-n}{(N-n)-(j-n)} = \binom{N-n}{N-j} \tag{20}$$

(where we used (6) for the second equality sign).

On the other hand, we can compute $|(\bigcup_{i=1}^{n} A_i)^c|$ using (12).

To do so, we fix a subset I of $[n]$. Then, $\bigcap_{i\in I} A_i$ is the set of all $S \in U$ that contain none of the $i \in I$. In other words, $\bigcap_{i\in I} A_i$ is the set of all j-element subsets of $[N]$

that contain none of the $i \in I$. In other words, $\bigcap_{i \in I} A_i$ is the set of all j-element subsets of $[N] \setminus I$. Hence,

$$\left| \bigcap_{i \in I} A_i \right| = \binom{|[N] \setminus I|}{j} = \binom{N - |I|}{j}. \tag{21}$$

Now, forget that we fixed I. Then, (12) yields

$$\begin{aligned}
\left| \left(\bigcup_{i=1}^{n} A_i \right)^c \right| &= \sum_{I \subseteq [n]} (-1)^{|I|} \underbrace{\left| \bigcap_{i \in I} A_i \right|}_{= \binom{N-|I|}{j} \text{ (by (21))}} = \sum_{I \subseteq [n]} (-1)^{|I|} \binom{N - |I|}{j} \\
&= \sum_{i=0}^{n} \sum_{I \subseteq [n];\ |I|=i} \underbrace{(-1)^{|I|} \binom{N - |I|}{j}}_{= (-1)^i \binom{N-i}{j} \text{ (since } |I|=i)} = \sum_{i=0}^{n} \sum_{I \subseteq [n];\ |I|=i} (-1)^i \binom{N-i}{j} \\
&= \sum_{i=0}^{n} \binom{n}{i} (-1)^i \binom{N-i}{j}
\end{aligned}$$

(because for any $i \geq 0$, there are exactly $\binom{n}{i}$ subsets I of $[n]$ satisfying $|I| = i$). Comparing this with (20), we find

$$\binom{N-n}{N-j} = \sum_{i=0}^{n} \binom{n}{i} (-1)^i \binom{N-i}{j} = \sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{N-i}{j}.$$

This proves Identity 17.1. □

18. Partitions of an integer

We have seen compositions: a composition of n is a vector $(a_1, \ldots, a_k)$ of positive integers with $\sum_{j=1}^{k} a_j = n$. Here the order of the a_i's matters; $(1, 3, 1)$ and $(3, 1, 1)$ are different compositions of 5. If we consider two compositions to be the same if they differ only up to a re-ordering of the components of the vector, then we are in the world of *partitions* (or, to avoid confusion with set partitions, *partitions of an integer*). Formally, a *partition* of the positive integer n is a vector $(a_1, \ldots, a_k)$ with the a_i's all positive integers, arranged in non-increasing order (that is, $a_1 \geq a_2 \geq \ldots \geq a_k$) that sum to n (that is, $a_1 + \ldots + a_k = n$). We will often abuse notation and say that the expression "$a_1 + \ldots + a_k$" (with $a_1 \geq \ldots \geq a_k \geq 1$ and $\sum a_i = n$) is a partition of n.

Rather than defining $p(n, k)$ to be the number of partitions of n into k parts, we simply jump straight to the analog of Bell numbers for partitions, and define, for $n \geq 1$,

$$p(n) = \text{ the number of partitions of } n.$$

For example, $p(1) = 1$, $p(2) = 2$, $p(3) = 3$ and $p(4) = 5$ (the five partitions of 4 being 4, $3 + 1$, $2 + 2$, $2 + 1 + 1$ and $1 + 1 + 1 + 1$).

The study of partitions has a long and glorious history, going back to Euler. No simple formula is known for $p(n)$ (though there is a summation formula that sums around $\sqrt{n}$ terms

to get the value of $p(n)$). A remarkable asymptotic formula was found in 1918 by Hardy and Ramanujan:

$$p(n) \sim \frac{1}{4n\sqrt{3}} \exp\left(\pi\sqrt{\frac{2n}{3}}\right).$$

We won't aim as high as this in these notes, contenting ourselves with a few nice facts about partitions that admit easy bijective proofs. To begin, notice that 4 has 2 partitions that consist exclusively of odd parts ($3+1$ and $1+1+1+1$), and 2 partitions that consist of distinct parts (4 and $3+1$). Stepping up one, 5 has the following partitions into odd parts: 5, $3+1+1$ and $1+1+1+1+1$, three of them, and the following partitions into distinct parts: 5, $4+1$ and $3+2$, also three of them. The pattern continues.

Proposition 18.1. *For each $n \geq 1$, n has exactly as many partitions into odd parts as it does partitions into distinct parts.*

Proof. We notice first that each positive integer has a unique representation in the form $2^x y$ for $x, y \in \mathbb{N}$ with y odd. (This follows from unique prime factorization, or simply because we can keep dividing the integer by 2 until it becomes odd.)

We'll construct a bijection from $\mathcal{O}_n$, the set of partitions of n into odd parts, to $\mathcal{D}_n$, the set of partitions of n into distinct parts. Let $a_1 + a_2 + \ldots + a_\ell$ be a partition of n into odd parts. Re-write this sum as $b_1 1 + b_3 3 + b_5 5 + \ldots$, where for each odd number k, b_k is the number of times that k occurs in the partition $a_1 + a_2 + \ldots + a_\ell$. Write each b_i in binary, as $b_i = b_{i1} + b_{i2} + \ldots$, where each b_{ij} is a power of 2, and $b_{i1} < b_{i2} < \ldots$. Next, distribute the odd numbers into the sums of powers of 2 to get

$$b_{i1}1 + b_{i2}1 + \ldots + b_{31}3 + b_{32}3 + \ldots + b_{51}5 + b_{52}5 + \ldots.$$

The terms in this sum are distinct (since each positive integer has a unique representation in the form $2^x y$ for $x, y \in \mathbb{N}$ with y odd) and add to n, so when they are rearranged in descending order, they form a partition of n into distinct parts.

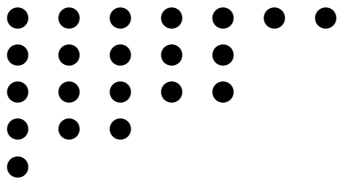
We thus have a map from $\mathcal{O}_n$ to $\mathcal{D}_n$. For injectivity, let p and p' be different partitions in $\mathcal{O}_n$. Let k be an odd number that appears b_k times in p, and b'_k times in p', $b_k \neq b'_k$. Because $b_k \neq b'_k$ there is a power of 2, say 2^s, that (without loss of generality) appears in the binary expansion of b_k but not in that of b'_k, so that the image of p has a part of the form $2^s k$, but the image of p' does not (here we use again that each positive integer has a unique representation in the form $2^x y$ for $x, y \in \mathbb{N}$ with y odd). For surjectivity, given a partition $q \in \mathcal{D}_n$, we can group together parts that have the same y in the representation $2^x y$ (with y odd), sum all the 2^x's for each y to get a multiplicity $m(y)$ for y (notice that all the 2^x's corresponding to the same y are distinct, because we are in $\mathcal{D}_n$), and look at the partition of n that has each odd y appearing with multiplicity $m(y)$; this is in $\mathcal{O}_n$, and evidently maps to q. □

As an example of the bijection described in the proof of Proposition 18.1, we have

$$\begin{aligned} 13+13+5+5+5+3 &= 1\cdot 3 + 3\cdot 5 + 2\cdot 13 \\ &= (1)3 + (1+2)5 + (2)13 \\ &= 3 + 5 + 10 + 26 \\ &= 26 + 10 + 5 + 3, \end{aligned}$$

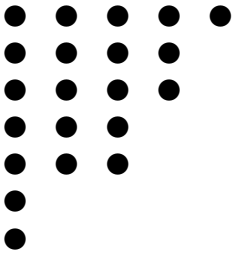
so the partition $(13, 13, 5, 5, 5, 3)$ of 44 maps to $(26, 10, 5, 3)$.

For the next two cute facts about partitions, it will be useful to introduce the *Ferrers diagram* of a partition. The Ferrers diagram of $(a_1, \ldots, a_\ell)$ consists of ℓ rows of dots, left justified, with the jth row (read from the top down) having a_j dots. For example, here is the Ferrers diagram of $(7, 5, 5, 3, 1)$:



Ferrers diagrams (which completely encode partitions; there is a one-to-one correspondence between partitions of n and Ferrers diagrams with n dots) play a major role in representation theory; we may say more about this later.

Each Ferrers diagram has a *conjugate*, obtained by rotating the diagram along the northwest-southeast diagonal that passes through the northwesternmost dot of the diagram. For example, here is the conjugate of the above Ferrers diagram:



Notice that this is itself a Ferrers diagram, the diagram of $(5, 4, 4, 3, 3, 1, 1)$. We refer to this as the *conjugate* of $(7, 5, 5, 3, 1)$.

The following two identities involving partitions of n can both be proven in the same way: in each case the conjugate map provides a bijection between the two sets in question. We leave the details as exercises.

Proposition 18.2. (1) *Fix $n \geq k \geq 1$. There are as many partitions of n into at least k parts, as there are partitions in which the largest part is at least k.*

(2) *Fix $n \geq 1$. There are as many partitions of n in which the first two parts are equal, as there are partitions in which all parts are at least 2.*

19. Some problems

(1) Prove Proposition 18.2.

20. The Twelvefold way

Gian-Carlo Rota identified a collection of twelve fundamental counting problems, that can all be framed as "in how many ways can n balls be placed into k boxes". The twelve different problems arise from considering

- whether the balls are distinguishable or not,
- whether the boxes are distinguishable or not, and
- whether no restriction is placed on the distribution, or the restriction that each box must get at least one ball, or the restriction that each box gets at most one ball.

More formally we can think of enumerating functions from X (the set of balls) to Y (the set of boxes), up to equivalence, with four notions of whether two functions are equivalent:

(1) functions are equivalent if they are equal (or identical) (balls and boxes both distinguishable);
(2) functions are equivalent if one can be obtained from the other by a permutation of X (balls are indistinguishable, boxes are distinguishable);
(3) functions are equivalent if one can be obtained from the other by a permutation of Y (balls are distinguishable, boxes are indistinguishable); and
(4) functions are equivalent if one can be obtained from the other by a permutation of X and a permutation of Y (balls and boxes both indistinguishable)

and with three possible restrictions on the kinds of functions we consider:

(1) no restriction (no restriction on how many balls can be placed in each box);
(2) functions are injective (no box can get more than one ball); and
(3) functions are surjective (every box must get at least one ball).

Notice that we do not include the restriction "bijective"; this only gives a non-zero count if $n = k$, in which case it is equivalent to both injectivity and surjectivity.

We have addressed all twelve counting problems. Here we summarize. We encode each problem by a pair (a)(b) where $1 \leq a \leq 4$ and $1 \leq b \leq 3$; the encoding being that (a)(b) refers to the problem obtained by considering the a$^{\text{th}}$ equivalence relation of the four defined above, and the b$^{\text{th}}$ restriction of the three listed above.

- (1)(1), distinguishable balls, distinguishable boxes, no restriction: k^n.
- (1)(2), distinguishable balls, distinguishable boxes, injective: $k(k-1)\dots(k-(n-1))$.
- (1)(3), distinguishable balls, distinguishable boxes, surjective: $k!\left\{ {n \atop k} \right\}$.
- (2)(1), indistinguishable balls, distinguishable boxes, no restriction: This is easily seen to be the problem of weak compositions of n into k parts, so $\binom{n+k-1}{k-1}$ when $n > 0$.
- (2)(2), indistinguishable balls, distinguishable boxes, injective: This is easily seen to be the problem of subsets of size n of a set of size k, so $\binom{k}{n}$.
- (2)(3), indistinguishable balls, distinguishable boxes, surjective: This is easily seen to be the problem of compositions of n into k parts, so $\binom{n-1}{k-1}$ when $n > 0$.
- (3)(1), distinguishable balls, indistinguishable boxes, no restriction: This is easily seen to be the problem of partitioning a set of size n into at most k non-empty sets, so $\sum_{j \leq k} \left\{ {n \atop j} \right\}$.
- (3)(2), distinguishable balls, indistinguishable boxes, injective: If $k \geq n$ this is 1, and it is 0 otherwise.
- (3)(3), distinguishable balls, indistinguishable boxes, surjective: This is easily seen to be the problem of partitioning a set of size n into k non-empty sets, so $\left\{ {n \atop k} \right\}$.
- (4)(1), indistinguishable balls, indistinguishable boxes, no restriction: This is easily seen to be the problem of partitioning the number n into at most k non-empty parts, so $\sum_{j \leq k} p(n, j)$.
- (4)(2), indistinguishable balls, indistinguishable boxes, injective: If $k \geq n$ this is 1, and it is 0 otherwise.
- (4)(3), indistinguishable balls, indistinguishable boxes, surjective: This is easily seen to be the problem of partitioning the number n into k non-empty parts, so $p(n, k)$.

21. Generating functions

To a sequence $(a_0, a_1, \ldots)$, we may associate a *generating function*

$$A(x) = a_0 + a_1 x + a_2 x^2 + \ldots.$$

This may be viewed as an element of the ring (over the complex numbers) of formal power series in x (an approach we will talk about later), or way may think of it as an ordinary power series in the complex variable x (which is the approach we will take initially). Viewing $A(x)$ as an ordinary power series raises questions of convergence; we will get over these by promising to only work with values of x that are inside the radius of convergence of $A(x)$ (we will not, at least initially, encounter any generating functions with radius of convergence 0). Often during the course of a single problem we will encounter multiple generating functions; our approach will be to manipulate them at will (via the operations of term-by-term integration and differentiation, for example), and then at the end think about the range of x that would allow all of the manipulations to be justified.

One might initially look at $A(x)$ and not see any value in its introduction, noting that we have simply replaced one infinite encoding of a sequence with another. By introducing power series, however, we have the opportunity to bring an analytic approach to bear on the study of the sequence, and therein lies the power of the method of generating functions.

A typical situation in which generating functions can be helpful, is when the sequence $(a_n)_{n\geq 0}$ is presented via a recurrence relation (a relation that specifies a_n in terms of the values $a_0, \ldots, a_{n-1}$, for all n at least as large as a certain threshold n_0, and also gives the initial data $a_0, \ldots, a_{n_0-1}$). It this case the method of generating proceeds as follows:

(1) Write down the generating function in the form given above, $A(x) = \sum_{n\geq 0} a_n x^n$.
(2) EITHER:
   - re-write the right-hand side of generating function, replacing $a_0, \ldots, a_{n_0-1}$ with their known values, and replacing a_n $(n \geq n_0)$ with the expression (in terms of earlier values of the sequence) given by the recurrence relation (this is the approach I prefer)

   OR:
   - write down the general recurrence relation $a_n = f(a_0, \ldots, a_{n-1})$, multiply both sides by x^n, and then sum both sides over all $n \geq n_0$ (this is the approach suggested by Wilf in his book *generatingfunctionology*).
(3) Rearrange terms in the identity you are looking at, to recover copies of $A(x)$, often by adding finitely many terms to an infinite sum that consists of all but finitely many terms from $A(x)$. Using my method, only the right-hand side will be modified; using Wilf's method, both sides will need to be, with in particular the left-hand side having $a_0 + a_1 x + \ldots + a_{n_0-1} x^{n_0-1}$ added to bring it up to $A(x)$.
(4) Solve the resulting functional equation to get a compact expression for $A(x)$.
(5) Use tools from analysis (in particular, tools relating to Taylor series) to extract the coefficient of x^n in this compact expression. This is an explicit expression for a_n.

In general generating function applications these last three steps may require a great deal of ingenuity, and indeed one or more of them may be impossible. There is one special but important case where the three steps are (at least in theory) easy: the case of linear, finite-depth recurrences.

Definition 21.1. *A* linear, depth k recurrence *is a recurrence of the form*

$$a_n = c_1 a_{n-1} + \ldots + c_k a_{n-k} \ (n \geq k)$$

(where the c_i's are constants, $c_k \neq 0$) with initial values $a_0, \ldots, a_{k-1}$ given.

Before dealing with the general theory, we consider the special case of the Fibonacci numbers, defined by $f_0 = 0, f_1 = 1$ and $f_n = f_{n-1} + f_{n-2}$, a linear, depth 2 recurrence. We begin with the generating function of the sequence $(f_n)_{n \geq 0}$ in the form

$$F(x) = f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \ldots .$$

Using my preferred method of "use recurrence where possible, initial conditions where necessary", we get

$$\begin{aligned} F(x) &= x + (f_1 + f_0)x^2 + (f_2 + f_1)x^3 + \ldots \\ &= x + (f_1 x^2 + f_2 x^3 + \ldots) + (f_0 x^2 + f_1 x^3 + \ldots) \\ &= x + x(f_1 x + f_2 x^2 + \ldots) + x^2(f_0 + f_1 x + \ldots) \\ &= x + x(F(x) - f_0) + x^2 F(x) \\ &= x + xF(x) + x^2 F(x) \end{aligned}$$

and so

$$F(x) = \frac{x}{1 - x - x^2}.$$

Using the method preferred by Wilf, "multiply recurrence by x^n and sum where valid", we get

$$\begin{aligned} f_n &= f_{n-1} + f_{n-2}, \text{ for } n \geq 2, \text{ so} \\ f_n x^n &= f_{n-1} x^n + f_{n-2} x^n, \text{ for } n \geq 2, \text{ so} \\ \sum_{n \geq 2} f_n x^n &= \sum_{n \geq 2} f_{n-1} x^n + \sum_{n \geq 2} f_{n-2} x^n, \text{ so} \\ F(x) - f_0 - f_1 x &= x \sum_{n \geq 1} f_n x^n + x^2 \sum_{n \geq 0} f_n x^n, \text{ so} \\ F(x) - x &= x(F(x) - f_0) + x^2 F(x), \text{ so} \\ F(x) - x &= xF(x) + x^2 F(x), \end{aligned}$$

and so

$$F(x) = \frac{x}{1 - x - x^2}.$$

To find the coefficient of x^n in $F(x)$, we use the method of partial fractions. Defining α_1, α_2 via

$$1 - x - x^2 = (1 - \alpha_1 x)(1 - \alpha_2 x), \tag{22}$$

the method of partial fractions tells us that there are constants A_1 and A_2 such that

$$\frac{x}{1 - x - x^2} = \frac{A_1}{1 - \alpha_1 x} + \frac{A_2}{1 - \alpha_2 x}.$$

We now use our knowledge of Taylor series — specifically, the fact that

$$\frac{1}{1 - z} = 1 + z + z^2 + \ldots ,$$

valid as long as $|z| < 1$, to conclude that the coefficient of x^n in $x/(1 - x - x^2)$ is

$$A_1\alpha_1^n + A_2\alpha_2^n. \tag{23}$$

What are the constants α_1, α_2, A_1 and A_2? Well, dividing both sides of (22) by x^2 and setting $z = 1/x$, we find that

$$z^2 - z - 1 = (z - \alpha_1)(z - \alpha_2);$$

in other words, α_1 and α_2 are the roots of the quadratic $z^2 - z - 1 = 0$, so are $(1 \pm \sqrt{5})/2$ (for definiteness, say $\alpha_1 = (1+\sqrt{5})/2$). To find A_1 and A_2, we have to do a little bit of linear algebra. Knowing $f_0 = 0$ and $f_1 = 1$, we get from (23) that

$$\begin{aligned} 0 &= A_1 + A_2 \\ 1 &= A_1\alpha_1 + A_2\alpha_2. \end{aligned}$$

Solving this system of simultaneous equations yields $A_1 = 1/\sqrt{5}$, $A_2 = -1/\sqrt{5}$, and so we get the famous Binet's formula:

$$f_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right).$$

Notice that if we are concerned about convergence, we can go back through the argument and observe that all of our power series manipulations were valid as long as both $|\alpha_1 x| < 1$ and $|\alpha_2 x| < 1$, that is, as long as $|x| < (-1+\sqrt{5})/2 \approx .618$, a quite reasonable open set in which to operate. Notice also that the entire process was analytic: if there was any combinatorics involved, it had to come in earlier than the present analysis, perhaps in the derivation of the recurrence for the Fibonacci numbers.

We now deal with the more general case of an arbitrary linear, depth k recurrence, as presented in Definition 21.1. In terms of deriving a closed form for the generating function $A(x) = a_0 + a_1x + a_2x^2 + \ldots$, there is no substantial difference between the general case and the Fibonacci example, so we simply state, and leave it as an exercise to verify, that either using my preferred method or Wilf's, we end up with

$$A(x) = \frac{P(x)}{Q(x)}$$

where

$$Q(x) = 1 - c_1x - c_2x^2 - \ldots - c_kx^k$$

and, writing $A_\ell(x)$ for the truncated generating function $a_0 + a_1x + \ldots + a_\ell x^\ell$,

$$P(x) = A_{k-1}(x) - \sum_{j=1}^{k-1} c_j x^j A_{k-j-1}(x).$$

Notice that $A(x)$ is a rational function whose numerator is a polynomial of degree at most $k-1$, and whose denominator has degree k.

How do we understand the Taylor series of such a rational function? One answer is to use the method of *partial fractions*.

Proposition 21.2. *Let $P(x)/Q(x)$ be a rational function with $Q(x)$ a polynomial of degree k, and $P(x)$ a polynomial of degree at most $k-1$. Suppose that $Q(x)$ factors as*

$$Q(x) = \prod_{i=1}^{r}(a_1 + b_ix)^{m_i}$$

where a_i, b_i are complex numbers. Then there exist constants A_{ij}, $i = 1, \ldots, r$, $j = 1, \ldots, m_i$ such that

$$\frac{P(x)}{Q(x)} = \sum_{i=1}^{r} \sum_{j=1}^{m_i} \frac{A_{ij}}{(a_i - b_i x)^j}.$$

Proof. We first need Bézout's identity for polynomials. Bézout's identity for integers states that if a and b are relatively prime integers (no common factors other than ± 1) then there are integers c and d such that $ca + db = 1$. The proof uses the string of equalities that appears in the Euclidean algorithm for finding the greatest common divisor of a and b, and finds the appropriate linear combination of a and b by working backwards through that string. The proof for polynomials over the complex numbers is identical; all we need to know is the *division algorithm for polynomials*: given two univariate polynomials $a(x)$ and $b(x) \neq 0$ defined over a field, there exist two polynomials $q(x)$ (the quotient) and $r(x)$ (the remainder) which satisfy

$$a(x) = b(x)q(x) + r(x)$$

and the degree of $r(x)$ is strictly less than the degree of $b(x)$ (the degree of the zero polynomial is -1). By repeated application of the division algorithm, we get the following chain of equalities:

$$\begin{aligned}
a &= bq_0 + r_0 \\
b &= r_0 q_1 + r_1 \\
r_0 &= r_1 q_2 + r_2 \\
&\cdots \\
r_{k-4} &= r_{k-3} q_{k-2} + r_{k-2} \\
r_{k-3} &= r_{k-2} q_{k-1} + r_{k-1} \\
r_{k-2} &= r_{k-1} q_k + r_k
\end{aligned}$$

for some finite k, where all expressions above are polynomials in x, and where r_k is 0 and r_{k-1} is not. Now notice that r_{k-1} divides r_{k-2} (from the last equality, since $r_k = 0$); and so it divides r_{k-3} (from the second-to-last equality); and proceeding up the chain we see that it divides both b and a. Since b and a are coprime, we must have $r_{k-1} = 1$. The second-to-last equality now allows 1 to be expressed as a $\mathbb{C}[x]$-linear combination of r_{k-3} and r_{k-2}; since the third-from-last equality allows r_{k-2} to be expressed as a $\mathbb{C}[x]$-linear combination of r_{k-4} and r_{k-3}, we get that 1 can be expressed as a $\mathbb{C}[x]$-linear combination of r_{k-4} and r_{k-3}; and proceeding up the chain we eventually can express 1 as a $\mathbb{C}[x]$-linear combination of b and a.

In summary: if $a(x)$ and $b(x)$ are two non-zero polynomials with no factor in common over the complex numbers, there are polynomials $c(x)$ and $d(x)$ with

$$c(x)a(x) + d(x)b(x) = 1.$$

This implies that if $e(x)/(a(x)b(x))$ is a rational function, with $a(x)$ and $b(x)$ two non-zero polynomials with no factor in common over the complex numbers, and $e(x)$ is another non-zero complex polynomial, then there exist polynomials $e_1(x)$ and $e_2(x)$ such that

$$\frac{e(x)}{a(x)b(x)} = \frac{e_1(x)}{a(x)} + \frac{e_2(x)}{b(x)}.$$

Indeed, if $c(x)$ and $d(x)$ are such that $c(x)a(x) + d(x)b(x) = 1$, then $e_1(x) = d(x)e(x)$ and $e_2(x) = c(x)e(x)$ work.

Applying to our present situation, factor $Q(x)$ over the complex numbers as $\prod_{i=1}^{r}(a_i - b_i x)^{m_i}$. By induction we find that there are polynomials $e_1(x), \ldots, e_r(x)$ such that

$$\frac{P(x)}{Q(x)} = \sum_{i=1}^{r} \frac{e_i(x)}{(a_i - b_i x)^{m_i}}.$$

By polynomial long division, we may write this as

$$\frac{P(x)}{Q(x)} = E(x) + \sum_{i=1}^{r} \frac{\widetilde{e}_i(x)}{(a_i - b_i x)^{m_i}}$$

where $E(x)$ is a polynomial and for each i, the degree of the polynomial $\widetilde{e}_i(x)$ is strictly less than m_i. Taking limits as $x \to \infty$ (or, alternatively, multiplying the equality through by $Q(x)$ and comparing degrees on both sides), we conclude that $E(x) = 0$.

It remains to show that for any polynomial of the form $\widetilde{e}(x)/(a+bx)^m$ (with the degree of $\widetilde{e}(x)$ strictly less than m) there is an expansion of the form

$$\frac{\widetilde{e}(x)}{(a+bx)^m} = \sum_{j=1}^{m} \frac{A_j}{(a+bx)^j}.$$

Multiplying through by $(a+bx)^m$, this is the same as

$$\widetilde{e}(x) = \sum_{j=1}^{m} A_j (a+bx)^{m-j}.$$

But such an expansion exists; the polynomials $(a+bx)^{m-1}, \ldots, (a+bx)^0$ form a basis for the space of polynomials of degree at most $m-1$, and $\widetilde{e}(x)$ is in this space. □

We now apply Proposition 21.2 to our rational expression for $A(x)$. Factoring

$$Q(x) = \prod_{i=1}^{r} (1 - \alpha_i x)^{m_i},$$

we find that there are constants A_{ij}, $i = 1, \ldots, r$, $j = 1, \ldots, m_i$ such that

$$A(x) = \sum_{i=1}^{r} \sum_{j=1}^{m_i} \frac{A_{ij}}{(1-\alpha_i x)^j}.$$

We have reduced the problem of understanding a_n to that of understanding the Taylor series of functions of the form $f(x) = (1-\alpha x)^{-s}$ for positive integers s. For this we need a generalization of the binomial theorem.

Theorem 21.3. *Fix $\alpha \in \mathbb{R}$. Define $f_\alpha(x) = (1+x)^\alpha$ on $(-1, 1)$. The sum*

$$\sum_{k \geq 0} \frac{f_\alpha^{(k)}(x)}{k!} x^k = \sum_{k \geq 0} \frac{(\alpha)_k}{k!} x^k$$

is absolutely convergent, and sums to $f_\alpha(x)$ (where here $f_\alpha^{(k)}(x)$ is the kth derivative of $f_\alpha(x)$ with respect to x, with the understanding that the 0th derivative of a function f is f itself).

We won't give a proof of this; it is simply a matter of working carefully with the remainder term in any version of Taylor's theorem.

If we extend the definition of the binomial coefficient to allow the upper number to be arbitrary, via

$$\binom{\alpha}{k} = \frac{(\alpha)_k}{k!},$$

then we get the following, which justifies referring to Theorem 21.3 as a generalization of the binomial theorem: for all real α and all x satisfying $|x| < 1$,

$$(1+x)^\alpha = \sum_{k \geq 0} \binom{\alpha}{k} x^k.$$

Notice that this does indeed generalize the binomial theorem: if $\alpha = n$, a positive integer, then $\binom{\alpha}{k} = \binom{n}{k}$ for all k.

Applying the binomial theorem with $\alpha = -s$, where s is a positive integer, we get

$$\begin{aligned}
(1-x)^{-s} &= \sum_{k \geq 0} (-1)^k \binom{-s}{k} x^k \\
&= \sum_{k \geq 0} (-1)^k \frac{(-s)_k}{k!} x^k \\
&= \sum_{k \geq 0} \frac{s^{(k)}}{k!} x^k \\
&= \sum_{k \geq 0} \binom{s+k-1}{k} x^k \\
&= \sum_{k \geq 0} \binom{s+k-1}{s-1} x^k
\end{aligned}$$

Notice that when $s = 1$ we recover the familiar

$$\frac{1}{1-x} = 1 + x + x^2 + \ldots,$$

and when $s = 2$ we get

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \ldots.$$

We are now ready to give an explicit solution to an arbitrary depth k linear recurrence.

Theorem 21.4. *Let the sequence $(a_n)_{n \geq 0}$ be defined by*

$$a_n = c_1 a_{n-1} + \ldots + c_k a_{n-k} \ (n \geq k)$$

(where the c_i's are constants, $c_k \neq 0$), with initial values $a_0, \ldots, a_{k-1}$ given. The generating function $A(x) = \sum_{n \geq 0} a_n x^n$ is a rational function, specifically

$$A(x) = \frac{P(x)}{Q(x)}$$

where

$$Q(x) = 1 - c_1 x - c_2 x^2 - \ldots - c_k x^k$$

and, writing $A_\ell(x)$ for the truncated generating function $a_0 + a_1x + \ldots + a_\ell x^\ell$,

$$P(x) = A_{k-1}(x) - \sum_{j=1}^{k-1} c_j x^j A_{k-j-1}(x).$$

Factoring

$$Q(x) = \prod_{i=1}^{r} (1 - \alpha_i x)^{m_i},$$

there are constants A_{ij}, $i = 1, \ldots, r$, $j = 1, \ldots, m_i$ such that

$$A(x) = \sum_{i=1}^{r} \sum_{j=1}^{m_i} \frac{A_{ij}}{(1 - \alpha_i x)^j}.$$

We have the following formula for a_n:

$$a_n = \sum_{i=1}^{r} \sum_{j=1}^{m_i} A_{ij} \binom{j+n-1}{j-1} \alpha_i^n. \tag{24}$$

where the α_i are the r distinct roots of

$$z^k - c_1 z^{k-1} - c_2 z^{k-2} - \ldots - c_k = 0,$$

and m_i is the multiplicity of α_i. The constants A_{ij} may be found by solving the k by k system of linear equations given by applying (24) in the cases $n = 0, 1, \ldots, k-1$.

Notice that the entire discussion is valid in the world of complex power series as long as $|x| < 1/\max_i |\alpha_i|$.

Examining the asymptotics of a_n in (24) as $n \to \infty$, we easily obtain the following.

Corollary 21.5. *Let the sequence $(a_n)_{n \geq 0}$ be defined by*

$$a_n = c_1 a_{n-1} + \ldots + c_k a_{n-k} \ (n \geq k)$$

(where the c_i's are constants, $c_k \neq 0$), with initial values $a_0, \ldots, a_{k-1}$ given. With the notation as in Theorem 21.4, consider only those α_i for which the corresponding coefficients $A_{i,1}, A_{i,2}, \ldots, A_{i,m_i}$ are not all zero. Suppose that among these α_i, there is a unique α_i of maximum modulus. Denote this α_i by α_ι. Also, let γ be the largest j such that $A_{\iota,j} \neq 0$. Then,

$$a_n \sim \frac{A_{\iota,\gamma}}{(\gamma - 1)!} n^{\gamma - 1} \alpha_\iota^n$$

as $n \to \infty$, that is,

$$\lim_{n \to \infty} \frac{a_n}{\frac{A_{\iota,\gamma}}{(\gamma-1)!} n^{\gamma-1} \alpha_\iota^n} = 1.$$

Notice that the α_ι in this corollary must be real if the initial values $a_0, \ldots, a_{k-1}$ and the constants $c_1, \ldots, c_k$ are real (by its uniqueness).

Let us do an example (carefully chosen so the numbers work out nicely). Suppose that a_n is given by the recurrence

$$a_n = 6a_{n-1} - 7a_{n-2} - 12a_{n-3} + 18a_{n-4}$$

for $n \geq 4$, with $a_0 = a_1 = a_2 = 0$ and $a_3 = 1$, so the sequence begins

$$0, 0, 0, 1, 6, 29, 120, 463, 1698, 6029, 20892, 71107, 238614.$$

We have

$$Q(x) = 1 - 6x + 7x^2 + 12x^3 - 18x^4 = (1-3x)^2(1-\sqrt{2}x)(1+\sqrt{2}x).$$

From Theorem 21.4 we get that

$$a_n = A_{11}3^n + A_{12}(n+1)3^n + A_{21}\sqrt{2}^n + A_{31}(-\sqrt{2})^n.$$

Plugging in $n = 0, 1, 2$ and 3 we obtain

$$\begin{array}{ccccccccc} A_{11} & + & A_{12} & + & A_{21} & + & A_{31} & = & 0 \\ 3A_{11} & + & 6A_{12} & + & \sqrt{2}A_{21} & - & \sqrt{2}A_{31} & = & 0 \\ 9A_{11} & + & 27A_{12} & + & 2A_{21} & + & 2A_{31} & = & 0 \\ 27A_{11} & + & 108A_{12} & + & 2\sqrt{2}A_{21} & - & 2\sqrt{2}A_{31} & = & 1. \end{array}$$

Solving this system yields

$$A_{11} = \frac{-25}{147},\ A_{12} = \frac{1}{21},\ A_{21} = \frac{33+18\sqrt{2}}{294\sqrt{2}},\ A_{31} = \frac{33-18\sqrt{2}}{294\sqrt{2}}$$

so that

$$a_n = \frac{n3^n}{21} - \frac{6\cdot 3^n}{49} + \frac{1}{294\sqrt{2}}\left((33+18\sqrt{2})2^{n/2} + (33-18\sqrt{2})(-2)^{n/2}\right),$$

and so in particular

$$a_n \sim \frac{n3^n}{21}$$

as $n \to \infty$.

If a term $f(n)$ is added to the recurrence — so that it becomes

$$a_n = c_1 a_{n-1} + \ldots + c_k a_{n-k} + f(n)\ (n \geq k),$$

then we get

$$A(x) = \frac{P(x) + F_{\geq k}(x)}{Q(x)}$$

where $F_{\geq k}(x) = \sum_{n \geq k} f(n)x^n$. If $F_{\geq k}$ happens to be a rational function of x, then again the method of partial fractions gives a way of solving explicitly for a_n. We look at one specific situation here, the case when $f(n)$ is a polynomial in n. This requires the study of

$$P^{(\ell)}(x) := \sum_{n \geq 0} n^\ell x^n$$

for each integer $\ell \geq 0$. For $\ell = 0$, the infinite series is $1 + x + x^2 + \ldots = 1/(1-x)$ (interpreting 0^0 as 1). For $\ell > 0$ we have the recurrence

$$P^{(\ell)}(x) = x\frac{d}{dx}\left(P^{(\ell-1)}(x)\right).$$

It follows that

$$P^{(\ell)}(x) = \left(x\frac{d}{dx}\right)^\ell \left(\frac{1}{1-x}\right).$$

We saw earlier that for any infinitely differentiable function $f(x)$,

$$\left(x\frac{d}{dx}\right)^\ell f(x) = \sum_{k=1}^{\ell} \left\{ {\ell \atop k} \right\} x^k \frac{d^k}{dx^k} f(x)$$

for $\ell \geq 1$. Since

$$\frac{d^k}{dx^k}\left(\frac{1}{1-x}\right) = \frac{k!}{(1-x)^{k+1}},$$

we get that

$$P^{(\ell)}(x) = \sum_{k=1}^{\ell} k! \left\{ {\ell \atop k} \right\} \frac{x^k}{(1-x)^{k+1}} \qquad \text{for } \ell \geq 1.$$

Observe what happens if we extract the coefficient of x^n from both sides. From the left-hand side we get n^ℓ. For the right-hand side, note that

$$[x^n]\left(\frac{x^k}{(1-x)^{k+1}}\right) = [x^{n-k}]\left(\frac{1}{(1-x)^{k+1}}\right) = \binom{(k+1)+(n-k)-1}{(k+1)-1} = \binom{n}{k},$$

and so we obtain the identity

$$n^\ell = \sum_{k=1}^{\ell} \left\{ {\ell \atop k} \right\} (n)_k \qquad \text{for } \ell \geq 1.$$

This also holds for $\ell = 0$ if we start the sum at $k = 0$ instead of at $k = 1$. Of course, we can change the upper bound of the sum to ∞, since every $k > \ell$ satisfies $\left\{ {\ell \atop k} \right\} = 0$. Thus we obtain precisely the identity (11); here we have rediscovered it through generating functions, using no combinatorics whatsoever, only the recurrence for the Stirling numbers. (This is assuming that we took the inductive, rather than the combinatorial, approach to Claim 14.6.)

We have shown:

Proposition 21.6. *Let $f(n) = c_0 + c_1 n + c_2 n^2 + \ldots + c_m n^m$. The generating function $F(x) = \sum_{n \geq 0} f(n) x^n$ is given by*

$$\sum_{\ell=0}^{m} c_\ell \sum_{k=0}^{\ell} k! \left\{ {\ell \atop k} \right\} \frac{x^k}{(1-x)^{k+1}}.$$

Since this is rational, we now have a method (at least in principle) of explicitly solving the recurrence

$$a_n = c_1 a_{n-1} + \ldots + c_k a_{n-k} + f(n) \ (n \geq k),$$

for arbitrary polynomials $f(n)$.

22. Some problems

(1) Here's a problem that came up in my research: a hotel corridor has n rooms in a row, numbered 1 through n (so $n \geq 1$). The rooms are to be painted, each one either red, or white, or blue, subject to the condition that a red room can't be immediately adjacent to a blue room. Let p_n be the number of different ways to paint the rooms. So, for example, $p_1 = 3$ and $p_2 = 7$ (the seven legitimate ways being WW, WR, WB, RR, RW, BB and BW). Find a recurrence for p_n, and use generating functions to find an explicit expression for p_n. (Your recurrence doesn't have to be constant-depth; the easiest one to find expresses p_n in terms of all of $p_1, p_2, \ldots, p_{n-1}$.)

Solution: We have $p_1 = 3$ and $p_2 = 7$; we'll take these as initial conditions. For $n \geq 3$, the collection of possible ways to paint the rooms can be decomposed according to the value of k, $0 \leq k \leq n$, the number of the first room painted white (with $k = 0$ meaning that no room is painted white). Corresponding to $k = 0$ there are 2 ways to

paint (all red or all blue). Corresponding to $k = 1$, there are p_{n-1}. Corresponding to $k = 2$ there are $2p_{n-2}$ (the factor of 2 accounting for the number of ways to paint the first room, the p_{n-2} accounting for the number of ways to paint the last $n-2$ rooms). In general, for $k \in \{2, \ldots, n-1\}$ there are $2p_{n-k}$ configurations. Finally, for $k = n$ there are two configurations (all red or all blue for the first $n-1$ rooms). This leads to the recurrence:

$$p_n = 2 + p_{n-1} + 2p_{n-2} + \ldots + 2p_1 + 2 = 4 + p_{n-1} + 2\sum_{i=1}^{n-2} p_i$$

valid for $n \geq 3$.

We form the generating function $P(x) = p_1 x + p_2 x^2 + p_3 x^3 + \ldots$. Apply the usual method, $P(x)$ breaks down into the sum of four pieces:

$$3x + 7x^2,$$

$$4(x^3 + x^4 + \ldots) = \frac{4x^3}{1-x},$$

$$p_2 x^3 + p_3 x^4 + \ldots = x(P(x) - 3x)$$

and

$$2p_1 x^3 + 2(p_1 + p_2)x^4 + 2(p_1 + p_2 + p_3)x^5 + \ldots = \frac{2x^2 P(x)}{1-x},$$

leading to the functional equation

$$P(x) = 3x + 7x^2 + \frac{4x^3}{1-x} + x(P(x) - 3x) + \frac{2x^2 P(x)}{1-x}$$

so that

$$P(x) = \frac{3x + x^2}{1 - 2x - x^2}.$$

(which should immediately tell us that p_n satisfies the simpler recurrence $p_n = 2p_{n-1} + p_{n-2}$, which can indeed by verified by induction!). This leads to the following explicit formula for p_n:

$$p_n = \frac{(1+\sqrt{2})^{n+1}}{2} + \frac{(1-\sqrt{2})^{n+1}}{2};$$

I'm omitting the algebra!

Here's a simple combinatorial derivation of the recurrence $p_n = 2p_{n-1} + p_{n-2}$: there are $3p_{n-2}$ colorings in which the second room is painted white (the first room can be painted in any of 3 ways, the last $n-2$ in any of p_{n-2} ways). If the second room is painted either red or blue there are 2 ways to paint the first room, and $p_{n-1} - p_{n-2}$ ways to paint the last $n-1$ (the $-p_{n-2}$ because we have to remove from the p_{n-1} all the ways of painting the last $n-1$ rooms that starts with a white, and there are evidently p_{n-2} such). So $p_n = 3p_{n-2} + 2(p_{n-1} - p_{n-2}) = 2p_{n-1} + p_{n-2}$.

23. Operations on power series

Write $(a_n)_{n\geq 0} \longleftrightarrow A(x)$ to indicate that $A(x)$ is the generating function of the sequence $(a_n)_{n\geq 0}$, that is, that $A(x) = \sum_{n=0}^{\infty} a_n x^n$, and that $(a_n)_{n\geq 0}$ is the coefficient sequence of $A(x)$, that is, $a_n = [x^n]A(x)$.

There are natural operations that can be performed on power series, and some of these correspond to natural operations on the associated coefficients sequences. We mention some of the more useful here.

Theorem 23.1. *Let $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ be complex sequences with $(a_n)_{n\geq 0} \longleftrightarrow A(x)$ and $(b_n)_{n\geq 0} \longleftrightarrow B(x)$, and let c and d be complex numbers. We have the following relations:*

(1) $(0, a_0, a_1, \ldots) \longleftrightarrow xA(x)$, *and more generally, for $k \geq 1$,*
$$(0, \ldots, 0, a_0, a_1, \ldots) \longleftrightarrow x^k A(x)$$
with the sequence on the left beginning with k zeros.

(2) $(a_1, a_2, \ldots) \longleftrightarrow (A(x) - a_0)/x$, *and more generally, for $k \geq 1$,*
$$(a_k, a_{k+1}, \ldots) \longleftrightarrow (A(x) - a_0 - a_1 x - \ldots - a_{k-1}x^{k-1})/x^k.$$

(3) $(a_n + b_n)_{n\geq 0} \longleftrightarrow A(x) + B(x)$, $(ca_n)_{n\geq 0} \longleftrightarrow cA(x)$, *and more generally*
$$(ca_n + db_n)_{n\geq 0} \longleftrightarrow cA(x) + dB(x).$$

(4) $((n+1)a_{n+1})_{n\geq 0} \longleftrightarrow A'(x)$, *and more generally, for $k \geq 1$*
$$((n+k)_k a_{n+k})_{n\geq 0} \longleftrightarrow \frac{d^k}{dx^k} A(x).$$

(5)
$$\left(0, a_0, \frac{a_1}{2}, \frac{a_2}{3}, \ldots\right) \longleftrightarrow \int_0^x A(t)\ dt.$$

(6) *With $c_n = \sum_{k=0}^{n} a_k b_{n-k}$ for each $n \geq 0$,*
$$(c_n)_{n\geq 0} \longleftrightarrow A(x)B(x).$$

(7) *With $d_n = \sum_{k=0}^{n} a_k$ for each $n \geq 0$,*
$$(d_n)_{n\geq 0} \longleftrightarrow \frac{A(x)}{1-x}.$$

Only the last of these needs some justification, but using $1/(1-x) = 1+x+x^2+\ldots$ it follows immediately from the second-to-last. The sequence $(c_n)_{n\geq 0}$ introduced in the second-to-last identity is the *convolution* of $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$, and may be thought of as a discrete analog of the convolution of two real functions $f(x)$, $g(x)$, which is defined to be $\int_{-\infty}^{\infty} f(t)g(x-t)\ dt$.

We give a quick application of the operation of dividing a generating function by $1-x$; let us derive a formula for $S_\ell(n)$, the sum of the first n perfect ℓth powers. Since
$$\left(n^\ell\right)_{n\geq 0} \longleftrightarrow \sum_{k=1}^{\ell} k! \left\{ {\ell \atop k} \right\} \frac{x^k}{(1-x)^{k+1}},$$
we get
$$\left(S_\ell(n)\right)_{n\geq 0} \longleftrightarrow \sum_{k=1}^{\ell} k! \left\{ {\ell \atop k} \right\} \frac{x^k}{(1-x)^{k+2}}.$$
Extracting coefficients of x^n from both sides, we get the following.

Proposition 23.2. *For each $\ell \geq 1$ and each $n \geq 1$,*

$$1^\ell + 2^\ell + \ldots + n^\ell = \sum_{k=1}^{\ell} k! \left\{ {\ell \atop k} \right\} \binom{n+1}{k+1}.$$

This identity can also be proven in a combinatorial way:

- The left-hand side counts all pairs (f, i), where f is a map $[\ell] \to [n]$, and where i is an element of $[n+1]$ greater than each element of the image of f. (Indeed, for each given $i \in [n+1]$, the number of maps $f : [\ell] \to [n]$ having this property is $(i-1)^\ell$; thus, the total number of pairs (f, i) is $\sum_{i \in [n+1]} (i-1)^\ell = 0^\ell + 1^\ell + \cdots + n^\ell = 1^\ell + 2^\ell + \cdots + n^\ell$.)
- The right-hand side counts the same pairs, but organized in a different way. Namely, we can construct such a pair (f, i) as follows: First decide on the size k of the image $f([\ell])$. This k can range from 1 to ℓ. Next, we choose a subset of $[n+1]$ which will serve as the union $f([\ell]) \cup \{i\}$; this must be a $(k+1)$-element subset (since i must be greater than each element of $f([\ell])$, and therefore cannot belong to $f([\ell])$), and thus can be chosen in $\binom{n+1}{k+1}$ ways. Having chosen this subset, we immediately know that its highest element will be i, while its remaining k elements will form the image $f([\ell])$. It remains to choose f in such a way that the image $f([\ell])$ of f is precisely the set of these remaining k elements; in other words, f should be a surjection from $[\ell]$ to this k-element set. The number of such surjections is $k! \left\{ {\ell \atop k} \right\}$ (by Claim 14.4); thus, f can be chosen in $k! \left\{ {\ell \atop k} \right\}$ ways. Altogether, we thus conclude that the number of all pairs (f, i), where f is a map $[\ell] \to [n]$, and where i is an element of $[n+1]$ greater than each element of the image of f, equals $\sum_{k=1}^{\ell} \binom{n+1}{k+1} k! \left\{ {\ell \atop k} \right\} = \sum_{k=1}^{\ell} k! \left\{ {\ell \atop k} \right\} \binom{n+1}{k+1}$.

Comparing these two counts, we recover Proposition 23.2.

24. The Catalan numbers

Now we move on to a more substantial application of sequence convolution. We first consider three separate counting problems.

(1) Let t_n be the number of different triangulations of a convex $(n+2)$-gon on vertex set $\{1, \ldots, n+2\}$, where two triangulations are considered different if the sets of associated triples (the unordered sets of vertices of each of the triangles in the triangulations) are different.

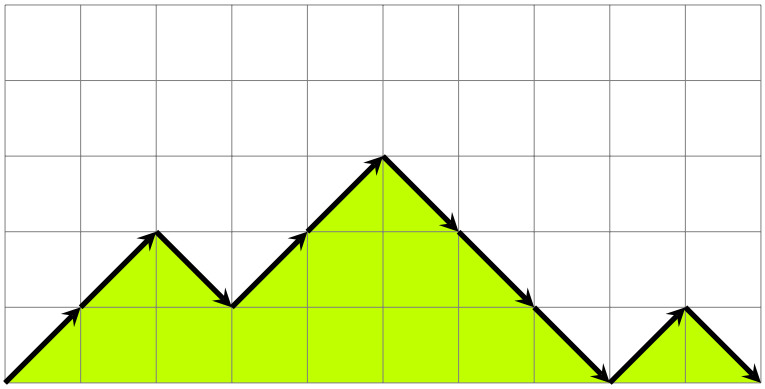
We have $t_1 = 1$, $t_2 = 2$ and $t_3 = 5$. For convenience we set $t_0 = 1$. The sequence $(t_n)_{n \geq 0}$ satisfies a non-linear recurrence. Fix $n \geq 1$, and fix one side, say 12, of the $(n+2)$-gon (the vertices are ordered 1 through $n+2$ in a clockwise direction). This side must belong to some triangle in the triangulation. If it belongs to triangle 123, then to complete the triangulation we must triangulate the $(n+1)$-gon with vertex set $[n+2] \setminus \{2\}$; there are t_{n-1} ways to do this. If it belongs to triangle 124, then to complete the triangulation we must independently triangulate the 3-gon with vertex set $\{2, 3, 4\}$, and the n-gon with vertex set $[n+2] \setminus \{2, 3\}$; there are $t_1 t_{n-2}$ ways to do this. Continuing around the $(n+2)$-gon, we find that

$$t_n = t_{n-1} + t_1 t_{n-2} + t_2 t_{n-3} + \ldots + t_{n-2} t_1 + t_{n-1} = \sum_{k=0}^{n-1} t_k t_{n-k-1}$$

(the last equality using $t_0 = 1$). This, valid for $n \geq 1$, together with the initial condition $t_0 = 1$, generates the full sequence.

(2) Let m_n be the number of words of length $2n$ over alphabet $\{U, D\}$, with the property that there word has n U's and n D's, and that every initial seqment (reading left-to-right) has at least as many U's as D's. These words are called *Dyck words*. For example, $UUDD$, $UDUD$ and $UDUUDD$ are Dyck words (and so is the empty word), whereas $UDDUUD$ is not (since the initial segment UDD has more D's than it has U's), and UUD is neither (since it has different numbers of U's and D's).

Dyck words are clearly in one-to-one correspondence with *Dyck paths*: paths in $\mathbb{R}^2$ that start at $(0, 0)$, proceed by taking steps of length $\sqrt{2}$ in the direction either of the vector $(1, 1)$ or of the vector $(1, -1)$, end on the x-axis, and never go below the axis. (The correspondence is given by mapping U's to steps in the direction of $(1, 1)$, and D's to $(1, -1)$.) I tend to think of Dyck paths as pictures of mountain ranges (hence the notation m_n for their number); for example, here is the Dyck path corresponding to the Dyck word $UUDUUDDDUD$:

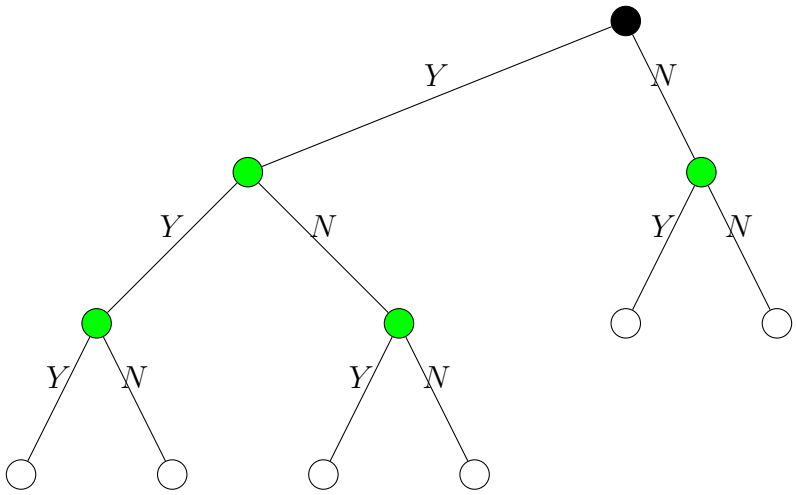


We have $m_0 = 1$, $m_1 = 1$, $m_2 = 2$ and $m_3 = 5$. We also have a recurrence: For any given $n > 0$, the set of mountain ranges (i.e., Dyck paths) ending at $(2n, 0)$ decomposes according to the value of k, $k = 0, \ldots, n-1$, such that $(2k, 0)$ is the last point (other than $(2n, 0)$) on the range to hit the x-axis. (For example, the Dyck path drawn above has $n = 5$ and $k = 4$.) There are m_k possible appearances for the mountain range up to $(2k, 0)$. Beyond $(2k, 0)$, the range must begin with a step parallel to $(1, 1)$, and then go to $(2n - 1, 1)$ before finally stepping down to $(2n, 0)$. Between $(2k + 1, 1)$ and $(2n - 1, 1)$ the range must not touch the x-axis, i.e., must not go below the parallel line with the equation $y = 1$; so, translating every point by $-(2k + 1, 1)$, this segment of the mountain range corresponds to a mountain range that starts at $(0, 0)$ and ends at $(2(n - k - 1), 0)$. There are m_{n-k-1} such ranges, so for $n \geq 1$,

$$m_n = \sum_{k=0}^{n-1} m_k m_{n-k-1}.$$

(3) A *binary decision tree* is a tree with one vertex, the root, distinguished, and some extra structure and properties. Namely, the root either has degree 0 or has degree 2, and if it has degree 2 then the two edges leaving the root are labeled Y and N (for "yes" and "no"). Each other vertex either has degree 1 or degree 3. For vertices with degree 3, the two edges that leave the vertex in the direction away from the root (i.e.,

the edges that don't lie on the path from the vertex to the root) are labeled Y and N. The vertices are not labeled, so (apart from the root) they are indistinguishable. Let b_n be the number of binary decision trees with n internal nodes. Here, an *internal node* means a vertex having degree 2 or 3. Here is an example of a binary decision tree with 5 internal nodes:



(The root is colored black, and the other four internal nodes are colored green.)

We have $b_0 = 1$, $b_1 = 1$, $b_2 = 2$ and $b_3 = 5$. We also have a recurrence: for $n \geq 1$, the root must have degree 2. We construct a binary decision tree with n internal nodes by selecting some k with $k = 0, \ldots, n-1$, then selecting binary decision trees with k and $n-k-1$ internal nodes, attaching the tree with k internal nodes to the endvertex of the Y edge from the root, and attaching the tree with $n-k-1$ internal nodes to the endvertex of the N edge from the root, with both attachments being made at the root of the tree being attached. It follows that for $n \geq 1$,

$$b_n = \sum_{k=0}^{n-1} b_k b_{n-k-1}.$$

Because all three counting problems lead to recursively defined sequences with the same initial conditions and recurrence relations, they have the same solution.

Definition 24.1. *The* Catalan numbers $(c_n)_{n \geq 0}$ *are defined by the recurrence*

$$c_n = \sum_{k=0}^{n-1} c_k c_{n-k-1}$$

for $n \geq 1$, *with initial condition* $c_0 = 1$.

Hence, $c_n = t_n = m_n = b_n$ for each $n \in \mathbb{N}$, where t_n, m_n and b_n have been introduced in the three counting problems above.

The Catalan numbers are probably the most frequently occurring sequence of numbers in combinatorics, after the binomial coefficients. In his book,[14] Stanley lists 66 different counting problems whose answer is "the Catalan numbers"; in an online addendum to the book[15], he extends this list to 223 interpretations!

[14] R. Stanley, *Enumerative Combinatorics*, Wadsworth & Brooks/Cole, 1986

[15] `http://www-math.mit.edu/~rstan/ec/catadd.pdf`

We can use generating functions, convolutions and the generalized binomial theorem to find an explicit expression for the nth Catalan number.

Theorem 24.2. *The nth Catalan number satisfies*

$$c_n = \frac{1}{n+1}\binom{2n}{n}.$$

Proof. Let $C(x) = \sum_{n\geq 0} c_n x^n$ be the generating function of the Catalan numbers. The recursive definition of the Catalan numbers then yields

$$\begin{aligned} C(x) &= 1 + (c_0c_0)x + (c_0c_1 + c_1c_0)x^2 + (c_0c_2 + c_1c_1 + c_2c_0)x^3 + \dots \\ &= 1 + xC^2(x) \end{aligned}$$

(where $C^2(x)$ means $(C(x))^2$, not $C(C(x))$). This is a quadratic equation in $C(x)$. Solving it, we get

$$C(x) = \frac{1 \pm \sqrt{1-4x}}{2x}.$$

So there are two possible solutions to the functional equation. We want one that satisfies $\lim_{x\to 0} C(x) = 1$, so

$$C(x) = \frac{1 - \sqrt{1-4x}}{2x}.$$

To find c_n from this, we first need to find the coefficient of x^{n+1} in $(1-4x)^{1/2}$; by the generalized binomial theorem, this is $(-1)^{n+1}4^{n+1}\binom{1/2}{n+1}$, and so

$$\begin{aligned} c_n &= (-1)^n 2^{2n+1}\binom{1/2}{n+1} \\ &= \frac{(-1)^n 2^{2n+1}(1(-1)(-3)\dots(-(2n-1)))}{2^{n+1}(n+1)!} \\ &= \frac{2^n(2n-1)(2n-3)\dots(5)(3)(1)}{(n+1)n!} \\ &= \frac{(2n)!}{(n+1)n!n!} \\ &= \frac{1}{n+1}\binom{2n}{n}. \end{aligned}$$

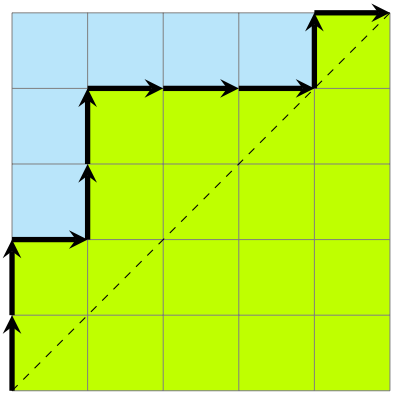
□

Notice two things: first, the derivation used no combinatorics, and second, it is completely valid for $|x| < 1/4$.

Generating functions have given us a nice, combinatorial-looking, formula for c_n. Is there a *combinatorial* proof? It will require some cleverness, as there is no obvious set whose size is $\binom{2n}{n}/(n+1)$ which we could put in bijection with something counted by the Catalan numbers; indeed, it is not even a priori obvious that $\binom{2n}{n}/(n+1)$ is an integer!

Here's one combinatorial proof, that uses the Dyck path interpretation of the Catalan numbers, modified slightly. A *staircase path* of length n is a path in $\mathbb{R}^2$ that starts at $(0,0)$,

ends at (n, n), takes steps of length 1 parallel either to $(1, 0)$ or $(0, 1)$, and never drops below the diagonal $x = y$. Here is an example of a staircase path of length 5:

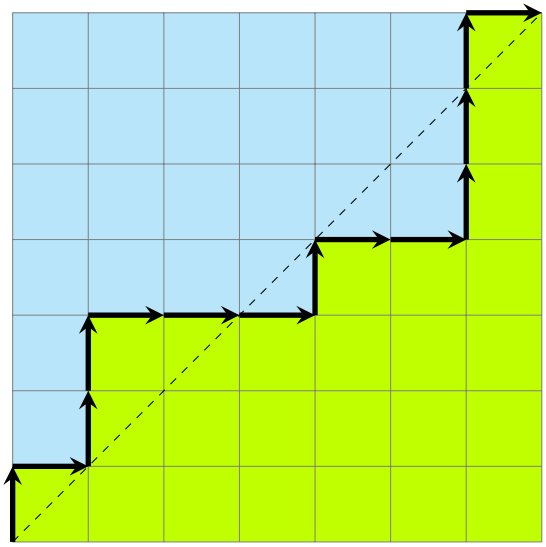


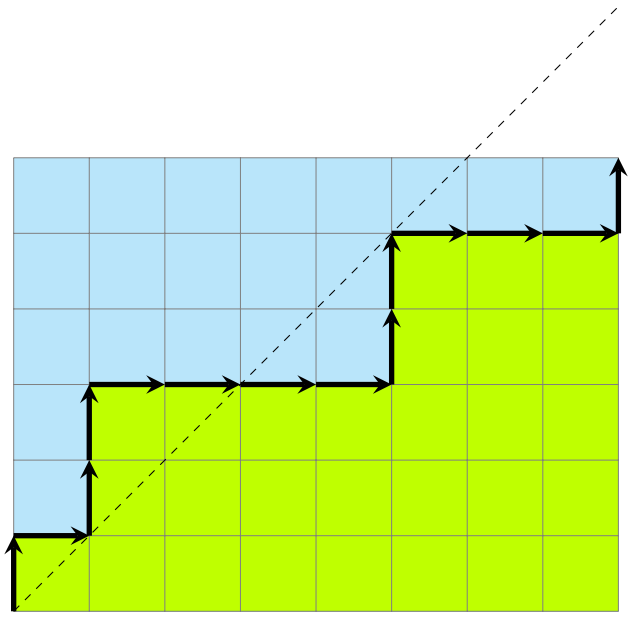
Clearly, staircase paths of length n are in one-to-one correspondence with Dyck words of length $2n$ (as defined in counting problem (2) above); namely, a staircase path p corresponds to the Dyck word whose i-th letter is $\begin{cases} U, & \text{if the } i\text{-th step of } p \text{ is in direction } (0,1); \\ D, & \text{if the } i\text{-th step of } p \text{ is in direction } (1,0) \end{cases}$. (For example, the staircase path drawn above corresponds to the Dyck word $UUDUUDDDUD$. Geometrically speaking, the correspondence between staircase paths and Dyck words is particularly simple: Given a Dyck word, draw the corresponding Dyck path, then rotate it by 45° counterclockwise and shrink it by the factor $\sqrt{2}$; the result will be the staircase path corresponding to the Dyck word.) Hence, the number of staircase paths of length n equals the number of Dyck words of length $2n$; but the latter is $m_n = c_n$, as we already know. Hence, the number of staircase paths is c_n. We may interpret the $2n + 1$ lattice points on a staircase path as the successive scores in a match between Arsenal and Liverpool that ends in an n-n tie, and in which Arsenal are never trailing. (Just consider any step in the direction $(1, 0)$ as a goal for Liverpool, and any step in the direction $(0, 1)$ as a goal for Arsenal.)

A staircase path is a special case of a lattice path (that we have seen earlier) — lattice paths do not carry the diagonal restriction (i.e., they are allowed to drop below the diagonal $x = y$). We consider here lattice paths that take n steps parallel to $(1, 0)$ and n steps parallel to $(0, 1)$. Say that such a path is *bad* if at some point it drops below the diagonal $x = y$. Let $\mathcal{B}_n$ be the set of bad lattice paths. Then, since there are $\binom{2n}{n}$ lattice paths (of the kind we are considering) in total, we have $c_n = \binom{2n}{n} - |\mathcal{B}_n|$. Armed with the knowledge that $c_n = \binom{2n}{n}/(n + 1)$, a little algebra shows that $|\mathcal{B}_n| = \binom{2n}{n+1}$.

We prove this fact about $|\mathcal{B}_n|$ combinatorially, using *André's reflection*. We know that $\mathcal{B}_n$ is the set of all lattice paths that start at $(0, 0)$ and end at (n, n) and pass below the main diagonal. Let $\mathcal{B}'_n$ be the set of all lattice paths that start at $(0, 0)$ and end at $(n + 1, n - 1)$ (with no restriction on their interaction with the diagonal). We have $|\mathcal{B}'_n| = \binom{2n}{n+1}$, so a bijection from $\mathcal{B}_n$ to $\mathcal{B}'_n$ will give $|\mathcal{B}_n| = \binom{2n}{n+1}$. For each $p \in \mathcal{B}_n$ there is a least k, with $0 \leq k \leq n - 1$, at which the path goes from (k, k) to $(k + 1, k)$ (this is the first time p passes below the main diagonal). Reflecting the portion of p at and beyond $(k + 1, k)$, where here and later in the paragraph the reflection is across the line through $(k+1, k)$ that is parallel to $x = y$, yields a new lattice path that ends at $((k+1)+(n-k), k+(n-k-1)) = (n+1, n-1)$.

For example,

(25) if $p =$  ,

then the new path is 

(in this example, we have $n = 7$ and $k = 3$, so the two paths diverge at point $(k+1, k) = (4, 3)$). The map that takes p to this new lattice path is therefore a map from $\mathcal{B}_n$ to $\mathcal{B}'_n$. It is evidently injective. For surjectivity, note that for each $p' \in \mathcal{B}'_n$ there is a least k, with $0 \leq k \leq n-1$, at which the path goes from (k, k) to $(k+1, k)$; reflecting the portion of p' at and beyond $(k+1, k)$ yields a path in $\mathcal{B}_n$ that is in the preimage of p'. Hence, we have found a bijection $\mathcal{B}_n \to \mathcal{B}'_n$, and therefore we have $|\mathcal{B}_n| = |\mathcal{B}'_n| = \binom{2n}{n+1}$, so that

$$c_n = \binom{2n}{n} - |\mathcal{B}_n| = \binom{2n}{n} - \binom{2n}{n+1} = \binom{2n}{n}/(n+1)$$

(by simple computations).

Notice that another way to describe the reflection in the previous paragraph is to say that beyond $(k+1, k)$, all steps parallel to $(1, 0)$ are replaced with steps parallel to $(0, 1)$, and vice-versa.

A slightly unsatisfactory aspect to this combinatorial proof is that it does not explain why $\binom{2n}{n}$ should be divisible by $n+1$ (its last step is computational). Here is a more complicated, but more satisfying, combinatorial argument that does give such an explanation.

Again we work with lattice and staircase paths. For a lattice path p that ends at (n,n), define the *exceedance* of p to be the number of horizontal steps that lie below the diagonal $x=y$ (this is the number of goals scored by Liverpool while they are either tied or in the lead). For instance, the path p in (25) has exceedance 3. Let $\mathcal{A}_i(n)$ be the set of paths that have exceedance i. We have

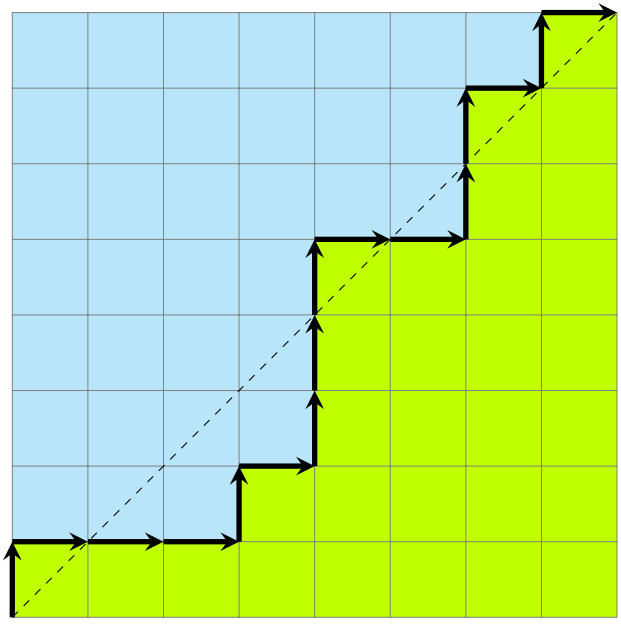
$$\sum_{i=0}^{n} |\mathcal{A}_i(n)| = \binom{2n}{n}$$

and $|\mathcal{A}_0(n)| = c_n$. We will construct, for each $i>0$, a bijection from $\mathcal{A}_i(n)$ to $\mathcal{A}_{i-1}(n)$. This will exhibit a decomposition of a set of $\binom{2n}{n}$ into $n+1$ equinumerous sets $\mathcal{A}_0(n), \mathcal{A}_1(n), \ldots, \mathcal{A}_n(n)$ (explaining combinatorially why $\binom{2n}{n}$ is divisible by $n+1$), and moreover show that

$$c_n = |\mathcal{A}_0(n)| = \frac{(n+1)|\mathcal{A}_0(n)|}{n+1} = \frac{\sum_{i=0}^{n} |\mathcal{A}_i(n)|}{n+1} = \frac{\binom{2n}{n}}{n+1}.$$

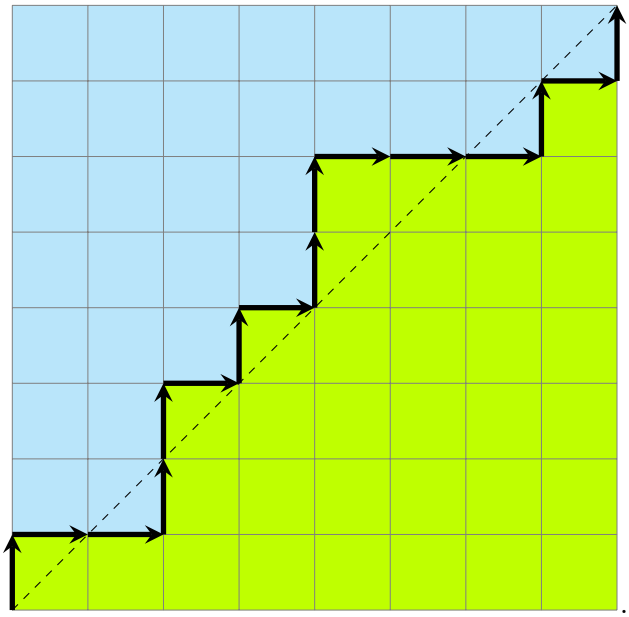
Each $p \in \mathcal{A}_i(n)$ can be written as the concatenation $p_1p_2p_3$, where p_1 is a lattice path ending at some point of the form $(k,k-1)$ $(1 \le k \le n)$, p_2 is the step from $(k,k-1)$ to (k,k), p_3 is a path starting at (k,k) and ending at (n,n), and the point (k,k) is the *first* point at which p hits the diagonal coming from below (here we use $1 \le i \le n$). For example, if p is the lattice path

(26)



then p_1 comprises the first 7 steps of p, whereas p_2 consists of the next (i.e., 8-th) step, and p_3 comprises the remaining 8 steps of p (and we have $k=4$). Consider now the path $p_3p_2p_1$ (more correctly, the path that begins with the translate of p_3 by the vector $(-k,-k)$, and continues with the necessary translates of p_2 and p_1 to make it continuous). For instance, if

p is as in (26), then $p_3p_2p_1$ is the lattice path



We leave it to the reader to verify the following facts about the map f that sends $p_1p_2p_3$ to $p_3p_2p_1$:

- $f(p_1p_2p_3) \in \mathcal{A}_{i-1}(n)$ (the main point here being that the only horizontal edge of p below the diagonal that does not correspond to a horizontal edge of $f(p)$ below the diagonal, is the first horizontal edge of p below the diagonal);
- $f : \mathcal{A}_i(n) \to \mathcal{A}_{i-1}(n)$ is an injection; and
- $f : \mathcal{A}_i(n) \to \mathcal{A}_{i-1}(n)$ is a surjection (the main point here is that $p' \in \mathcal{A}_{i-1}(n)$ can be written as the concatenation $p'_1p'_2p'_3$, where p'_1 is a lattice path ending at some point of the form (k, k) $(0 \leq k \leq n-1)$, p'_2 is the step from (k, k) to $(k, k+1)$, p'_3 is a path starting at $(k, k+1)$ and ending at (n, n), and the point (k, k) is the *last* point at which p' leaves the diagonal moving up (here we use $0 \leq i-1 \leq n-1$); then $p'_3p'_2p'_1$ is in $\mathcal{A}_i(n)$, and is in the preimage of p').

We conclude that for $1 \leq i \leq n$, $|\mathcal{A}_i(n)| = |\mathcal{A}_{i-1}(n)|$.

25. Some problems

(1) $2n$ people sit around a circular table. Let h_n be the number of ways that they can pair off into n pairs, in such a way that the n pairs can shake hands simultaneously without there being any pair of handshakers with crossing hands. For example, if $n = 3$ and the six people are a, b, c, d, e, f, in that order, there are five possible pairings: $\{ab, cd, ef\}$, $\{ab, cf, de\}$, $\{ad, bc, ef\}$, $\{af, bc, de\}$, $\{af, be, cd\}$ (an arrangement like $\{ab, ce, df\}$ is forbidden, since ce and df would have crossing hands). With $h_0 = 1$ by definition, show that h_n is the nth Catalan number.

Solution: Number the $2n$ cyclically $1, 2, \ldots, 2n$. Person 1 can shake hands with any of: person $2, 4, 6, \ldots, 2n$ (otherwise, if 1 shakes hands with person k for odd k, there are an odd number of people on each side of the line joining 1 to k, and so no legal configuration is possible).

If person 1 shakes hands with person $2k$, $k = 1, \ldots, n$, then there are $2k-2$ people on one side of the line joining 1 to $2k$, and these people must shake hands among themselves to avoid crossing the 1-k line; they can do this in h_{k-1} ways. There are $2n-2k$ people on the other side of the line joining 1 to k, and these people must shake hands among themselves to avoid crossing the 1-k line; they can do this in h_{n-k} ways. Any arrangement on one side of the 1-k line can be coupled with any arrangement on the other side to form a valid configuration, and all configurations are obtained in this way. It follows that

$$h_n = \sum_{k=1}^{n} h_{k-1} h_{n-k}$$

for $n \geq 1$, which is exactly the Catalan recurrence.

(2) An arrangement of the numbers 1 through n in a row is said to be 231-*avoiding* if it is not the case that are three numbers a, b, c occurring in that order (not necessarily consecutively) with $b > a > c$. For example, 7654321 is a 231-avoiding arrangement of $[n]$, but 73654231 is not, since $3, 5, 1$ form a "231" pattern.

Show that a_n, the number of 231-avoiding arrangements of $[n]$, is the nth Catalan number.

Solution: Evidently $a_0 = 1$. For $n \geq 1$, consider the position of the number n in the row: it goes in position k, $k = 1, \ldots, n$. Every number that goes to the left of n must be smaller than every number that goes to the right, otherwise we would have a 231 pattern, so the numbers $1, \ldots, k-1$ must go to the left, and $k, \ldots, n-1$ to the right. The numbers $1, \ldots, k-1$ must be arranged in a 231-avoiding arrangement, and so must the numbers $k, \ldots, n-1$; any pair of such arrangements can be combined to form a 231-avoiding arrangement of $1, \ldots, n$; and all arrangements arise by this process. It follows that

$$a_n = \sum_{k=1}^{n} a_{k-1} a_{n-k}$$

for $n \geq 1$, which is exactly the Catalan recurrence.

(3) In an election between two candidates N and M, N gets n votes and M gets m votes, $m \geq n$. In how many ways can the $n+m$ votes be ordered, so that if the votes were counted in that order, M would never trail N in the count? (Votes for the same candidate are indistinguishable, so it does not matter in what order they are counted.)

(If $m = n$, the answer is the nth Catalan number. Generalize the reflection principle we saw in the first combinatorial proof of the Catalan number formula. Don't forget the reality check: if your answer doesn't reduce to $\binom{2n}{n}/(n+1)$ when $m = n$ then something is wrong.)

26. Some examples of two-variable generating functions

We can use the method of generating functions to analyze two- (or three-, or four-) variable recurrences. We won't form any kind of general theory here, just present some illustrative examples.

27. Binomial coefficients

We begin with the binomial coefficients. For $n, k \geq 0$, let $b_{n,k}$ denote the number of subsets of size k of a set of size n. We have a recurrence relation, Pascal's identity — $b_{n,k} = b_{n-1,k-1} + b_{n-1,k}$ — valid for $n, k \geq 1$, and initial conditions $b_{n,0} = 1$ (for $n \geq 0$) and $b_{0,k} = 0$ (for $k \geq 1$). We form the two-variable generating function $B(x,y) = \sum_{n\geq 0, k\geq 0} b_{n,k} x^n y^k$. Using the method of "use recurrence where possible, initial conditions where necessary", we get

$$\begin{aligned} B(x,y) &= \sum_{n\geq 0} x^n + \sum_{n\geq 1, k\geq 1} (b_{n-1,k-1} + b_{n-1,k})\, x^n y^k \\ &= \frac{1}{1-x} + xy \sum_{n\geq 1,k\geq 1} b_{n-1,k-1} x^{n-1} y^{k-1} + x \sum_{n\geq 1,k\geq 1} b_{n-1,k} x^{n-1} y^k \\ &= \frac{1}{1-x} + xyB(x,y) + x\left(B(x,y) - \frac{1}{1-x}\right) \\ &= 1 + x(y+1)B(x,y), \end{aligned}$$

so

$$B(x,y) = \frac{1}{1-x(y+1)}. \tag{27}$$

One thing we can do immediately with this is to extract the coefficient of x^n from both sides, to get

$$\sum_{k\geq 0} b_{n,k} y^k = (y+1)^n,$$

the binomial theorem! The coefficient of y^k on the right-hand side is, by Taylor's theorem, $(n)_k/k!$, and so we recover the familiar algebraic expression of the binomial coefficients,

$$b_{n,k} = \frac{(n)_k}{k!}.$$

We could also try to extract the coefficient of y^k directly from both sides of (27). Rewriting the right-hand side of (27) as

$$\frac{1}{1-x}\left(\frac{1}{1-\left(\frac{x}{1-x}\right)y}\right)$$

we get

$$\sum_{n\geq 0} b_{n,k} x^n = \frac{x^k}{(1-x)^{k+1}},$$

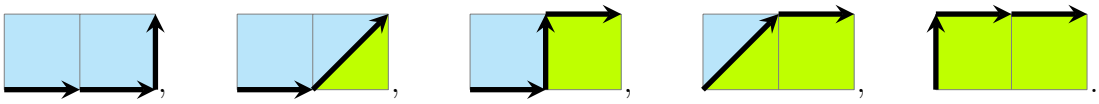
an identity that was key for solving linear recurrence relations.

28. Delannoy numbers

As a second simple example, we mention the *Delannoy numbers*.

Definition 28.1. *For $n, m \geq 0$, we define the* Delannoy number $d_{n,m}$ *to be the number of paths from $(0,0)$ to (n,m) in which each step of the path is either a step up (U) one unit [length 1, parallel to $(0,1)$], across (A) one unit [length 1, parallel to $(1,0)$], or diagonal (D) one unit [length $\sqrt{2}$, parallel to $(1,1)$]. Paths of this kind are called* Delannoy paths.

For example, $d_{2,1} = 5$, the five Delannoy paths being AAU, AD, AUA, DA and UAA. Here are these five paths drawn in the plane:



There is an obvious recurrence relation satisfied by the Delannoy numbers, that is the analog of the lattice path recurrence $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$, namely

$$d_{n,m} = d_{n-1,m} + d_{n,m-1} + d_{n-1,m-1}$$

valid when $m, n \geq 1$, with initial conditions $d_{n,0} = d_{0,m} = 1$ for all $m, n \geq 0$.

Setting $D(x,y) = \sum_{n,m\geq 0} d_{n,m} x^n y^m$ we get the functional equation

$$\begin{aligned} D(x,y) &= \frac{1}{1-x} + \frac{y}{1-y} + \sum_{n,m\geq 1} (d_{n-1,m} + d_{n,m-1} + d_{n-1,m-1}) x^n y^m \\ &= \frac{1}{1-x} + \frac{y}{1-y} + x\left(D(x,y) - \frac{1}{1-x}\right) + y\left(D(x,y) - \frac{1}{1-y}\right) \\ &= 1 + (x + y + xy) D(x,y), \end{aligned}$$

so that

$$\begin{aligned} D(x,y) &= \frac{1}{1-x-y-xy} \\ &= \frac{1}{1-y}\left(\frac{1}{1-\left(\frac{1+y}{1-y}\right)x}\right) \\ &= \sum_{n\geq 0} \frac{(1+y)^n}{(1-y)^{n+1}} x^n \\ &= \sum_{n\geq 0} \sum_{k\geq 0} \binom{n}{k} \frac{y^k}{(1-y)^{n+1}} x^n. \end{aligned}$$

It follows that $d_{n,m}$ is the coefficient of y^m in $\sum_{k\geq 0} \binom{n}{k} y^k/(1-y)^{n+1}$. Now the coefficient of y^m in $y^k/(1-y)^{n+1}$ is the same as the coefficient of y^{m-k} in $1/(1-y)^{n+1}$, which is[16] $\binom{n+m-k}{n}$, so

$$d_{n,m} = \sum_{k\geq 0} \binom{n}{k} \binom{n+m-k}{n}.$$

Could we have seen this formula combinatorially? Yes! To get from $(0,0)$ to (n,m) via a Delannoy path, it is necessary to take k diagonal steps for some $k \geq 0$, $n-k$ horizontal steps, and $m-k$ vertical steps, for a total of $n+m-k$ steps. We completely specify a Delannoy path with k diagonal steps by specifying first the locations, in the list of $n+m-k$ steps, of the n steps that are either diagonal or horizontal ($\binom{n+m-k}{n}$ choices), and then specifying the locations, in the list of n steps that are either diagonal or horizontal, of the k steps that are diagonal ($\binom{n}{k}$ choices); so $d_{n,m} = \sum_{k\geq 0} \binom{n}{k}\binom{n+m-k}{n}$.

[16]Here we are assuming that $k \leq n$ (so that $n+m-k$ is nonnegative). The case when $k > n$ is irrelevant, since $\binom{n}{k} = 0$ in this case (causing the whole addend to vanish).

The Delannoy numbers are connected to another set of numbers that at first blush seem completely unrelated. Let $B_n(m)$ be the set of all lattice points in $\mathbb{Z}^n$ with the property that the sum of the absolute values of the coordinate values is at most m (i.e.,

$$B_n(m) = \{(x_1, \ldots, x_n) : x_i \in \mathbb{Z} \text{ for each } i,\ \sum_{i=1}^{n} |x_i| \leq m\}).$$

This set $B_n(m)$ is the *Hamming ball* or *ℓ_1-norm ball* in $\mathbb{Z}^n$ of radius m. Let $b_{n,m} = |B_n(m)|$.

Claim 28.2. *For all $n \geq 0$ and $m \geq 0$, we have $b_{n,m} = \sum_{k=0}^{n} \binom{n}{k}\binom{m}{k}2^k = \sum_{k\geq 0} \binom{n}{k}\binom{m}{k}2^k$, and so $b_{n,m} = b_{m,n}$ (that is, there are exactly as many points in $\mathbb{Z}^n$ at ℓ_1-distance at most m from the origin, as there are points in $\mathbb{Z}^m$ at ℓ_1-distance at most n from the origin!).*

Proof. We can specify a point in $B_n(m)$ by first deciding k, the number of non-zero coordinates in the point ($k = 0, \ldots, n$), then deciding which k coordinates are non-zero ($\binom{n}{k}$ options), then deciding the signs of the non-zero coordinates (2^k options), and then deciding the absolute values of the non-zero coordinates. For this last problem, we have to specify (in order) k positive integers that sum to at most m; this is the same as specifying $k+1$ positive integers that sum to exactly $m+1$ (the $(k+1)$st being a dummy to bring the sum up to $m+1$), which in turn is the same as specifying a composition of $m+1$ into $k+1$ parts; there are $\binom{m}{k}$ such. It follows that

$$b_{n,m} = \sum_{k=0}^{n} \binom{n}{k} 2^k \binom{m}{k} = \sum_{k=0}^{n} \binom{n}{k}\binom{m}{k} 2^k$$

as claimed. We can re-write this as

$$b_{n,m} = \sum_{k=0}^{\infty} \binom{n}{k}\binom{m}{k} 2^k$$

(using $\binom{n}{k} = 0$ for $k \geq n+1$). It is now clearly symmetric in m and n, so $b_{n,m} = b_{m,n}$. □

What has this to do with Delannoy numbers? Let us form the two-variable generating function of the numbers $b_{n,m}$:

$$\begin{aligned}
B(x,y) &= \sum_{n,m,k\geq 0} \binom{n}{k}\binom{m}{k} 2^k x^n y^m \\
&= \sum_{k\geq 0} 2^k \left(\sum_{n\geq 0} \binom{n}{k} x^n\right)\left(\sum_{m\geq 0} \binom{m}{k} y^m\right) \\
&= \sum_{k\geq 0} \frac{(2xy)^k}{(1-x)^{k+1}(1-y)^{k+1}} \\
&= \frac{1}{(1-x)(1-y)} \left(\frac{1}{1-\frac{2xy}{(1-x)(1-y)}}\right) \\
&= \frac{1}{1-x-y-xy} \\
&= D(x,y).
\end{aligned}$$

So in fact, the numbers $b_{n,m}$ and $d_{n,m}$ coincide!

This tells us that the numbers $b_{n,m}$ satisfy the recurrence

$$b_{n,m} = b_{n-1,m} + b_{n,m-1} + b_{n-1,m-1},$$

a fact that is not completely obvious; it also suggests that there should be a bijection from Delannoy paths to (n, m) to points in $B_n(m)$. Further, the above discussion suggests the existence of a bijection from points in $B_n(m)$ to points in $B_m(n)$. Finding all these bijections is left to the reader!

29. Some problems

(1) Show combinatorially that for $n, m \geq 1$ we have

$$b_{n,m} = b_{n-1,m} + b_{n,m-1} + b_{n-1,m-1}.$$

(That is, show that the Hamming ball of radius m in $\mathbb{Z}^n$ can be decomposed into three disjoint parts, one of which is in bijection with the Hamming ball of radius m in $\mathbb{Z}^{n-1}$, another of which is in bijection with the Hamming ball of radius $m-1$ in $\mathbb{Z}^n$, and the last of which is in bijection with the Hamming ball of radius $m-1$ in $\mathbb{Z}^{n-1}$).

Solution: $B_n(m)$ naturally decomposes into $B_n(m-1)$ and $D_n(m)$, where $D_n(m)$ is the set of all integer points in $\mathbb{Z}^n$, the sum of whose coordinates, in absolute value, is exactly m. In $D_n(m)$, some points have non-negative first coordinate; these points correspond bijectively with $B_{n-1}(m)$, via the map that removes the first coordinate. The remaining points have strictly negative first coordinate; these points correspond bijectively with $B_{n-1}(m-1)$, via the same map. The claimed recurrence follows.

(2) Exhibit a bijection from $D_{n,m}$ to $B_n(m)$, where $D_{n,m}$ is the set of Delannoy paths to (n, m).

Solution: We exhibit a bijection from Delannoy paths to (n, m) to points in $B_n(m)$ (ball of radius m in $\mathbb{Z}^n$). From the path P, form the integer list $(b_1, \ldots, b_n)$ with $\sum |b_i| \leq m$: namely, $|b_i|$ is the number of vertical or diagonal steps from the point of first reaching the line $x = i-1$ to first reaching the line $x = i$; let the sign of b_i be $+$ if the final step to the line $x - i$ is horizontal, and $-$ if it is diagonal. This gives a function from paths to points in the ball. A little thought shows that this is a bijection (note that as a path goes from $x = i-1$ to $x = i$, it must follow a string of verticals followed by either one diagonal or one horizontal; so the number-sign combination completely determines the shape of the path).

(3) Exhibit a bijection from $B_n(m)$ to $B_m(n)$.

Solution: We proceed in several steps.

Step 1. For every $k \geq 0$, we let $B_n(m; k)$ denote the set of all lattice points $p \in B_n(m)$ such that exactly k coordinates of p are nonzero. Furthermore, for every $k \geq 0$, we let $B_n^+(m; k)$ denote the set of all lattice points $p \in B_n(m; k)$ whose coordinates are all nonnegative (so that exactly k of them are positive, and the remaining $n-k$ are zero). Finally, for every $k \geq 0$, we let $S_k(n)$ be the set of all k-element subsets of $[n]$.

Step 2. For every $s \geq 0$, there is a natural bijection between subsets of $\{1, \ldots, s\}$ of size k, and solutions to $a_1 + a_2 + \ldots + a_k \leq s$ in positive integers $a_1, a_2, \ldots, a_k$: if the subset is $\{b_1, \ldots, b_k\}$ with $b_1 < b_2 < \ldots < b_k$, then the corresponding solution is given by $a_1 = b_1$, $a_2 = b_2 - b_1$, $a_3 = b_3 - b_2$, etc.. Conversely, any solution $(a_1, a_2, \ldots, a_k)$ of $a_1 + a_2 + \ldots + a_k \leq s$ in positive integers gives rise to

the subset $\{a_1, a_1+a_2, a_1+a_2+a_3, \ldots, a_1+a_2+\cdots+a_k\}$, which we shall denote by $P(a_1, a_2, \ldots, a_k)$.

Step 3. Now, fix $k \geq 0$. There is a bijection $\beta_{n,m} : B_n^+(m;k) \to S_k(n) \times S_k(m)$. This bijection $\beta_{n,m}$ takes a lattice point $(p_1, p_2, \ldots, p_n) \in B_n^+(m;k)$ to the pair (U, V), where $U = \{u_1 < u_2 < \cdots < u_k\}$ is the subset $\{i \in [n] \mid p_i > 0\}$ of $[n]$ (this is the set of all positions in which our lattice point has a nonzero coordinate), and where V is the subset $P(p_{u_1}, p_{u_2}, \ldots, p_{u_k})$ of $[m]$ (which, as we know from Step 2, uniquely determines the k-tuple $(p_{u_1}, p_{u_2}, \ldots, p_{u_k})$ and therefore the positive coordinates of the lattice point). For example, if $n = 5$, $m = 7$ and $k = 3$, then our bijection $\beta_{n,m}$ sends the lattice point $(0, 3, 2, 0, 2) \in B_n^+(m;k)$ to $(\{2, 3, 5\}, \{3, 5, 7\}) \in S_k(n) \times S_k(m)$, because the nonzero coordinates of the lattice point appear in positions $2, 3, 5$ and because $P(3, 2, 2) = \{3, 5, 7\}$.

Step 4. There is furthermore a bijection $\gamma_{n,m} : B_n(m;k) \to B_n^+(m;k) \times \{-1, 1\}^k$ which sends any lattice point $p = (p_1, p_2, \ldots, p_n) \in B_n(m;k)$ to the pair $((|p_1|, |p_2|, \ldots, |p_n|), (s_1, s_2, \ldots, s_k)) \in B_n^+(m;k) \times \{-1, 1\}^k$, where $s_i = 1$ if the i-th nonzero entry of p is positive and $s_i = -1$ otherwise. For example, if $n = 5$, $m = 7$ and $k = 3$, then our bijection $\gamma_{n,m}$ sends the lattice point $(0, -3, 2, 0, -2) \in B_n(m;k)$ to $((0, 3, 2, 0, 2), (-1, 1, -1)) \in B_n^+(m;k) \times \{-1, 1\}^k$.

Step 5. In Step 3, we have found a bijection $\beta_{n,m} : B_n^+(m;k) \to S_k(n) \times S_k(m)$. It clearly gives rise to a bijection $\beta_{n,m} \times \mathrm{id} : B_n^+(m;k) \times \{-1, 1\}^k \to S_k(n) \times S_k(m) \times \{-1, 1\}^k$. Composing this with the bijection $\gamma_{n,m}$ from Step 4, we obtain a bijection

$$\delta_{n,m} = (\beta_{n,m} \times \mathrm{id}) \circ \gamma_{n,m} : B_n(m;k) \to S_k(n) \times S_k(m) \times \{-1, 1\}^k.$$

The same argument, with n and m interchanged, yields a bijection

$$\delta_{m,n} : B_m(n;k) \to S_k(m) \times S_k(n) \times \{-1, 1\}^k.$$

But of course, there is also a bijection

$$\chi_{n,m} : S_k(n) \times S_k(m) \times \{-1, 1\}^k \to S_k(m) \times S_k(n) \times \{-1, 1\}^k,$$
$$(U, V, s) \mapsto (V, U, s).$$

These three bijections form a diagram

$$\begin{array}{ccc} B_n(m;k) & \xrightarrow{\delta_{n,m}} & S_k(n) \times S_k(m) \times \{-1, 1\}^k \\ & & \downarrow{\scriptstyle \chi_{n,m}} \\ B_m(n;k) & \xrightarrow[\delta_{m,n}]{} & S_k(m) \times S_k(n) \times \{-1, 1\}^k \end{array}.$$

Composing the arrows in the appropriate order, we obtain a bijection $B_m(n;k) \to B_n(m;k)$.

Step 6. Thus we have found a bijection $B_m(n;k) \to B_n(m;k)$ for each $k \geq 0$. These bijections, taken all together, form a bijection $B_m(n) \to B_n(m)$ (since the set $B_m(n)$ is decomposed into its disjoint subsets $B_m(n;k)$ for $k \geq 0$, and since the set $B_n(m)$ is decomposed into its disjoint subsets $B_n(m;k)$ for $k \geq 0$).

30. Stirling numbers of the second kind

We now move on to a more complicated example, the Stirling numbers of the second kind. Let $s_{n,k}$ denote the number of partitions of $[n]$ into k (non-empty) blocks. Again we have a recurrence relation

$$s_{n,k} = s_{n-1,k-1} + k s_{n-1,k}$$

valid for $n, k \geq 1$, and initial conditions $s_{n,0} = s_{0,k} = 0$ (for $n, k \geq 1$) and $s_{0,0} = 1$. We form the two-variable generating function $S(x,y) = \sum_{n \geq 0, k \geq 0} s_{n,k} x^n y^k$. Using the method of "use recurrence where possible, initial conditions where necessary", we get

$$\begin{aligned} S(x,y) &= 1 + \sum_{n \geq 1, k \geq 1} \left(s_{n-1,k-1} + k s_{n-1,k} \right) x^n y^k \\ &= 1 + xy \sum_{n \geq 1, k \geq 1} s_{n-1,k-1} x^{n-1} y^{k-1} + x \sum_{n \geq 1, k \geq 1} k s_{n-1,k} x^{n-1} y^k \\ &= 1 + xyS(x,y) + xy \sum_{n \geq 0, k \geq 1} k s_{n,k} x^n y^{k-1} \\ &= 1 + xyS(x,y) + xy \frac{\partial}{\partial y} S(x,y) \end{aligned} \tag{28}$$

or

$$\frac{\partial}{\partial y} S(x,y) = \left(\frac{1}{xy} - 1 \right) S(x,y) - \frac{1}{xy}.$$

This functional equation, being a non-linear two-variable partial differential equation, is not as easy to solve as the equivalent equation for the sequence of binomial coefficients, but we can still work with it. For example, we can attempt to extract the coefficient of y^k from both sides of (28).

Set, for $k \geq 0$, $B_k(x) = \sum_{n \geq 0} s_{n,k} x^n$, so that

$$S(x,y) = \sum_{k \geq 0} B_k(x) y^k$$

and

$$\frac{\partial}{\partial y} S(x,y) = \sum_{k \geq 0} (k+1) B_{k+1}(x) y^k,$$

and, from (28),

$$B_k(x) = x B_{k-1}(x) + x k B_k(x)$$

for $k \geq 1$. This leads to the recurrence

$$B_k(x) = \frac{x B_{k-1}(x)}{1 - kx}$$

for $k \geq 1$, with initial condition $B_0(x) = 1$. This is easily solved explicitly:

$$B_k(x) = \frac{x^k}{(1-x)(1-2x)(1-3x)\ldots(1-kx)}. \tag{29}$$

In other words, for each fixed k, the sequence $(s_{n,k})_{n \geq 0}$ has a rational generating function, which can (in principle at least) be tackled by the method of partial fractions.

The easiest way to do the partial fraction analysis is as follows. There are constants $A_1, \ldots, A_k$ such that

$$\frac{1}{(1-x)(1-2x)(1-3x)\ldots(1-kx)} = \sum_{r=1}^{k} \frac{A_r}{1-rx}. \tag{30}$$

Multiplying through by $(1-x)(1-2x)(1-3x)\ldots(1-kx)$, we obtain

$$1 = \sum_{r=1}^{k} A_r(1-x)\ldots\widehat{(1-rx)}\ldots(1-kx).$$

For any given $s \in [k]$, we can evaluate both sides of this equality at $x = 1/s$, which causes all but one addend in the sum to vanish (since the factor $1-sx$ becomes 0); thus, the equality becomes

$$\begin{aligned} 1 &= A_s(1-1/s)\ldots\widehat{(1-s/s)}\ldots(1-k/s) = A_s \cdot \frac{(s-1)\ldots\widehat{(s-s)}\ldots(s-k)}{s^{k-1}} \\ &= A_s \cdot \frac{(s-1)!\cdot(-1)^{k-s}\,(k-s)!}{s^{k-1}}, \end{aligned}$$

so that

$$A_s = (-1)^{k-s}\frac{s^{k-1}}{(s-1)!(k-s)!}.$$

Hence, (30) becomes

$$\frac{1}{(1-x)(1-2x)(1-3x)\ldots(1-kx)} = \sum_{r=1}^{k}(-1)^{k-r}\frac{r^{k-1}}{(r-1)!(k-r)!}\cdot\frac{1}{1-rx}. \tag{31}$$

Now we can extract the coefficient of x^n from both sides of (29):

$$\begin{aligned} s_{n,k} &= [x^n]\left(\frac{x^k}{(1-x)(1-2x)(1-3x)\ldots(1-kx)}\right) \\ &= [x^{n-k}]\left(\frac{1}{(1-x)(1-2x)(1-3x)\ldots(1-kx)}\right) \\ &= \sum_{r=1}^{k}(-1)^{k-r}\frac{r^{k-1}}{(r-1)!(k-r)!}r^{n-k} \\ &= \sum_{r=1}^{k}(-1)^{k-r}\frac{r^n}{r!(k-r)!}. \end{aligned}$$

This is just a minor re-writing of (19), this time derived purely analytically.

We could also have attempted to extract the coefficient of x^n from both sides of the functional equation for $S(x,y)$. Here it is convenient to set $A_n(y) = \sum_{k\geq 0} s_{n,k}y^k$, so that the functional equation becomes

$$\sum_{n\geq 0} A_n(y)x^n = 1 + xy\sum_{n\geq 0} A_n(y)x^n + xy\sum_{n\geq 0}\frac{d}{dy}\left(A_n(y)\right)x^n.$$

At $n = 0$ we get $A_0(y) = 1$, and for $n \geq 1$ we get the recurrence

$$A_n(y) = yA_{n-1}(y) + y\frac{d}{dy}A_{n-1}(y) = (y + yD_y)\,A_{n-1}(y) \tag{32}$$

(where D_y is differentiation with respect to y) and so, by induction,

$$A_n(y) = (y + yD_y)^n\,1.$$

From this we can immediately see that $A_n(y)$ is a polynomial of degree n in y, so that $s_{n,k} = 0$ for $k > n$.

31. Unimodality, log-concavity and asymptotic normality

We use (32) as an illustrative example in a digression into unimodality, log-concavity and asymptotic normality.

Very often we cannot understand the terms of a combinatorially defined sequence explicitly, so we must resort to understanding qualitative behavior; perhaps the asymptotic growth rate of the terms of the sequence, or perhaps (as in this section) the rough "shape" of the sequence. A property shared by many combinatorial sequences that relates to rough shape is unimodality: the property of (essentially) rising monotonely to a single peak, and then falling away monotonely.

Definition 31.1. *A sequence $(a_n)_{n\geq 0}$ is said to be* unimodal*, with* mode *m, if*

$$a_0 \leq a_1 \leq \ldots \leq a_{m-1} \leq a_m \geq a_{m+1} \geq \ldots.$$

Note that the mode may be 0 (in which case the sequence is monotonically decreasing) or infinity (in which case the sequence is monotonically increasing); note also that the mode may not be unique (for example, for a constant sequence every index is a mode).

It's easy to check algebraically that for even $n = 2m \geq 0$

$$\binom{n}{0} < \ldots < \binom{n}{m-1} < \binom{n}{m} > \binom{n}{m+1} > \ldots > \binom{n}{n},$$

while for odd $n = 2m + 1 \geq 0$

$$\binom{n}{0} < \ldots < \binom{n}{m} = \binom{n}{m+1} > \ldots > \binom{n}{n},$$

so that the sequence of binomial coefficients is unimodal with unique mode $n/2$ (if n is even), and modes $(n-1)/2$ and $(n+1)/2$ (if n is odd). It's far from clear how to show this combinatorially: what natural function maps the subsets of $\{1, \ldots, 17\}$ of size 6 into the subsets of $\{1, \ldots, 17\}$ of size 7 injectively, for example?

A little experimentation suggests that the sequence $(\left\{{n \atop k}\right\})_{k\geq 0}$ is unimodal for each $n \geq 0$, but here an algebraic proof seems out of reach, as none of our expressions for $\left\{{n \atop k}\right\}$ seem particularly amenable to an analysis focussing on inequalities. A combinatorial proof seems out of reach too, as this would require describing a map from the partitions of $[n]$ into k non-empty blocks, to the partitions of $[n]$ into $k+1$ non-empty blocks, for each $k = 0, \ldots, n-1$, that is injective for all $k \leq m$ for some m, and surjective for all $k > m$; but it is far from clear what this value of m should be, at which the functions flip from injectivity to surjectivity. (It is known that $m \sim n/\ln n$, but no exact formula is known; nor is it known whether in general the sequence of Stirling numbers of the second kind has a unique mode).

An approach to unimodality is through the stronger property of *log-concavity*.

Definition 31.2. *A sequence* $(a_n)_{n\geq 0}$ *is said to be* log-concave *if for each* $k \geq 1$

$$a_k^2 \geq a_{k-1}a_{k+1}.$$

Note that this is the same as saying that $\log a_k \geq (\log a_{k-1} + \log a_{k+1})/2$, that is, that the sequence $(\log a_k)_{k\geq 0}$ is concave; hence the terminology. Note also that if all the terms are strictly positive, log-concavity is equivalent to the monotonicity of successive ratios:

$$\frac{a_1}{a_0} \geq \frac{a_2}{a_1} \geq \frac{a_3}{a_2} \geq \dots,$$

while if a_k is positive for all $k = 0, \dots, n$, and zero elsewhere, then log-concavity is equivalent to

$$\frac{a_1}{a_0} \geq \frac{a_2}{a_1} \geq \frac{a_3}{a_2} \geq \dots \geq \frac{a_n}{a_{n-1}}.$$

Proposition 31.3. *If a log-concave sequence* $(a_k)_{k\geq 0}$ *is such that* a_k *is positive for all* $k = 0, \dots, n$*, and zero elsewhere, then it is unimodal.*

Proof. If the sequence is not unimodal, there is k, $1 \leq k \leq n-1$ such that $a_{k-1} > a_k$ and $a_k < a_{k+1}$. But then (using strict positivity) $a_{k-1}a_{k+1} > a_k^2$, contradicting log-concavity. □

The sequence of binomial coefficients is readily seen to be log-concave (and so unimodal), using the algebraic formula. A combinatorial proof of log-concavity is much harder to come by. For the Stirling numbers, log-concavity seems to hold for small values, but neither an algebraic proof nor a combinatorial one is easily found. It's worth noting, however, that log-concavity should be easier to prove combinatorially than unimodality, as it requires an *injection* for *every* $k \geq 1$ from $\mathcal{A}_{k-1} \times \mathcal{A}_{k+1}$ into $\mathcal{A}_k \times \mathcal{A}_k$ (where $\mathcal{A}_m$ is a set of objects counted by a_m) — there is no flipping from injections to surjections around the mode.

We continue to strengthen our notions. From here on we work with finite sequences, and assume if it is necessary that the sequences extend to infinity with the addition of zeros.

Definition 31.4. *A sequence* $(a_k)_{k=0}^n$ *of positive terms (with perhaps* $a_0 = 0$*) is said to be* ultra-log-concave *if for each* $k = 1, \dots, n-1$*,*

$$\left(\frac{a_k}{\binom{n}{k}}\right)^2 \geq \left(\frac{a_{k-1}}{\binom{n}{k-1}}\right)\left(\frac{a_{k+1}}{\binom{n}{k+1}}\right).$$

Evidently the sequence of binomial coefficients is (just) ultra-log-concave; and since the ultra-log-concavity condition is equivalent to

$$a_k^2 \geq a_{k-1}a_{k+1}\left(1 + \frac{1}{k}\right)\left(1 + \frac{1}{n-k}\right)$$

we easily see that an ultra-log-concave sequence is log-concave.

Our final, and strongest, notion is the *real-roots* property.

Definition 31.5. *A sequence* $(a_k)_{k=0}^n$ *of positive terms (with perhaps* $a_0 = 0$*) is said to have the* real-roots *property if the polynomial equation*

$$a_0 + a_1 x + \dots + a_n x^n = 0$$

has only real solutions.

By the binomial theorem the sequence of binomial coefficients has the real-roots property. Notice also that since we are working with sequences of positive terms (except possible a_0), having only real roots is equivalent (when $a_0 \neq 0$) to having only real, negative roots, that is, to admitting a factorization of the form

$$a_0 + a_1 x + \ldots + a_n x^n = a_0 \prod_{i=1}^{n} (1 + \alpha_i x),$$

with the α_i all real and positive (but not necessarily distinct); or (when $a_0 = 0$) to having only real, non-positive roots, with 0 a root of multiplicity 1, that is, to admitting a factorization of the form

$$a_1 x + \ldots + a_n x^n = a_1 x \prod_{i=1}^{n-1} (1 + \alpha_i x),$$

with the α_i all real and positive (but not necessarily distinct).

The following theorem goes back to Newton.

Theorem 31.6. *Let $(a_k)_{k=0}^n$ be a sequence of positive terms (with perhaps $a_0 = 0$) with the real-roots property. Then $(a_k)_{k=0}^n$ is ultra-log-concave.*

Proof. We need two easy facts.

- If the polynomial $p(x)$ has only real roots, then the same is true of its derivative $p'(x)$. Here's why this is true: if $p(x) = c \prod_{i=1}^{k} (x - \alpha_i)^{m_i}$ (so $p(x)$ has distinct roots α_i, $i = 1, \ldots, k$, with multiplicities m_i), then it is easy to check via the product rule for differentiation that $\prod_{i=1}^{k} (x - \alpha_i)^{m_i - 1}$ divides $p'(x)$. But also, between any two consecutive roots of $p(x)$ there is a point where $p'(x) = 0$ (this is Rolle's theorem, or the mean value theorem); so if $\alpha_1 < \ldots < \alpha_k$, then there are $\beta_1, \ldots, \beta_{k-1}$ with

  $$\alpha_1 < \beta_1 < \alpha_2 < \beta_2 < \ldots < \alpha_{k-1} < \beta_{k-1} < \alpha_k$$

  and with $p'(\beta_i) = 0$ for each i. We have accounted for $(k-1) + \sum_{i=1}^{k} (m_i - 1) = -1 + \sum_{i=1}^{k} m_i$ roots of $p'(x)$; these are all the roots, so $p'(x)$ has all real roots. [This fact is a special case of the more general statement that for any polynomial $p(x)$ over $\mathbb{C}$, the roots of $p'(x)$ lie in the convex hull of the roots of $p(x)$.]
- If the polynomial $p(x) = a_0 + \ldots + a_n x^n$ ($a_n \neq 0$) has only real roots, then so does the polynomial $r_p(x) = a_n + \ldots + a_0 x^n$ (the "reverse" polynomial). Here's why this is true: first suppose $a_0 \neq 0$, so all the roots of $p(x)$ are non-zero. If $p(\alpha) = 0$ then $r_p(1/\alpha) = p(\alpha)/\alpha^n = 0$, so if the roots of $p(x)$ are $\alpha_1, \ldots, \alpha_k$ (with multiplicities $m_1, \ldots, m_k$), then among the roots of $r_p(x)$ are $1/\alpha_1, \ldots, 1/\alpha_k$ (with multiplicities $m_1, \ldots, m_k$), and since $p(x)$ and $r_p(x)$ have the same degree, this accounts for all the roots. If $a_0 = 0$, then let k_0 be the least integer such that $a_{k_0} \neq 0$. We have $r_p(x) = a_n + \ldots + a_{k_0} x^{n-k_0}$, which is also $r_q(x)$ for $q(x) = a_{k_0} + \ldots + a_n x^{n-k_0}$. Since the roots of $p(x)$ are real, so are the roots of $q(x)$ (the roots of $q(x)$ are exactly the non-zero roots of $p(x)$), so by the previous case the roots of $r_q(x)$ and hence $r_p(x)$ are real.

Now consider the polynomial $p(x) = a_0 + a_1 x + \ldots + a_n x^n$. We may assume $n \geq 2$, since the result is trivial for $n = 0, 1$. Note that in the case $n = 2$, that $p(x)$ has real roots implies

that $a_1^2 \geq 4a_0a_2$, which in turn implies that

$$\left(\frac{a_1}{\binom{2}{1}}\right)^2 \geq \left(\frac{a_0}{\binom{2}{0}}\right)\left(\frac{a_2}{\binom{2}{2}}\right),$$

which is exactly the ultra-log-concavity relation. We do not need this observation for the proof, but it motivates how we proceed: in the general case we try to reduce to the quadratic case and apply the quadratic formula.

Fix k, $1 \leq k \leq n-1$. Differentiating $p(x)$ $k-1$ times and applying the first fact above, we get that

$$(k-1)!a_{k-1} + k_{(k-1)}a_k x + (k+1)_{(k-1)}a_{k+1}x^2 + \ldots + n_{(k-1)}a_n x^{n-k+1}$$

has all real roots. Applying the second fact above, we get that

$$n_{(k-1)}a_n + \ldots + (k+1)_{(k-1)}a_{k+1}x^{n-k-1} + k_{(k-1)}a_k x^{n-k} + (k-1)!a_{k-1}x^{n-k+1}$$

has all real roots. Now differentiating $n-k-1$ times and applying the first fact again, we get that

$$(n-k-1)!(k+1)_{(k-1)}a_{k+1} + (n-k)_{((n-k-1))}k_{(k-1)}a_k x + (n-k+1)_{(n-k-1)}(k-1)!a_{k-1}x^2$$

or

$$\frac{(n-k-1)!(k+1)!}{2}a_{k+1} + (n-k)!k!a_k + \frac{(n-k+1)!(k-1)!}{2}a_{k-1}$$

has all real roots, and so, using the quadratic formula and rearranging terms a bit,

$$\left(\frac{a_k}{\binom{n}{k}}\right)^2 \geq \left(\frac{a_{k-1}}{\binom{n}{k-1}}\right)\left(\frac{a_{k+1}}{\binom{n}{k+1}}\right),$$

as required. □

We now have a chain of implications for a sequence $(a_k)_{k=0}^n$ of positive terms (with perhaps $a_0 = 0$):

$$\text{real-roots property} \implies \text{ultra-log-convacity} \implies \text{log-convacity} \implies \text{unimodality}.$$

None of these implications can be reversed: $(1, 2, 5)$ is a unimodal sequence that is not log-concave; $(1, 1, 1)$ is a log-concave sequence that is not ultra-log-concave; and $(2, 6, 6, 1)$ is an ultra-log-concave sequence that does not have the real-roots property.

The property of real-rootedness, being in some sense a global rather than a local property, is often easier to establish than the weaker log-concavity. This can be particularly true in the case when we are working with a recursively-defined family of polynomials. Here it might be possible to derive the real-rootedness of later polynomials from that of earlier ones. The following theorem, first proved by Harper, illustrates this idea and brings us back to the Stirling numbers of the second kind.

Theorem 31.7. *For each $n \geq 0$, the polynomial*

$$A_n(y) = \sum_{k \geq 0} \left\{ n \atop k \right\} y^k$$

has all real roots. As a corollary, the sequence $(\left\{ n \atop k \right\})_{k \geq 0}$ is unimodal for all $n \geq 0$.

Proof. The result is trivial for $n = 0$ and easy to verify for $n = 1$. For $n \geq 2$, we will prove a stronger statement, by induction on n. Let $P(n)$ be the following proposition: $A_n(y)$, a polynomial of degree n, has constant term 0 and otherwise has positive coefficients, has n distinct roots, $0 = y_1 > y_2 > \ldots > y_n$, all real; $A_{n-1}(y)$, a polynomial of degree $n-1$, has constant term 0 and otherwise has positive coefficients, has $n-1$ distinct roots, $0 = y'_1 > y'_2 > \ldots > y'_{n-1}$, all real; and the roots of $A_{n-1}(y)$ interleave those of $A_n(y)$, that is, $0 = y_1 = y'_1 > y_2 > y'_2 > y_3 > y'_3 > \ldots > y_{n-1} > y'_{n-1} > y_n$.

We prove $P(n)$ by induction on $n \geq 2$, with the base case $n = 2$ easy to verify. For $n \geq 3$, the statement that $A_{n-1}(y)$ is a polynomial of degree $n-1$ with $n-1$ distinct roots, $0 = y'_1 > y'_2 > \ldots > y'_{n-1}$, all real, and has constant term 0 and otherwise has positive coefficients, is part of the induction hypothesis. Now we use (32), which immediately tells us that $A_n(y)$ is a polynomial of degree n with a root at 0 (so 0 constant term), and otherwise has positive coefficients. Next, we use (32) to assess the sign of $A_n(y)$ at y'_2, the least negative root of $A_{n-1}(y)$. We know that $A_{n-1}(y)$ is negative between y'_2 and $y'_1 = 0$ and has no repeated roots, so its derivative is negative at y'_2; hence $y'_2(dA_{n-1}(y)/dy)|_{y=y'_2}$ is positive, and since $y'_2 A_{n-1}(y'_2)$ is zero, (32) tells us that $A_n(y'_2)$ is positive. The same argument shows that $A_n(y'_3)$ is negative, $A_n(y'_4)$ is positive, and so on.

This tells us that $A_n(y)$ has roots $0 = y_1 > y_2 > \ldots > y_{n-1}$, all real, that satisfy

$$0 = y_1 = y'_1 > y_2 > y'_2 > y_3 > y'_3 > \ldots > y_{n-1} > y'_{n-1}.$$

But now by continuity, $A_n(y)$ has the same sign as $A_{n-1}(y)$ in some open interval $(y'_{n-1} - \varepsilon, y'_{n-1})$, but the signs of $A_n(y)$ and $A_{n-1}(y)$ are different in the limit as y goes to $-\infty$; hence, it follows that $A_n(y)$ has a root y_n satisfying $y'_{n-1} > y_n$. This accounts for all n roots of $A_n(y)$, and completes the induction. □

This idea of establishing the real-rootedness of a sequence of polynomials by demonstrating inductively that successive polynomials have interleaving roots is a quite versatile one; we may see more examples later.

We end our digression by introducing the idea of *asymptotic normality*. A sequence of sequences $((a_{n,k})_{k\geq 0})_{n\geq 0}$ is asymptotically normal if the histogram of $(a_{n,k})_{k\geq 0}$ (the set of points $\{(k, a_{n,k}) : k \geq 0\}$), suitably normalized, approaches the density function of the standard normal as n grows. We give a formal definition only in the special case where $a_{n,k} = 0$ for $k > n$, all $a_{n,k} \geq 0$, and for each n there is at least one k with $a_{n,k} > 0$. The definition involves the random variable X_n, supported on $\{0, \ldots, n\}$, with mass function $P(X_n = k) \propto a_{n,k}$. If $a_{n,k}$ counts the number of objects in the kth of n bins, then X_n can be thought of as the bin number of an object selected uniformly at random from all objects in all bins. We write

$$E(X_n) = \frac{\sum_{k=0}^{n} k a_{n,k}}{\sum_{k'=0}^{n} a_{n,k'}}$$

and

$$\mathrm{Var}(X_n) = \frac{\sum_{k=0}^{n} k^2 a_{n,k}}{\sum_{k'=0}^{n} a_{n,k'}} - \left(\frac{\sum_{k=0}^{n} k a_{n,k}}{\sum_{k'=0}^{n} a_{n,k'}} \right)^2$$

for the expectation and variance of X_n.

Definition 31.8. *The sequence* $(a_{n,k})_{n,k\geq 0}$ *is* asymptotically normal *if, for each* $x \in \mathbb{R}$*, we have*

$$\lim_{n\to\infty} \Pr\left(\frac{X_n - E(X_n)}{\sqrt{\mathrm{Var}(X_n)}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt,$$

with the convergence uniform in x*.*

For example, the DeMoivre-Laplace theorem is the statement that the sequence of binomial coefficients in asymptotically normal.

Here is the connection between asymptotic normality and real-rootedness. If $p_n(x) = \sum_{i=0}^{m} a_i x^i$ has all real roots, then there are constants p_i, $i = 1, \ldots, n$, all in $(0, 1]$, such that

$$\frac{p_n(x)}{p_n(1)} = \prod_{i=1}^{n} ((1 - p_i) + p_i x)$$

(if the roots of $p_n(x) = 0$ are $\alpha_1, \ldots, \alpha_n$, then each p_i is $1/(1-\alpha_i)$). The left-hand side above is the probability generating function of X_n (the function $\sum_{k=1}^{n} P(X_n = k)x^k$), while the right-hand side is the product of the probability generating functions of Bernoulli random variables with parameters p_i (random variables taking the value 1 with probability p_i and otherwise taking the value 0). It follows from standard results in probability that X_n can be written as $X_1 + \ldots + X_n$, where the X_i's are independent, and X_i is a Bernoulli with parameter p_i. This representation of X_n as the sum of independent random variables raises the possibility of a central limit theorem (i.e., asymptotic normality). We omit the proof of the following precise result.

Theorem 31.9. *If* $\mathrm{Var}(X_n) \to \infty$ *as* $n \to \infty$ *then* $(a_{n,k})_{n,k\geq 0}$ *is asymptotically normal. Moreover, we have the* local limit theorem

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left| \sqrt{\mathrm{Var}(X_n)} \Pr\left(X_n = \left\lfloor E(X_n) + x\sqrt{\mathrm{Var}(X_n)} \right\rfloor\right) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| = 0.$$

The advantage of a local limit theorem is that it provides quantitative information about the $a_{n,k}$'s that asymptotical normality does not; we won't discuss this any further here.

It is an easy exercise to show that if X_n is a binomial random variable with parameter $1/2$ ($P(X_n = k) \propto \binom{n}{k}$ for $k = 0, \ldots, n$), then $\mathrm{Var}(X_n) \to \infty$ as $n \to \infty$, and so Theorem 31.9 generalizes the symmetric DeMoivre-Laplace theorem. It is much less easy to show that if X_n measures the number of parts in a partition of $[n]$ selected uniformly at random, then $\mathrm{Var}(X_n) \to \infty$ as $n \to \infty$, but as first shown by Harper this is indeed the case, and so we have as a corollary:

Corollary 31.10. *The sequence* $(\left\{{n \atop k}\right\})_{n,k\geq 0}$ *of Stirling numbers of the second kind is asymptotically normal.*

32. Some problems

(1) Let $m_{n,k}$ denote the number of ways of extracting k disjoint pairs from a set of size n (here the order of the pairs does not matter, nor does order within pairs matter; so, for example, $m_{5,2} = 15$, the ways being $\{ab, cd\}$, $\{ab, ce\}$, $\{ab, de\}$, $\{ac, bd\}$, $\{ac, be\}$, $\{ac, de\}$, $\{ad, bc\}$, $\{ad, be\}$, $\{ad, ce\}$, $\{ae, bc\}$, $\{ae, bd\}$, $\{ae, cd\}$, $\{bc, de\}$, $\{bd, ce\}$, $\{be, cd\}$). Find a recurrence relation for $m_{n,k}$, with initial conditions. Set $P_n(y) = \sum_{k\geq 0} m_{n,k} y^k$. Use the recurrence relation for $m_{n,k}$ to find a recurrence for

$P_n(y)$, with initial conditions. Reality check: $P_6(y) = 1+15y+45y^2+15y^3$. Following the proof of Theorem 31.7 show that $P_n(y)$ always has all real roots (and, hint, that moreover the roots of $P_n(y)$ and $P_{n-1}(y)$ are each distinct, and distinct from each other interleave, with the least negative root of $P_n(y)$ being closer to 0 than the least negative root of $P_{n-1}(y)$).

Solution: We have $m_{0,0} = 1$ (and $m_{0,k} = 0$ for all $k > 0$). We also have $m_{1,0} = 1$ and $m_{1,k} = 0$ for all $k > 0$. Next, we have $m_{n,0} = 1$ for all $n \geq 0$. We'll take these as initial conditions. For $n \geq 2$ and $k \geq 1$, consider a fixed element, a say, of the set of size n. There are $m_{n-1,k}$ ways of extracting k pairs from the set, with a not involved in any of the pairs. For each of the $n-1$ other elements of the set, there are $m_{n-2,k-1}$ ways of extracting k pairs from the set, with a as part of a pair together with the designated other element. So we get the recurrence:

$$m_{n,k} = m_{n-1,k} + (n-1)m_{n-2,k-1}$$

valid for $n \geq 2, k \geq 1$.

For the generating polynomial, we have $P_0(y) = P_1(y) = 1$ as initial conditions. For $n \geq 2$,

$$\begin{aligned} P_n(y) &= m_{n,0} + \sum_{k\geq 1} m_{n,k} y^k \\ &= 1 + \sum_{k\geq 1} (m_{n-1,k} + (n-1)m_{n-2,k-1}) y^k \\ &= 1 + \sum_{k\geq 1} m_{n-1,k} y^k + (n-1) \sum_{k\geq 1} m_{n-2,k-1} y^k \\ &= m_{n-1,0} + \sum_{k\geq 1} m_{n-1,k} y^k + (n-1)y \sum_{k\geq 1} m_{n-2,k-1} y^{k-1} \\ &= P_{n-1}(y) + (n-1)yP_{n-2}(y). \end{aligned}$$

So $P_2(y) = 1 + y$ and $P_3(y) = 1 + 3y$, etc..

We now prove, by induction on $n \geq 3$, the proposition $\mathcal{P}(n)$: $P_n(y)$ and $P_{n-1}(y)$ both have all their roots real, negative and distinct; they both take the value 1 at $y = 0$ and are increasing at $y = 0$; and their roots interlace, Specifically, if the roots of $P_{n-1}(y)$ are $0 > r_1 > r_2 > \ldots$ (by this we mean that $r_1, r_2, r_3, \ldots$ are the roots, not that $0, r_1, r_2, \ldots$ are the roots) and the roots of $P_n(y)$ are $0 > s_1 > s_2 > \ldots$, then

$$0 > s_1 > r_1 > s_2 > r_2 > \ldots.$$

The case $n = 3$ is evident. For $n > 3$, the parts about the behavior at $y = 0$ are evident (no induction hypothesis needed). For the rest, we have

$$P_n(y) = P_{n-1}(y) + (n-1)yP_{n-2}(y)$$

and, if the roots of $P_{n-2}(y)$ are $0 > r_1 > r_2 > \ldots$ and the roots of $P_{n-1}(y)$ are $0 > s_1 > s_2 > \ldots$, then

$$0 > s_1 > r_1 > s_2 > r_2 > \ldots$$

(this by induction). Evaluating $P_n(y)$ at s_1: $P_{n-1}(s_1) = 0$, $(n-1)s_1 < 0$, $P_{n-2}(s_1) > 0$ (the last because the first negative root of $P_{n-2}(y)$, r_1, hasn't been reached yet), so $P_n(s_1) < 0$. Evaluating at s_2: $P_{n-1}(s_2) = 0$, $(n-1)s_2 < 0$, $P_{n-2}(s_2) < 0$ (the last

because s_2 lies between r_2 and r_1), so $P_n(s_2) > 0$. Continuing, we find $P_n(s_3) < 0$, $P_n(s_4) > 0$, etc..

By the intermediate value theorem, $P_n(y)$ has at least one root in each of the intervals: $(s_1, 0), (s_2, s_1), (s_3, s_2), \ldots$. If n is odd, so $P_n(y)$ and $P_{n-1}(y)$ have the same degree, this accounts for all of the roots of $P_n(y)$ and so we are done. If n is even, so $P_n(y)$ has degree one greater than $P_{n-1}(y)$, then this accounts for all but one of the roots of $P_n(y)$; the last root is to the left of the last root of $P_{n-1}(y)$, since if $\lim_{y\to-\infty} P_{n-1}(y) = +\infty$ then $\lim_{y\to-\infty} P_n(y) = -\infty$, and if $\lim_{y\to-\infty} P_{n-1}(y) = -\infty$ then $\lim_{y\to-\infty} P_n(y) = +\infty$.

Either way the induction step is complete.

33. Back to the generating function of Stirling numbers

We used generating functions to derive the identity

$$\left\{ n \atop k \right\} = \sum_{r=1}^{k} (-1)^{k-r} \frac{r^n}{r!(k-r)!}. \tag{33}$$

This was valid for $n \geq 0, k \geq 1$, but can be made to be valid for $n, k \geq 0$ simply by extending the summation to include $r = 0$, and adapting the convention $0^0 = 1$. Summing over all $k \geq 0$, we get an expression for the nth Bell number:

$$B(n) = \sum_{k=0}^{n} \sum_{r=0}^{k} (-1)^{k-r} \frac{r^n}{r!(k-r)!} = \sum_{k=0}^{n} \sum_{r=0}^{k} \frac{1}{k!} (-1)^{k-r} r^n \binom{k}{r}.$$

But notice that we can extend the inner summation to infinity, since for all $r > k$ the $\binom{k}{r}$ terms ensures we add 0; and we can extend the outer summation to infinity, too, since (33) is valid for $k > n$ (in which range it must give 0, although this is not obvious). Making both these extensions, and reversing the order of summation, we get

$$\begin{aligned} B(n) &= \sum_{r=0}^{\infty} \frac{r^n}{r!} \sum_{k=0}^{\infty} (-1)^{k-r} \frac{(k)_r}{k!} \\ &= \sum_{r=0}^{\infty} \frac{r^n}{r!} \sum_{k=r}^{\infty} \frac{(-1)^{k-r}}{(k-r)!} \\ &= \frac{1}{e} \sum_{r=0}^{\infty} \frac{r^n}{r!}. \end{aligned}$$

This is *Dobinski's formula*, a remarkable formula that has something which is evidently an integer on the left, but an infinite sum of transcendentals on the right! As an aside, note that this formula says that if X is a Poisson random variable with parameter 1, then $E(X^n) = B(n)$, a fact we alluded to earlier.

Now that we have an explicit formula for the Bell numbers, we could try to use this to form the generating function $B_0(x) = \sum_{n\geq 0} B(n)x^n$; the fact that Dobinski's formula expresses $B(n)$ as a summation raises the possibility that we can learn something new from

the generating function by re-expressing it with the order of summations reversed. We have

$$\begin{aligned} B_0(x) &= \frac{1}{e}\sum_{n\geq 0}\sum_{r\geq 0}\frac{(xr)^n}{r!} \\ &= \frac{1}{e}\sum_{r\geq 0}\frac{1}{r!}\sum_{n\geq 0}(xr)^n \\ &= \frac{1}{e}\sum_{r\geq 0}\frac{1}{r!(1-xr)}, \end{aligned}$$

at which point we run out of steam somewhat.

We now arrive at our first instance of what turns out to be a very important paradigm. When we are dealing with a sequence $(a_n)_{n\geq 0}$ in which a_n counts the number of *labelled* structures on a set of size n, it is often helpful to consider not the generating function $A(x) = \sum_{n\geq 0} a_n x^n$, as we have been considering, but rather to consider the function $\hat{A}(x) = \sum_{n\geq 0} \frac{a_n x^n}{n!}$. We will see the utility of this idea in examples, but we will also glimpse some sound theoretical reasons for the paradigm.

Definition 33.1. *Given a sequence $(a_n)_{n\geq 0}$, its* ordinary generating function *is the power series*

$$A(x) = \sum_{n\geq 0} a_n x^n,$$

and we write $(a_n)_{n\geq 0} \overset{ogf}{\longleftrightarrow} A(x)$ to indicate that $A(x)$ is the ordinary generating function (ogf) of $(a_n)_{n\geq 0}$. The exponential generating function *of the sequence is*

$$A(x) = \sum_{n\geq 0} \frac{a_n x^n}{n!}$$

and we write $(a_n)_{n\geq 0} \overset{egf}{\longleftrightarrow} A(x)$ to indicate that $A(x)$ is the exponential generating function (egf) of $(a_n)_{n\geq 0}$.

Note that if $(a_n)_{n\geq 0} \overset{ogf}{\longleftrightarrow} A(x)$ then $a_n = [x^n]A(x)$, whereas if $(a_n)_{n\geq 0} \overset{egf}{\longleftrightarrow} A(x)$ then $a_n/n! = [x^n]A(x)$; we can get over this slight awkwardness by thinking of the exponential power series as being written in terms of $x^n/n!$, and writing

$$a_n = \left[\frac{x^n}{n!}\right] A(x).$$

Let us see what happens when we use Dobinski's formula in the exponential generating function of the Bell numbers. We have

$$\begin{aligned} B(x) &= \sum_{n\geq 0} \frac{B(n)x^n}{n!} \\ &= \frac{1}{e}\sum_{n\geq 0}\sum_{r\geq 0}\frac{(xr)^n}{r!n!} \\ &= \frac{1}{e}\sum_{r\geq 0}\frac{1}{r!}\sum_{n\geq 0}\frac{(xr)^n}{n!} \\ &= \frac{1}{e}\sum_{r\geq 0}\frac{e^{xr}}{r!} \\ &= \frac{1}{e}\left(e^{e^x}\right) \\ &= e^{e^x-1}. \end{aligned}$$

So the nth Bell number is the coefficient of $x^n/n!$ in the power series of e^{e^x-1}.

Taking logarithms and differentiating with respect to x, we get

$$B'(x) = B(x)e^x.$$

The coefficient of x^n on the left-hand side is $B(n+1)/n!$. The coefficient of x^n on the right-hand side is the convolution sum

$$\sum_{k=0}^{n}\frac{B(k)}{k!(n-k)!}.$$

Equating the two, and multiplying both sides by $n!$, we get back a familiar recurrence:

$$B(n+1) = \sum_{k=0}^{n}\binom{n}{k}B(k).$$

We have gone from a recurrence (for the Stirling numbers of the second kind) to a generating function, to explicit summation formulae for the Stirling numbers and the Bell numbers, to a generating function for the Bell numbers, back to a recurrence for the Bell numbers!

34. Operations on exponential generating functions

Earlier we saw that some natural operations on power series correspond to operations of the sequences for which those power series are the (ordinary) generating functions. The same goes for exponential generating functions.

Theorem 34.1. *Let $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ be complex sequences with $(a_n)_{n\geq 0} \overset{egf}{\longleftrightarrow} A(x)$ and $(b_n)_{n\geq 0} \overset{egf}{\longleftrightarrow} B(x)$. We have the following relations:*

(1) $(a_0+b_0, a_1+b_1, \ldots) \overset{egf}{\longleftrightarrow} A(x)+B(x)$, *and more generally, for $c, d \in \mathbb{C}$,*

$$(ca_0+db_0, ca_1+db_1, \ldots) \overset{egf}{\longleftrightarrow} cA(x)+dB(x).$$

(2) $(a_1, a_2, \ldots) \overset{egf}{\longleftrightarrow} (d/dx)A(x)$, *and more generally, for* $k \geq 1$,

$$(a_k, a_{k+1}, \ldots) \overset{egf}{\longleftrightarrow} (d^k/dx^k)A(x).$$

(3) $(0, a_0, a_1, \ldots) \overset{egf}{\longleftrightarrow} \int_0^x A(t)\ dt$.

(4) *Set* $c_n = \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}$ *for* $n \geq 0$. *Then* $(c_n)_{n\geq 0} \overset{egf}{\longleftrightarrow} A(x)B(x)$. *More generally, if* $(a_n^i)_{n\geq 0} \overset{egf}{\longleftrightarrow} A_i(x)$ *for* $i = 1, \ldots, \ell$, *then*

$$\left(\sum_{x_1+\ldots+x_\ell=n,\ x_i\geq 0} \binom{n}{x_1, \ldots, x_\ell} a_{x_1}^1 a_{x_2}^2 \ldots a_{x_\ell}^\ell \right)_{n\geq 0} \overset{egf}{\longleftrightarrow} \prod_{i=1}^{\ell} A_i(x).$$

Only the fourth of these requires justification. We have

$$\begin{aligned} \sum_{n\geq 0} \frac{c_n x^n}{n!} &= \sum_{n\geq 0} \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k} \frac{x^n}{n!} \\ &= \sum_{n\geq 0} \sum_{k=0}^{n} \frac{a_k x^k}{k!} \frac{b_{n-k} x^{n-k}}{(n-k)!} \\ &= \left(\sum_{n\geq 0} \frac{a_n x^n}{n!} \right) \left(\sum_{n\geq 0} \frac{b_n x^n}{n!} \right). \end{aligned}$$

The more general statement in (4) is proved similarly.

As an example of some of these operations, consider the Fibonacci numbers $f_0 = 0, f_1 = 1$ and $f_n = f_{n-1} + f_{n-2}$ for $n \geq 2$. Let $(f_n)_{n\geq 0} \overset{egf}{\longleftrightarrow} F(x)$. We have $(a_1, a_2, \ldots) \overset{egf}{\longleftrightarrow} F'(x)$ and $(a_2, a_3, \ldots) \overset{egf}{\longleftrightarrow} F''(x)$, but also

$$(a_2, a_3, \ldots) = (a_0 + a_1, a_1 + a_2, \ldots) \overset{egf}{\longleftrightarrow} F'(x) + F(x),$$

so that $F''(x) = F'(x) + F(x)$. It follows that

$$F(x) = A_1 e^{\varphi_1 x} + A_2 e^{\varphi_2 x}$$

where φ_1, φ_2 are the solutions to $x^2 - x - 1 = 0$, that is, $\varphi_1 = (1+\sqrt{5})/2$ and $\varphi_2 = (1-\sqrt{5})/2$, so that

$$f_n = A_1 \varphi_1^n + A_2 \varphi_2^n.$$

Using $f_0 = 0, f_1 = 1$ we find that $A_1 + A_2 = 0$ and $A_1\varphi_1 + A_2\varphi_2 = 1$, so $A_1 = 1/\sqrt{5}$, $A_2 = -1/\sqrt{5}$, and we recover Binet's formula.

More generally, if the sequence $(a_n)_{n\geq 0}$ be defined by

$$a_n = c_1 a_{n-1} + \ldots + c_k a_{n-k} \ (n \geq k)$$

(where the c_i's are constants, $c_k \neq 0$), with initial values $a_0, \ldots, a_{k-1}$ given, then $(a_n)_{n\geq 0} \overset{egf}{\longleftrightarrow} A(x)$ where $A(x)$ is a solution to the differential equation

$$A^{(k)}(x) = c_1 A^{(k-1)}(x) + c_2 A^{(k-2)}(x) + \ldots + c_k A^{(0)}(x).$$

Which solution can be determined using the initial conditions.

As an example of the product formula, consider D_n, the number of derangements of $[n]$. Since each permutation of $[n]$ has k fixed points, for some $k = 0, \ldots, n$, and the remaining $n - k$ elements form a derangement of a set of size $n - k$, we have the recurrence

$$n! = \sum_{k=0}^{n} \binom{n}{k} D_{n-k},$$

valid for all $n \geq 0$ (notice that the initial condition $D_0 = 1$ is an instance of this relation). The exponential generating function of $(n!)_{n\geq 0}$ is $1/(1-x)$. The exponential generating function of the sequence on the right-hand side above is the product of the exponential generating functions of the sequences $(1, 1, 1, \ldots)$ (which is e^x) and $(D_n)_{n\geq 0}$ (which we shall call $\mathcal{D}(x)$). Thus, from the above recurrence, we obtain

$$\frac{1}{1-x} = e^x \cdot \mathcal{D}(x),$$

so that

$$\mathcal{D}(x) = \frac{e^{-x}}{1-x},$$

and so

$$\frac{D_n}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!},$$

a formula we previously derived using inclusion-exclusion.

The product rule for generating functions suggests when we might use the ordinary generating functions, versus when we might use exponential generating functions. Suppose we have three combinatorially defined families (A_n), (B_n), (C_n) indexed by $n \in \mathbb{N}$ (we think of A_n as consisting of objects of "size" n, in some appropriate sense, etc.). Consider the following process: for each k, $k = 0, \ldots, n$, take one object from B_k and one from C_{n-k}, and put these together. If, with appropriate interpretation of "put together", the resulting set can be identified with A_n, then $|A_n| = \sum_{k=0}^{n} |B_k||C_{n-k}|$, and $A(x) = B(x)C(x)$ where $A(x), B(x), C(x)$ are the ordinary generating functions of (A_n), (B_n), (C_n).

For example, suppose that A_n is the set of all subsets of size n of a set of size $\ell_1 + \ell_2$, B_n is the set of all subsets of size n of a set of size ℓ_1, and C_n is the set of all subsets of size n of a set of size ℓ_2. It's evident that we can construct an element of A_n by taking the disjoint union of an element of B_k with an element of C_{n-k}, for some choice of $k = 0, \ldots, n$, and that this process captures all of A_n, so

$$|A_n| = \sum_{k=0}^{n} |B_k||C_{n-k}|$$

and $A(x) = B(x)C(x)$. This corresponds to the algebraic identity

$$(1+x)^{\ell_1+\ell_2} = (1+x)^{\ell_1}(1+x)^{\ell_2}.$$

Now suppose the objects counted by A_n, etc., are "labelled" (each object of size n has n nodes labelled 1 through n; two objects are different if they have different labelling). Consider the following process: for each k, $k = 0, \ldots, n$, select k of the n labels; associate to those labels an object from B_k; associate to the remaining labels an object from C_{n-k}. Put together these two objects in an appropriate way. If the set of resulting objects can be put in bijective correspondence with A_n, then $|A_n| = \sum_{k=0}^{n} \binom{n}{k} |B_k||C_{n-k}|$ and so $A(x) = B(x)C(x)$ where $A(x), B(x), C(x)$ are the exponential generating functions of (A_n), (B_n), (C_n).

For example, suppose that A_n is the set of all words of length n from an alphabet of size $\ell_1 + \ell_2$, B_n is the set of all words of length n from an alphabet of size ℓ_1, and C_n is the set of all words of length n from an alphabet of size ℓ_2. It's evident that we can construct an element of A_n by selecting some k, $k = 0, \ldots, n$, selecting a subset of size k of the n positions into which the letters of a word of length n fall, filling those k positions using an element of B_k, and filling the remaining $n-k$ positions using an element of C_{n-k} (making sure that the two alphabets are disjoint), and that this process captures all of A_n, so

$$|A_n| = \sum_{k=0}^{n} \binom{n}{k} |B_k||C_{n-k}|$$

and $A(x) = B(x)C(x)$, where now $A(x)$, $B(x)$, $C(x)$ are exponential generating functions. Since

$$\sum_{n\geq 0} \frac{A_n x^n}{n!} = \sum_{n\geq 0} \frac{(\ell_1+\ell_2)^n x^n}{n!} = e^{(\ell_1+\ell_2)x}$$

in this case, this corresponds to the algebraic identity

$$e^{(\ell_1+\ell_2)x} = e^{\ell_1 x} e^{\ell_2 x}.$$

35. The exponential formula

A situation in which exponential generating functions come in useful is the following, which we shall call the *component process*. Suppose that for each $n \geq 1$, there are c_n different connected objects that can be placed on a set of n labeled nodes, with, by convention, $c_0 = 0$. Let $a_{n,k}$ be the number of ways of taking a set of nodes labelled 1 through n, partitioning it into k non-empty blocks (with no regard to the order of the blocks), and putting a connected object on each block (so, if the k blocks are $A_1, \ldots, A_k$, then the number of ways of placing connected objects is $c_{|A_1|} \cdot c_{|A_2|} \ldots c_{|A_k|}$). The $a_{n,k}$'s make perfect sense combinatorially for all $n, k \geq 0$. Note that $a_{0,0} = 1$ (since the empty set can only be partitioned into zero non-empty blocks, and then we have no choices); also, $a_{n,0} = a_{0,k} = 0$ for $n, k \geq 1$ (since a non-empty set cannot be partitioned into zero blocks, nor an empty set into a positive number of non-empty blocks).

We give a few examples of situations of this kind.

- If $c_i = 1$ for all $i \geq 1$, then there is only one connected structure that can be placed down on a labelled set of nodes, and so $a_{n,k}$ is the number of ways of partitioning a set of size n into k non-empty blocks; that is, $a_{n,k}$ is the Stirling number $\left\{ {n \atop k} \right\}$.
- Suppose that c_i is the number of labelled trees on i nodes (so, by Cayley's formula, $c_i = i^{i-2}$). Then $a_{n,k}$ counts the number of labelled forests on n nodes that have k components. (A *forest* is defined as an acyclic graph — i.e., a graph that has no cycles. Equivalently, a forest is a graph whose all connected components are trees.)
- Suppose that $c_i = i!$ for each $i \geq 1$. We can then interpret $a_{n,k}$ as the number of ways of partitioning $[n]$ into k non-empty *lists*: subsets endowed with a total order. These numbers are usually referred to as the *Lah numbers*, written $L(n,k)$.
- Suppose that $c_i = (i-1)!$ for each $i \geq 1$. Since there are $(i-1)!$ ways of arranging i objects in cyclic order, $a_{n,k}$ is the number of ways of partitioning a set of size n into k non-empty cyclically ordered blocks. These numbers are usually referred to as the *Stirling numbers of the first kind*, written $\left[{n \atop k} \right]$ — we will have much more to say about these numbers later.

- Suppose that c_i is the number of connected graphs on i labelled vertices. Then $a_{n,k}$ is the number of k-component graphs on n labelled vertices.

To try and understand the $a_{n,k}$'s, we can form the two-variable generating function

$$A(x,y) = \sum_{n,k\geq 0} \frac{a_{n,k}}{n!} x^n y^k.$$

Notice that this is a mixture of an ordinary and exponential generating function; it is ordinary in y because there is no labelling of the set of blocks, while it is exponential in x because the underlying set from which the blocks are created is labelled.

There should be a connection between $A(x,y)$ and the generating function of the c_i's. The connection is given by the *exponential formula*.

Theorem 35.1. *With the notation as above,*

$$A(x,y) = e^{yC(x)}$$

where $(c_i)_{i\geq 0} \overset{egf}{\longleftrightarrow} C(x)$.

Proof. Extracting coefficients of y^k in $A(x,y)$ and $e^{yC(x)}$, it suffices to show that for each $k \geq 0$,

$$k! \sum_{n\geq 0} a_{n,k} \frac{x^n}{n!} = C^k(x),$$

for which it suffices to show that for each $n \geq 0$ the coefficient of $x^n/n!$ in $C^k(x)$ is $k!a_{n,k}$. For $k = 0$ this is trivial, and for $k > 0$ and $n = 0$ it's also trivial, so from now on we take $n, k \geq 1$.

From the product rule for exponential generating functions, we know that the coefficient of $x^n/n!$ in $C^k(x)$ is

$$\sum_{x_1+\ldots+x_k=n,\ x_i\geq 1} \binom{n}{x_1,\ldots,x_k} c_{x_1} c_{x_2} \ldots c_{x_k}$$

(we can take all x_i's to be at least 1 since $c_0 = 0$). This is the number of ways of decomposing $[n]$ into a list of non-empty sets $A_1, A_2, \ldots, A_k$, and putting one of $c_{|A_i|}$ structures on A_i. Each object counted by $a_{n,k}$ appears $k!$ times in this sum (once for each permutation of the blocks $A_1, A_2, \ldots, A_k$), so the sum is indeed $k!a_{n,k}$. □

We give some examples of the use of the exponential formula. A quick example is to the Stirling numbers of the second kind. Here $c_i = 1$ for all $i \geq 1$, so $C(x) = e^x - 1$, and so

$$\sum_{n,k\geq 0} \left\{ {n \atop k} \right\} \frac{x^n}{n!} y^k = e^{y(e^x-1)}.$$

Setting $y = 1$ we immediately recover our closed-form expression for the exponential generating function of the Bell numbers. But we can do more by utilizing the parameter y. Recall that we arrived at the exponential generating function of the Bell numbers via the Dobinski

formula. Reversing the steps, we get a refinement of Dobinski:

$$\begin{aligned}\sum_{n,k\geq 0}\left\{{n \atop k}\right\}\frac{x^n}{n!}y^k &= e^{y(e^x-1)} = e^{-y}e^{ye^x}\\ &= e^{-y}\sum_{r\geq 0}\frac{y^r e^{rx}}{r!}\\ &= e^{-y}\sum_{r\geq 0}\frac{y^r}{r!}\sum_{n\geq 0}\frac{r^n x^n}{n!}\\ &= e^{-y}\sum_{n\geq 0}\frac{x^n}{n!}\sum_{r\geq 0}\frac{y^r r^n}{r!},\end{aligned}$$

so, extracting coefficients of $x^n/n!$ from both sides,

$$\sum_{k\geq 0}\left\{{n \atop k}\right\}y^k = \sum_{r\geq 0} r^n\left(\frac{y^r}{r!}e^{-y}\right).$$

Dobinski's formula is the case $y = 1$ of this. For general y, note that if X_y is a Poisson random variable with parameter y (so $\Pr(X_y = r) = (y^r/r!)e^{-y}$ for $r = 0, 1, \ldots$), then the right-hand side above is the expected value of X_y^n. So we learn that for each $n \geq 1$, the polynomial

$$\sum_{k\geq 0}\left\{{n \atop k}\right\}y^k = \sum_{k=0}^{n}\left\{{n \atop k}\right\}y^k$$

(we know that's a polynomial, since in this pass through the Stirling numbers we have begun not with a recurrence but with a combinatorial definition) is exactly $E(X_y^n)$ where $X_y \sim \text{Poisson}(y)$.

We can go further:

$$\begin{aligned}\sum_{k\geq 0}\left\{{n \atop k}\right\}y^k &= \sum_{r\geq 0} r^n\left(\frac{y^r}{r!}e^{-y}\right)\\ &= \sum_{r\geq 0}\frac{r^n y^r}{r!}\sum_{\ell\geq 0}(-1)^\ell\frac{y^\ell}{\ell!}\\ &= \sum_{\ell,r\geq 0}(-1)^\ell\frac{r^n y^{\ell+r}}{\ell! r!}.\end{aligned}$$

Extracting the coefficient of y^k from both sides, we recover a variant of an old, familiar summation formula:

$$\left\{{n \atop k}\right\} = \sum_{\ell+r=k}(-1)^\ell\frac{r^n}{\ell! r!}.$$

36. Stirling numbers of the first kind

We begin with some basics of permutations.

Definition 36.1. *A* permutation *is a bijection f from $[n]$ to $[n]$. The* one-line *representation of a permutation f is the juxtaposition*

$$f(1)f(2)\ldots f(n).$$

A cycle representation *of a permutation f is a bracketed list*

$$(a_{11}a_{12}\dots a_{1c_1})(a_{21}a_{22}\dots a_{2c_2})\dots(a_{k1}a_{k2}\dots a_{kc_k})$$

where

- *each list $(a_{i1}, a_{i2}, \dots, a_{ic_i})$ forms a* cycle *in f (that is, the numbers $a_{i1}, a_{i2}, \dots, a_{ic_i}$ are distinct and satisfy $f(a_{i1}) = a_{i2}$, $f(a_{i2}) = a_{i3}$, …, $f(a_{ic_i}) = a_{i1}$), and*
- *the sets $\{a_{i1}, a_{i2}, \dots, a_{ic_i}\}$, $i = 1, \dots, k$ form a decomposition of $[n]$.*

Here c_i is referred to as the length *of the ith cycle.*

A canonical cycle representation *of a permutation f is one in which*

- *we have $a_{11} < a_{21} < \dots < a_{k1}$, and*
- *for each i, the number a_{i1} is the least element in the set $\{a_{i1}, a_{i2}, \dots, a_{ic_i}\}$.*

The cycle type *of a cycle representation is the vector $(a_1, a_2, \dots)$, where the multiset $\{c_1, \dots, c_k\}$ has a_j occurrences of j for each $j \geq 1$. (That is, the permutation has a_1 cycles of length 1, a_2 of length 2, etc.).*

The following theorem is straightforward and the proof is left as an exercise.

Theorem 36.2. *Every permutation has a cycle representation. All cycle representations of the same permutation have the same cycle type, and any one can be obtained from any other by cyclically permuting the elements within pairs of consecutive brackets, and permuting the bracketed lists. Each permutation has a unique canonical cycle representation.*

In particular, we can talk about the cycle type and number of cycles of a permutation. The main point in the proof is this: by the pigeon-hole principle, for each $i \in [n]$ the sequence $(i, f(i), f^2(i), \dots)$ must be eventually periodic; and because f is a bijection, if k is the least positive integer such that $f^k(i)$ appears in the set $\{i, f(i), \dots, f^{k-1}(i)\}$, then $f^k(i) = i$ (so in fact the sequence is periodic, not just eventually periodic).

Definition 36.3. *For $n, k \geq 0$, the* Stirling number of the first kind $\begin{bmatrix} n \\ k \end{bmatrix}$ *is the number of permutations of $[n]$ that have exactly k cycles.*

Note that $\begin{bmatrix} n \\ 0 \end{bmatrix} = 0$ for any $n > 0$ (since no permutation of $[n]$ has 0 cycles), and $\begin{bmatrix} 0 \\ k \end{bmatrix} = 0$ for $k > 0$ (since no permutation of $[0]$ has more than 0 cycles), but $\begin{bmatrix} 0 \\ 0 \end{bmatrix} = 1$ (since the identity permutation of $[0]$ has 0 cycles: its canonical cycle representation is an empty list).

For any $n \geq 1$, there are $(n-1)!$ permutations of $[n]$ with exactly one cycle (this may be seen by the double counting argument earlier in these notes, or simply by observing that there are $n-1$ choices for the image of 1 in such a permutation, then $n-2$ choices for the image of the image of 1, etc.). It follows that the numbers $\begin{bmatrix} n \\ k \end{bmatrix}$ are generated from the sequence $(0, 0!, 1!, \dots, (n-1)!, \dots)$ by the component process. Since the exponential generating function of this sequence is

$$\sum_{n \geq 1} \frac{(n-1)!}{n!} x^n = \sum_{n \geq 1} \frac{x^n}{n} = -\log(1-x),$$

it follows from the exponential formula that

$$\begin{aligned}\sum_{n,k\geq 0}\begin{bmatrix}n\\k\end{bmatrix}\frac{x^n}{n!}y^k &= e^{-y\log(1-x)}\\ &= (1-x)^{-y}\\ &= \sum_{n\geq 0}\frac{y^{(n)}}{n!}x^n.\end{aligned}$$

Extracting the coefficient of $x^n/n!$ from both sides, we get the nice relation

$$y(y+1)\dots(y+n-1)=\sum_{k\geq 0}\begin{bmatrix}n\\k\end{bmatrix}y^k. \tag{34}$$

This, in turn, leads to a recurrence relation for $\begin{bmatrix}n\\k\end{bmatrix}$. For $n\geq 1$, we have

$$\begin{aligned}\sum_{k\geq 0}\begin{bmatrix}n\\k\end{bmatrix}y^k &= y(y+1)\dots(y+(n-1)-1)(y+n-1)\\ &= (y+(n-1))\sum_{k\geq 0}\begin{bmatrix}n-1\\k\end{bmatrix}y^k\\ &= \sum_{k\geq 0}\begin{bmatrix}n-1\\k\end{bmatrix}y^{k+1}+\sum_{k\geq 0}(n-1)\begin{bmatrix}n-1\\k\end{bmatrix}y^k\\ &= \sum_{k\geq 1}\begin{bmatrix}n-1\\k-1\end{bmatrix}y^k+\sum_{k\geq 0}(n-1)\begin{bmatrix}n-1\\k\end{bmatrix}y^k\\ &= (n-1)\begin{bmatrix}n-1\\0\end{bmatrix}y^0+\sum_{k\geq 1}\left(\begin{bmatrix}n-1\\k-1\end{bmatrix}+(n-1)\begin{bmatrix}n-1\\k\end{bmatrix}\right)y^k.\end{aligned}$$

So, equating coefficient of y^k, we obtain the equality

$$\begin{bmatrix}n\\k\end{bmatrix}=\begin{bmatrix}n-1\\k-1\end{bmatrix}+(n-1)\begin{bmatrix}n-1\\k\end{bmatrix} \tag{35}$$

for all $n\geq 1$ and $k\geq 1$. So the Stirling numbers of the first kind are completely specified by the recurrence (35) for $n,k\geq 1$, together with initial conditions $\begin{bmatrix}n\\0\end{bmatrix}=\begin{bmatrix}0\\k\end{bmatrix}=0$ for $n,k>0$, $\begin{bmatrix}0\\0\end{bmatrix}=1$.

The recurrence (35) can also be seen combinatorially — in a permutation of $[n]$ with k cycles, either element n is in a cycle on its own (there are $\begin{bmatrix}n-1\\k-1\end{bmatrix}$ such permutations), or it is in a cycle with some other elements (in which case removing "n" from the cycle representation leaves one of $\begin{bmatrix}n-1\\k\end{bmatrix}$ permutations of $[n-1]$, and element n can be inserted into such a permutation in any of $n-1$ ways).

The relation (34) can also be interpreted combinatorially, in the case when y is a positive integer. A *y-labelled* permutation of $[n]$ is a permutation of $[n]$ together with an assignment of a label to each cycle in the permutation, with labels coming from $\{1,\dots,y\}$ (and with distinct cycles not required to receive distinct labels). Let $\mathcal{S}$ be the set of y-labelled permutations of

$[n]$. Evidently

$$|\mathcal{S}| = \sum_{k\geq 0} \begin{bmatrix} n \\ k \end{bmatrix} y^k.$$

But we may enumerate $\mathcal{S}$ in another way. First, decide what label the cycle involving element 1 receives; there are y options. Element 2 then either is part of a new cycle (to which a label has to be given, y options), or joins element 1 in its cycle (one more option); so there are $y+1$ options for the fate of element 2. In general element k either is part of a new cycle (to which a label has to be given, y options), or joins an existing cycle ($k-1$ more options; element k can be inserted immediately after any of the previous $k-1$ elements in the cycle representation); so there are $y+k-1$ options for the fate of element k. It follows that

$$|\mathcal{S}| = y(y+1)\dots(y+n-1),$$

and (34) follows by the polynomial principle.

At first blush this argument may not seem to be too similar to our proof of (11) in Claim 14.3, but there is a recasting of the argument that bears a close resemblance to the more "colorful" proof of (11).

In how many ways can n people sit at circular tables at a bar, each table with a pitcher of beer (different tables are allowed to choose the same beer), if the bar has y brands of beer available? One way this can be achieved is as follows: the first person enters the bar, chooses a beer and starts a table, y options. The second person enters, and either chooses a beer and starts a new table, or sits immediately to the left of the first person, $y+1$ choices. In general, the kth person enters, and either chooses a beer and starts a new table, or sits immediately to the left of someone already seated, $y+(k-1)$ choices. Thought of this way, the counting problem has $y(y+1)\dots(y+(n-1))$ solutions. Another strategy is for the n people together to decide how many tables they will form (k), then form k tables ($\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right]$), then one after the other each table chooses a beer (y^k). Thought of this way, the counting problem has $\sum_{k\geq 1} \left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right] y^k$ solutions. The polynomial principle completes the proof of the identity.

The similar-seeming relations (34) and (11) become even more similar if we replace y with $-x$ throughout (34), in which case it becomes

$$(x)_n = \sum_{k\geq 0} (-1)^{n-k} \begin{bmatrix} n \\ k \end{bmatrix} x^k \tag{36}$$

for $n \geq 0$, which seems very much like an "inverse" to our earlier

$$x^n = \sum_{k\geq 0} \begin{Bmatrix} n \\ k \end{Bmatrix} (x)_k. \tag{37}$$

Theorem 36.4. *Let $S_2 = (\left\{\begin{smallmatrix} n \\ k \end{smallmatrix}\right\})_{n,k\geq 0}$ and $S_1 = ((-1)^{n-k}\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right])_{n,k\geq 0}$ be the doubly-infinite matrices of Stirling numbers of the second kind and (signed) Stirling numbers of the first kind, respectively. These matrices are inverses to one another, i.e.,*

$$S_1 S_2 = S_2 S_1 = I$$

where I is the identity matrix $(i_{n,k})_{n,k\geq 0}$ with $i_{n,k} = 1$ if $n = k$ and 0 otherwise.

Proof. This is standard linear algebra: from (37) we see that S_2 is the matrix that represent the change of basis in the space of one-variable polynomials from $1, x, x^2, x^3, \dots$ to $1, x, (x)_2, (x)_3, \dots$, while from (36) we see that S_1 represents the change of basis in the other direction.

If we do not want to draw on the theory of infinite-dimensional vector space, we can simply insert one of (36), (37) into the other. For example, for $n \geq 0$ we have

$$\begin{aligned} x^n &= \sum_{k \geq 0} \left\{ n \atop k \right\} (x)_k \\ &= \sum_{k \geq 0} \left\{ n \atop k \right\} \sum_{j \geq 0} (-1)^{k-j} \left[k \atop j \right] x^j \\ &= \sum_{j \geq 0} \left[\sum_{k \geq 0} \left\{ n \atop k \right\} (-1)^{k-j} \left[k \atop j \right] \right] x^j. \end{aligned}$$

Comparing coefficients of different powers of x we find that

$$\sum_{k \geq 0} \left\{ n \atop k \right\} (-1)^{k-j} \left[k \atop j \right] = \begin{cases} 1 & \text{if } j = n \\ 0 & \text{otherwise.} \end{cases}$$

This is exactly the statement that $S_2 S_1 = I$; substituting in the other direction (and using that $\{1, x, (x)_2, (x)_3, \ldots\}$ forms a basis of the space of one-variable polynomials) gives $S_1 S_2 = I$. □

37. Bell-type numbers for the component process

For any counting problem that arises from the component process, there is an analog of the Bell numbers: we can ask how many objects can be constructed on label set $[n]$, without regard to the number of components. By setting $y = 1$ in the exponential formula we can answer this question.

Theorem 37.1. *Let c_n be the number of connected objects that can be placed on a label set of size n, and let $a_{n,k}$ the number of ways of partitioning $[n]$ into k non-empty blocks and putting a connected object on each block. Let $a_n := \sum_{k \geq 0} a_{n,k}$. With $(a_n)_{n \geq 0} \overset{egf}{\longleftrightarrow} A(x)$ and $(c_n)_{n \geq 0} \overset{egf}{\longleftrightarrow} C(x)$ we have*

$$A(x) = e^{C(x)}$$

and so (taking logarithms and differentiating)

$$A'(x) = A(x) C'(x)$$

and (extracting the coefficient of $x^n/n!$ from both sides)

$$a_{n+1} = \sum_{k=0}^{n} \binom{n}{k} a_k c_{n-k+1}$$

for $n \geq 0$ (with $a_0 = 1$).

For example, let g_n be the number of labelled graphs on n vertices, and c_n the number of *connected* labelled graphs; since $g_n = 2^{\binom{n}{2}}$, we can easily produce the sequence $(c_n)_{n \geq 0}$ from the identity

$$2^{\binom{n+1}{2}} = \sum_{k=0}^{n} \binom{n}{k} 2^{\binom{k}{2}} c_{n-k+1}$$

valid for $n \geq 0$. Specifically, we have the recurrence

$$c_{n+1} = 2^{\binom{n+1}{2}} - \sum_{k=1}^{n} \binom{n}{k} 2^{\binom{k}{2}} c_{n-k+1}$$

for $n \geq 1$, with initial condition $c_1 = 1$. We can easily calculate from this that, for example, $c_2 = 1$, $c_3 = 4$ and $c_4 = 38$. (The full sequence is A001187 in the online encyclopedia of integer sequences.)

We now give a more substantial application of Theorem 37.1. Recall that a *forest* is a graph with no cycles and a *tree* is a forest with just one component. A tree is *rooted* if one vertex is identified as the root, and a forest is rooted if each component has a root. Let t_n be number of rooted trees on $[n]$ and f_n the number of rooted forests, and let $T(x)$ and $F(x)$ be the respective exponential generating functions. Declaring $t_0 = 0$ and $f_0 = 1$, Theorem 37.1 tells us that

$$F(x) = e^{T(x)}.$$

There is another simple relationship between $F(x)$ and $T(x)$: we have $t_{n+1} = (n+1)f_n$ for $n \geq 0$. Indeed, consider the following bijection from pairs (k, F) where $k \in [n+1]$ and F is a rooted forest on $[n]$ to rooted trees on $[n+1]$. Given a pair (k, F), then for each vertex in F with label k or greater, add 1 to its label, then add a new vertex labeled k to F, declare it the root of a tree, and join it to all the roots of F. It follows that

$$F(x) = \sum_{n \geq 0} \frac{t_{n+1}}{(n+1)!} x^n = \frac{1}{x} T(x)$$

(using $t_0 = 0$) and so

$$T(x) = x e^{T(x)}.$$

How do we get $T(x)$ from this? Using the method of taking logarithms and then derivatives we get the following nice recurrence for t_n:

$$t_{n+1} = \frac{n+1}{n} \sum_{k=1}^{n} \binom{n}{k} t_k t_{n+1-k} \tag{38}$$

valid for $n \geq 1$, with initial conditions $t_0 = 0$ and $t_1 = 1$.

Recall that we already know, from Cayley's formula, that $t_n = n^{n-1}$. It's not immediately clear how to extract this from (38); we will now see the Lagrange inversion formula, that solves this problem.

38. Lagrange inversion

If $f(x) = a_0 + a_1 x + a_2 x^2 + \ldots$ is a power series with $a_0 \neq 0$, then the reciprocal function $1/f(x)$ can also be expanded as a power series about 0; indeed, if we set

$$\frac{1}{a_0 + a_1 x + a_2 x^2 + \ldots} = b_0 + b_1 x + b_2 x^2 + \ldots$$

then we can solve for the b_i one-by-one via

$$\begin{aligned} a_0b_0 = 1 \quad &\text{so} \quad b_0 = 1/a_0 \\ a_0b_1 + a_1b_0 = 0 \quad &\text{so} \quad b_1 = (-a_1b_0)/a_0 \\ &\dots \\ a_0b_k + a_1b_{k-1} + \ldots + a_kb_0 = 0 \quad &\text{so} \quad b_k = (-a_1b_{k-1} - \ldots - a_kb_0)/a_0 \\ &\dots; \end{aligned}$$

notice that if $a_0 = 0$ then this process breaks down quickly.

If $a_0 = 0$ then we can still take the reciprocal, but now we have to move from power series to *Laurent series* — series of the form $\sum_{n\in\mathbb{Z}} a_n x^n$ with only finitely many of the a_n with $n < 0$ being non-zero. Indeed, if k is the first index with $a_k \neq 0$, we have

$$\begin{aligned} \frac{1}{a_kx^k + a_{k+1}x^{k+1} + a_{k+2}x^{k+2} + \ldots} &= \frac{1}{x^k}\left(\frac{1}{a_k + a_{k+1}x^1 + a_{k+2}x^2 + \ldots}\right) \\ &= \frac{1}{x^k}(b_0 + b_1x + b_2x^2 + \ldots) \\ &= b_0x^{-k} + b_1x^{-k+1} + \ldots \end{aligned}$$

where $b_0 + b_1x + b_2x^2 + \ldots$ is the reciprocal of $a_k + a_{k+1}x^1 + a_{k+2}x^2 + \ldots$.

If $f(x) = a_0 + a_1x + a_2x^2 + \ldots$ and $g(x) = p_1x + p_2x^2 + p_3x^3 + \ldots$ are two power series such that $g(x)$ has constant term 0, then the composition function $f(g(x))$ makes perfect sense as a power series. Indeed, the expression

$$a_0 + a_1(p_1x + p_2x^2 + p_3x^3 + \ldots) + a_2(p_1x + p_2x^2 + p_3x^3 + \ldots)^2 + \ldots$$

is a perfectly valid power series, since for $n \geq 0$ the coefficient of x^n is the finite sum

$$\sum_{k=0}^{n} a_k \sum_{i_1+\ldots+i_k=n,\ i_j>0} p_{i_1} \cdots p_{i_k}. \tag{39}$$

If $f(x) = a_0 + a_1x + a_2x^2 + \ldots$ is a power series with $a_0 = 0$ and $a_1 \neq 0$, then there is a unique power series $g(x) = p_1x + p_2x^2 + \ldots$ that is the *compositional inverse* of $f(x)$, that is, that satisfies $f(g(x)) = g(f(x)) = x$. Indeed, let us first try to find a power series $g(x) = p_1x + p_2x^2 + \ldots$ satisfying $f(g(x)) = x$. Since the coefficient of x^n in $f(g(x))$ is given by (39), this requires us to solve the system of equations

$$\sum_{k=0}^{n} a_k \sum_{i_1+\ldots+i_k=n,\ i_j>0} p_{i_1} \cdots p_{i_k} = \begin{cases} 1, & \text{if } n = 1; \\ 0, & \text{if } n \neq 1 \end{cases}$$

for all $n \geq 0$. In view of $a_0 = 0$, this system simplifies to

$$\sum_{k=1}^{n} a_k \sum_{i_1+\ldots+i_k=n,\ i_j>0} p_{i_1} \cdots p_{i_k} = \begin{cases} 1, & \text{if } n = 1; \\ 0, & \text{if } n > 1 \end{cases}$$

for all $n \geq 1$ (because $a_0 = 0$ guarantees that the equation for $n = 0$ is always satisfied). Because of $a_1 \neq 0$, we can solve these equations for p_n recursively, obtaining $p_1 = 1/a_1$ and

$$p_n = \frac{-1}{a_1} \sum_{k=2}^{n} a_k \sum_{i_1+\ldots+i_k=n,\ i_j>0} p_{i_1} \cdots p_{i_k} \qquad \text{for } n \geq 2. \tag{40}$$

Thus, there exists a power series $g(x) = p_1x + p_2x^2 + \dots$ satisfying $f(g(x)) = x$. Similarly, there exists a power series $h(x)$ satisfying $h(f(x)) = x$. We can now see that the two power series $g(x)$ and $h(x)$ must be identical. Indeed, substituting $g(x)$ for x in $h(f(x)) = x$ yields $h(f(g(x))) = g(x)$, whence $g(x) = h(f(g(x))) = h(x)$ (because $f(g(x)) = x$). This shows that the power series $g(x)$ satisfies $f(g(x)) = g(f(x)) = x$. We thus have shown that such a power series exists; its uniqueness follows from a similar argument as we just used to prove $g(x) = h(x)$.

The formula (40) is quite involved, involving sums over compositions, so is not terribly practical for identifying the coefficients of the compositional inverse of a power series. A much more useful tool exists, the Lagrange inversion formula.

Theorem 38.1. *Let $f(x) = a_1x + a_2x^2 + \dots$ be a power series with $a_1 \neq 0$, and let $g(x) = p_1x + p_2x^2 + \dots$ be its compositional inverse. Then for $n \geq 1$*

$$p_n = \frac{1}{n}[x^{n-1}]\left(\frac{x}{f(x)}\right)^n.$$

Note that $x/f(x)$ is the reciprocal of $f(x)/x$, which has constant term $a_1 \neq 0$, so $x/f(x)$ is an ordinary power series. Note also that if an unknown function $g(x)$ satisfies a functional equation

$$g(x) = x\varphi(g(x))$$

for some function $\varphi(x)$ that has an ordinary power series expansion, then, considering $f(x) = x/\varphi(x)$ we see that $g(x)$ is the compositional inverse of $f(x)$ and so

$$[x^n]g(x) = \frac{1}{n}[x^{n-1}]\varphi^n(x).$$

Before proving the formula, we pause to give an example. With t_n counting the number of rooted trees on n labelled vertices, and $T(x) = \sum_{n\geq 1} t_n x^n/n!$, we have derived the functional equation

$$T(x) = xe^{T(x)}.$$

It follows that $T(x)$ is the compositional inverse of the function $f(x) = xe^{-x}$. By Theorem 38.1 we have

$$\begin{aligned}\frac{t_n}{n!} &= \frac{1}{n}[x^{n-1}]\left(\frac{x}{xe^{-x}}\right)^n \\ &= \frac{1}{n}[x^{n-1}]e^{nx} \\ &= \frac{1}{n}\frac{n^{n-1}}{(n-1)!}\end{aligned}$$

and so $t_n = n^{n-1}$. Since t_n/n counts the number of labelled (unrooted) trees on n labelled vertices, we have found a new proof of Cayley's formula.

Proof. (Theorem 38.1) The definition of g shows that

$$x = g\left(f\left(x\right)\right) = \sum_{i\geq 1} p_i f^i(x).$$

After differentiating and dividing through by $f^n(x)$, this becomes

$$\frac{1}{f^n(x)} = \sum_{i\geq 1} ip_i f^{i-n-1}(x)f'(x). \tag{41}$$

Note that both sides above are Laurent series.

We claim that

$$[x^{-1}] \sum_{i \geq 1} i p_i f^{i-n-1}(x) f'(x) = n p_n, \tag{42}$$

which yields the Lagrange inversion formula, since then

$$p_n = \frac{1}{n}[x^{-1}] \sum_{i \geq 1} i p_i f^{i-n-1}(x) f'(x) = \frac{1}{n}[x^{-1}] \left(\frac{1}{f^n(x)} \right) = \frac{1}{n}[x^{n-1}] \left(\frac{x}{f(x)} \right)^n$$

(where we have used (41) for the second equality sign).

To see (42), note that at $i = n$ we have

$$\begin{aligned}
[x^{-1}] i p_i f^{i-n-1}(x) f'(x) &= n p_n [x^{-1}] \left(\frac{f'(x)}{f(x)} \right) \\
&= n p_n [x^0] \left(x \cdot \frac{a_1 + 2a_2 x + 3a_3 x^2 + \ldots}{a_1 x + a_2 x^2 + a_3 x^3 + \ldots} \right) \\
&= n p_n [x^0] \left(\frac{a_1 + 2a_2 x + 3a_3 x^2 + \ldots}{a_1 + a_2 x + a_3 x^2 + \ldots} \right) \\
&= n p_n .
\end{aligned}$$

On the other hand, for $i \neq n$ we have

$$i p_i f^{i-n-1}(x) f'(x) = \frac{i p_i}{i-n} \frac{d}{dx} f^{i-n}(x)$$

(since $\frac{d}{dx} f^{i-n}(x) = (i-n) f^{i-n-1}(x) f'(x)$). Now $f^{i-n}(x)$ is a Laurent series, and it is evident that the coefficient of x^{-1} in the derivative of a Laurent series is 0. □

Here is another quick application of Lagrange inversion. We know from Galois theory (specifically the Abel-Ruffini theorem) that there is no procedure that finds the solutions to the general quintic $a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 = 0$ over the complex numbers, that uses only addition, subtraction, multiplication, division and extraction of roots. However, some progress can be made on the general quintic. Bring and Jerrard discovered a procedure (albeit a quite involved one) that reduces the problem of solving the general quintic to that of solving a quintic in so-called *Bring-Jerrard form*,

$$x^5 - x + a = 0. \tag{43}$$

The first step of this procedure, namely scaling by a_0 and then substituting $x - a \cdot \frac{1}{5} a_0$ for x, to eliminate the coefficient of x^4 is easy; the remaining steps are quite intricate[17].

Here is an approach to solving (43). We know that a as a function of x is $a(x) = x - x^5$, so that viewing x as a function of x, $x = x(a)$, we have the functional relation $a = x(a) - x(a)^5$. This says that $x(a)$ is the compositional inverse of the function $a \mapsto a - a^5$, and we can use

[17]See for example the discussion at `http://math.stackexchange.com/questions/542108/how-to-transform-a-general-higher-degree-five-or-higher-equation-to-normal-form`.

Lagrange inversion to expand $x(a)$ as a power series in a: For any $n \geq 1$, we obtain

$$\begin{aligned}[a^n]x(a) &= \frac{1}{n}[a^{n-1}]\left(\frac{1}{1-a^4}\right)^n \\ &= \frac{1}{n}[a^{n-1}]\sum_{\ell\geq 0}\binom{n+\ell-1}{\ell}a^{4\ell}.\end{aligned}$$

This is non-zero only when $n = 4k+1$ for some $k \geq 0$, in which case it is $\binom{5k}{k}/(4k+1)$. It follows that

$$x(a) = \sum_{k\geq 0}\frac{\binom{5k}{k}}{4k+1}a^{4k+1},$$

a power series with radius of convergence around 1.869. (The sequence $\left(\binom{5k}{k}/(4k+1)\right)_{k\geq 0}$, which is clearly an analog of the Catalan sequence, is the 5th *Fuss-Catalan* sequence, A002294 in the online encyclopedia of integer sequences.)

At $a = 0$ we are working with the quintic $x^5 - x = 0$, which has roots at 0, ± 1 and $\pm i$. The power series we have derived is capturing how one of those roots, specifically the root 0, varies as a varies around 0 (specifically, as a varies around the ball of radius about 1.869 around 0 in the complex plane).

The requirement that the series we work with converges is a little unfortunate; but it is possible to rewrite the series in terms of hypergeometric functions, and then use a general theory to analytically continue the series to the whole complex plane, thereby allowing for an understanding of how the 0 root of $x^5 - x + a = 0$ at $a = 0$ varies as a varies over the $\mathbb{C}$.

39. Finding averages with generating functions

Suppose that $(\mathcal{A}_n)_{n\geq 0}$ is a sequence of finite sets, and that each $\mathcal{A}_n$ comes with a decomposition

$$\mathcal{A}_n = \bigcup_{k=0}^{n}\mathcal{A}_{n,k}.$$

We think of $\mathcal{A}_n$ of being the nth level of some larger combinatorial family $\mathcal{A} = \bigcup_{n\geq 0}\mathcal{A}_n$, and of $\mathcal{A}_{n,k}$ as the number of objects of "size" k in level n. Let μ_n be the average size of an object from level n; that is, suppose we pick an object from $\mathcal{A}_n$ uniformly at random, observe its size, and let μ_n be the expected value of this random variable.

Let $a_{n,k}$ denote $|\mathcal{A}_{n,k}|$, and let $P_n(x) = \sum_{k=0}^{n} a_{n,k}y^k$ be the generating function of the sequence $(a_{n,k})_{k=0}^{n}$. Then

$$\begin{aligned}\mu_n &= \frac{0a_{n,0} + 1a_{n,1} + \ldots + na_{n,n}}{a_{n,0} + a_{n,1} + \ldots + a_{n,n}} \\ &= \frac{P_n'(1)}{P_n(1)} \\ &= \left\{(\log P_n(x))'\right\}_{x=1}.\end{aligned}$$

Moveover, we have

$$\frac{P_n''(1) + P_n'(1)}{P_n(1)} = \frac{0^2a_{n,0} + 1^2a_{n,1} + \ldots + n^2a_{n,n}}{a_{n,0} + a_{n,1} + \ldots + a_{n,n}}$$

so that σ_n^2, the variance of an object chosen uniformly from level n, is

$$\begin{aligned}\sigma_n^2 &= \frac{P_n''(1)+P_n'(1)}{P_n(1)} - \left(\frac{P_n'(1)}{P_n(1)}\right)^2 \\ &= \left\{(\log P_n(x))' + (\log P_n(x))''\right\}_{x=1}\end{aligned}$$

For example, if $\mathcal{A}_n$ is the set of all subsets of $[n]$, and $\mathcal{A}_{n,k}$ is the set of all those subsets of size k, then μ_n is the average size of a set chosen randomly from a set of size n, and σ_n^2 is the variance; that is, μ_n and σ_n^2 are the mean and variance of the binomial random variable with parameters n and $1/2$. Since $P_n(y) = (1+y)^n$ in this case, we immediately get

$$\mu_n = \frac{n}{2},\ \sigma_n^2 = \frac{n}{4}.$$

For a more substantial example, consider the Stirling numbers of the first kind. Let μ_n be the average number of cycles in a permutation of $[n]$ chosen uniformly at random. Since in this case $P_n(y) = \sum_{k=0}^n \left[{n \atop k}\right] y^k$, and we know from (34) that this equals $y^{(n)}$, we have

$$\begin{aligned}\mu_n &= \frac{1}{1^{(n)}}\frac{d}{dy}y^{(n)}|_{y=1} \\ &= \frac{1}{n!}\left(\frac{y^{(n)}}{y} + \frac{y^{(n)}}{y+1} + \ldots + \frac{y^{(n)}}{y+n-1}\right)|_{y=1} \\ &= \frac{1}{n!}\left(\frac{n!}{1} + \frac{n!}{2} + \ldots + \frac{n!}{n}\right) \\ &= \frac{1}{1} + \frac{1}{2} + \ldots + \frac{1}{n}.\end{aligned}$$

This is the *nth Harmonic number* H_n, and is approximately $\log n$. More precisely, there is a constant $\gamma = .577\ldots$, the *Euler-Mascheroni constant*, such that

$$H_n - \log n \to \gamma$$

as $n \to \infty$.

We could have also obtained this result by observing that

$$\log y^{(n)} = \log y + \log(y+1) + \ldots + \log(y+n-1)$$

so that

$$\left(\log y^{(n)}\right)' = \frac{1}{y} + \frac{1}{y+1} + \ldots + \frac{1}{y+n-1}$$

which clearly evaluates to H_n at $y=1$. Via this approach the variance is also very simple:

$$\left(\log y^{(n)}\right)'' = \frac{1}{y^2} + \frac{1}{(y+1)^2} + \ldots + \frac{1}{(y+n-1)^2}$$

which evaluates to $\sum_{i=1}^n \frac{1}{i^2}$ at $y=1$, a number which is uniformly bounded in n by $\pi^2/6$, so that the variance of the number of cycles in a randomly selected permutation of $\{1,\ldots,n\}$ is asymptotically $\log n$.

Incidentally, since $\log n \to \infty$ as $n \to \infty$, and the generating polynomial of $\left(\left[{n \atop k}\right]\right)_{k\geq 0}$ evidently has all real roots, this shows via Theorem 31.9 that the sequence of Stirling numbers of the first kind is asymptotically normal.

40. Some problems

(1)

- part a)[18]. Let a_n be the number of permutations of $[n]$ with only odd-length cycles. Use the exponential formula to show that $A(x)$, the exponential generating function of $(a_n)_{n\geq 0}$, satisfies

$$A(x) = \sqrt{\frac{1+x}{1-x}}.$$

Solution: By the exponential formula (in the special case $y = 1$), $A(x)$ is $e^{C(x)}$ where $C(x)$ is the exponential generating function of the sequence $(0, 0!, 0, 2!, 0, 4!, \ldots)$, that is

$$\begin{aligned} C(x) &= \frac{0!x}{1!} + \frac{2!x^3}{3!} + \frac{4!x^5}{5!} + \ldots \\ &= x + \frac{x^3}{3} + \frac{x^5}{5} + \ldots. \end{aligned}$$

Now

$$-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \ldots$$

and so

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots,$$

and so

$$\frac{\log(1+x) - \log(1-x)}{2} = x + \frac{x^3}{3} + \frac{x^5}{5} + \ldots$$

so that

$$C(x) = \log\sqrt{\frac{1+x}{1-x}}$$

and indeed

$$A(x) = \sqrt{\frac{1+x}{1-x}}.$$

- part b). Let b_n be the number of permutations of $[n]$ with an even number of cycles, all odd-length. Use the result of part a) to find $B(x)$, the exponential generating function of $(b_n)_{n\geq 0}$.
Solution: If n is odd, then $b_n = 0$ (by parity). If n is even, then $b_n = a_n$. So

$$\begin{aligned} B(x) &= a_0 + a_2\frac{x^2}{2!} + a_4\frac{x^4}{4!} + \ldots \\ &= \frac{A(x) + A(-x)}{2} \\ &= \frac{1}{2}\left(\sqrt{\frac{1+x}{1-x}} + \sqrt{\frac{1-x}{1+x}}\right) \\ &= (1-x^2)^{-1/2}. \end{aligned}$$

[18]The first and third problems below are taken from *generatingfunctionology* by Herb Wilf.

- part c). Let p be the probability that a permutation selected uniformly at random from all permutations of $[n]$ has an even number of cycles, all odd-length. Let q be the probability that a fair coin tossed n times comes up heads exactly $n/2$ times. Use the result of part b) to show that $p = q$.
  Solution: We have

$$q = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \binom{n}{n/2} 2^{-n} & \text{if } n \text{ is even.} \end{cases}$$

  On the other hand, the coefficient of x^n in $B(x)$ is $b_n/n!$, which is exactly p. It's clear that $[x^n](1-x^2)^{-1/2} = 0$ if n is odd, so it remains to show that $[x^n](1-x^2)^{-1/2} = \binom{n}{n/2}2^{-n}$ if n is even.
  By the binomial theorem, for even n we have

$$\begin{aligned} [x^n](1-x^2)^{-1/2} &= (-1)^{n/2}\binom{-\frac{1}{2}}{n/2} = (-1)^{n/2}\frac{\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\cdots\left(-\frac{2n-1}{2}\right)}{(n/2)!} \\ &= \frac{\left(\frac{1}{2}\right)\left(\frac{3}{2}\right)\cdots\left(\frac{2n-1}{2}\right)}{(n/2)!} = \frac{1\cdot 3\cdot \ldots \cdot (n-1)}{(n/2)!2^{n/2}} \\ &= \frac{(1\cdot 3\cdot \ldots \cdot (n-1))(2\cdot 4\cdot \ldots \cdot n)}{(n/2)!2^{n/2}(2\cdot 4\cdot \ldots \cdot n)} \\ &= \frac{(1\cdot 3\cdot \ldots \cdot (n-1))(2\cdot 4\cdot \ldots \cdot n)}{(n/2)!(n/2)!2^{n}} \\ &= \frac{n!}{(n/2)!(n/2)!2^n} = \binom{n}{n/2}2^{-n}, \end{aligned}$$

  as required.

(2) The Lah number $L(n,k)$ is the number of ways to partition $[n]$ into k non-empty *lists* (ordered sets). For example, $L(4,2) = 36$ (each of the 4 partitions into a singleton and a set of size 3 gives rise to 6 partitions into lists, one for each of the 6 ways of ordering the block of size 3, and each of the 3 partitions into two blocks of size 2 gives rise to 4 partitions into lists). By convention $L(0,0) = 1$.
  - part a) Use the exponential formula to get a closed-form expression for the mixed generating function $L(x,y) = \sum_{n,k\geq 0} L(n,k)\frac{x^n}{n!}y^k$, and in turn use this to find an explicit expression for $L(n,k)$ (which should turn out to be very simple, involving no summation).
    Solution: The Lah numbers $L(n,k)$ are the result of the component process applied to the sequence $(c_n)_{n\geq 1}$ given by $c_n = n!$. The exponential generating function $C(x)$ of this sequence is

$$C(x) = \sum_{n\geq 1}\frac{n!x^n}{n!} = \frac{x}{1-x},$$

    so by the exponential formula

$$L(x,y) = e^{\frac{xy}{1-x}}.$$

Extracting the coefficient of y^k ($k \geq 0$) from both sides we find

$$\begin{aligned}\sum_{n\geq 0} L(n,k)\frac{x^n}{n!} &= \frac{1}{k!}\left(\frac{x^k}{(1-x)^k}\right) \\ &= \frac{1}{k!}x^k \sum_{n\geq 0}\binom{n+k-1}{k-1}x^n\end{aligned}$$

(for $k > 0$). Extracting the coefficient of x^n ($n \geq 0$) from both sides we find

$$\frac{L(n,k)}{n!} = \frac{1}{k!}\binom{n-1}{k-1},$$

so

$$L(n,k) = \frac{n!}{k!}\binom{n-1}{k-1}$$

(again for $k > 0$).

- part b) Give a combinatorial explanation for the formula you obtained in part a).

  Solution: We'll show combinatorially that

  $$k!L(n,k) = n!\binom{n-1}{k-1}$$

  by showing that both sides count the number of ways of decomposing $[n]$ into a *list* (ordered) of k non-empty lists. That the left-hand side counts this is immediate (since each partition of $[n]$ into k non-empty lists corresponds to $k!$ distinct decompositions of $[n]$ into a list of k non-empty lists). For the right-hand side: we can decompose $[n]$ into a list of k non-empty lists by first writing down a permutation on $[n]$ in one-line notation, then selecting a composition $a_1 + \ldots + a_k = n$ of n into $k-1$ parts, and then letting the first a_1 elements of the one-line permutation be the entries in the first list (listed in the order given by the permutation), then letting the next a_2 elements of the permutation be the entries in the second list, and so on. This process produces each list of lists exactly once, and there are $n!\binom{n-1}{k-1}$ of them ($\binom{n-1}{k-1}$ being the number of compositions of n into $k-1$ parts).

- part c) Either combinatorially, or from the generating function, show that for all $n \geq 1$

  $$x^{(n)} = \sum_{k=1}^{n} L(n,k)(x)_k$$

  (so that the Lah matrix $(L(n,k))_{n,k\geq 0}$ moves one from the falling-power basis of the space of polynomials to the rising-power basis).

  Solution: Here's a combinatorial proof:

  Consider a bar with x different beers available, and with lots of linear tables. (The tables are indistinguishable.) The right-hand side above counts the number of ways that n people can enter the bar, arrange themselves (in a linear order) at some tables, and each table order a *different* pitcher of beer, by first deciding on k, the number of tables to be occupied, then deciding on the occupation of the tables (that's the $L(n,k)$), and then each table deciding which beer to order. The left-hand side counts the same thing by the following process: The first person

enters the bar, selects a beer for his table (x options), and sits at a table. The second person comes in, and either decides to form a new table, in which case he chooses a beer for it, and sits at an unoccupied table ($x-1$ options), or he joins an existing table (2 options for where to sit at the table), for $x+1$ options in total. When the kth person comes in, he either decides to form a new table, in which case he chooses a beer for it, and sits at an unoccupied table ($x-\ell$ options, where ℓ is the number of tables currently occupied), or he joins an existing table (ℓ options if he decides to sit in the first spot of a table, and $k-1$ options if he decides to sit in a later spot — in this latter case he has to sit next to one of the $k-1$ people already seated), for $x+k-1$ options in total. This leads to a grand total of $x^{(n)}$ options.

This proves the identity for all positive integers x, and so we are done by the polynomial principle.

- part d) Deduce that

$$(x)_n = \sum_{k=1}^{n} (-1)^{n-k} L(n,k) x^{(k)}$$

(so that the matrix that moves one in the other direction is the signed Lah matrix $\left((-1)^{n-k} L(n,k)\right)_{n,k\geq 0}$).

Solution: Making the substitution of $-x$ for x, the identity in part c) immediately yields this identity.

(3) In the coupon collector's problem we imagine that we would like to get a complete collection of photos of movie stars, where each time we buy a box of cereal we acquire one such photo, which may of course duplicate one that is already in our collection. Suppose there are d different photos in a complete collection. Let p_n be the probability that exactly n trials are needed in order, for the first time, to have a complete collection.

- part a) Find a very simple formula for p_n involving Stirling numbers of the second kind.

Solution: There are d^n ways in which a sequence of n trials could turn out. The number of ways in which a particular photo appears for the first time on the nth trial, and in which it is the dth photo to appear, is $(d-1)!\left\{ {n-1 \atop d-1} \right\}$ (the $\left\{ {n-1 \atop d-1} \right\}$ partitions the first $n-1$ trials into $d-1$ non-empty blocks, each block representing the trials on which a particular photo appears; the $(d-1)!$ assigns particular photos to the blocks). Since there are d photos that can be chosen as the one to appear for the first time last, we get that there are $d(d-1)!\left\{ {n-1 \atop d-1} \right\} = d!\left\{ {n-1 \atop d-1} \right\}$ ways for exactly n trials to be needed to see the d photos, and so

$$p_n = \frac{d!\left\{ {n-1 \atop d-1} \right\}}{d^n} = \frac{(d-1)!\left\{ {n-1 \atop d-1} \right\}}{d^{n-1}}.$$

- part b) Let $p(x)$ be the ordinary generating function of $(p_n)_{n\geq d}$. Show that

$$p(x) = \frac{(d-1)!x^d}{(d-x)(d-2x)\cdots(d-(d-1)x)}.$$

(You can use anything we know about the generating function of Stirling numbers of the second kind.)

Solution: We know from (29) that for each $d \geq 1$,

$$\sum_{n \geq d-1} \left\{ {n-1 \atop d-1} \right\} x^{n-1} = \frac{x^{d-1}}{(1-x)(1-2x)\dots(1-(d-1)x)}.$$

so that

$$\begin{aligned} \sum_{n \geq d-1} \left\{ {n-1 \atop d-1} \right\} (x/d)^{n-1} &= \frac{(x/d)^{d-1}}{(1-x/d)(1-2(x/d))\dots(1-(d-1)(x/d))} \\ &= \frac{x^{d-1}}{(d-x)(d-2x)\cdots(d-(d-1)x)}. \end{aligned}$$

Multiplying both sides by $(d-1)!x$ yields the claimed identity.

- part c) Find, directly from the generating function $p(x)$, the average number of trials that are needed to get a complete collection of all d coupons.
  Solution: Letting μ be the average, we have

$$\begin{aligned} \mu &= \left\{ (\log p(x))' \right\}_{x=1} \\ &= \left\{ \frac{d}{x} + \frac{1}{d-x} + \frac{2}{d-2x} + \dots + \frac{d-1}{d-(d-1)x} \right\}_{x=1} \\ &= d + \sum_{k=1}^{d-1} \frac{k}{d-k}. \end{aligned}$$

  This seems a little unwieldy, but with a little algebra can be simplified to

$$\mu = dH_d$$

  where $H_d = \sum_{k=1}^{d} \frac{1}{k}$ is the dth harmonic number.
- part d) Similarly, using $p(x)$, find the standard deviation of the number of trials.
  Solution: Letting σ^2 be the variance, we have

$$\begin{aligned} \sigma^2 &= \left\{ (\log p(x))' + (\log p(x))'' \right\}_{x=1} \\ &= dH_d + \left\{ -\frac{d}{x^2} + \frac{1}{(d-x)^2} + \frac{4}{(d-2x)^2} + \dots + \frac{(d-1)^2}{(d-(d-1)x)^2} \right\}_{x=1} \\ &= dH_d - d + \sum_{k=1}^{d-1} \frac{k^2}{(d-k)^2}. \end{aligned}$$

  This doesn't have a particularly nice simple form.
- part e) If there are 10 different kinds of pictures, how many boxes of cereal would you expect to have to buy in order to collect all 10?
  Solution: $10H_{10} = 7381/252 \approx 29.2897$.

41. Pulling out arithmetic progressions

Suppose we have a sequence $(a_n)_{n \geq 0}$, with a known generating function $A(x)$, and for some reason we become interested in the sequence $(a_0, 0, a_2, 0, a_4, \dots)$. How do we identify its generating function without starting a new computation from scratch? Here's one way: since

$$A(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 + \dots,$$

we have

$$A(-x) = a_0 - a_1x + a_2x^2 - a_3x^3 + a_4x^4 - a_5x^5 + \ldots,$$

and so

$$\frac{A(x) + A(-x)}{2} = a_0 + 0x + a_2x^2 + 0x^3 + a_4x^4 + 0x^5 + \ldots,$$

or, more succinctly,

$$(a_0, 0, a_2, 0, a_4, 0, \ldots) \overset{ogf}{\longleftrightarrow} \frac{A(x) + A(-x)}{2}.$$

For example, what is $\sum_{k\geq 0} \binom{n}{2k}$? We have

$$\left(\binom{n}{k}\right)_{k\geq 0} \overset{ogf}{\longleftrightarrow} (1+x)^n,$$

and so

$$\left(\binom{n}{0}, 0, \binom{n}{2}, 0, \binom{n}{4}, 0, \ldots\right) \overset{ogf}{\longleftrightarrow} \frac{(1+x)^n + (1-x)^n}{2}.$$

Setting $x = 1$ we get $\sum_{k\geq 0} \binom{n}{2k} = 2^{n-1}$ whenever $n \geq 1$.

If instead we wanted to pick out all the *odd-indexed* terms of a sequence, we would use

$$(0, a_1, 0, a_3, 0, a_5, \ldots) \overset{ogf}{\longleftrightarrow} \frac{A(x) - A(-x)}{2}.$$

For example,

$$\left(0, \binom{n}{1}, 0, \binom{n}{3}, 0, \ldots\right) \overset{ogf}{\longleftrightarrow} \frac{(1+x)^n - (1-x)^n}{2},$$

and setting $x = 1$ we get $\sum_{k\geq 0} \binom{n}{2k+1} = 2^{n-1}$ whenever $n \geq 1$.

What if we wanted to pick out every *third* term of a sequence, or every mth? This can be easily accomplished using roots of unity. Recall that in the complex plane, there are m distinct numbers z with the property that $z^m = 1$; these are given by $e^{2k\pi i/m}$ with $k = 0, \ldots, n-1$. Let us set, for $m \geq 1$ and $0 \leq k \leq m-1$,

$$\omega_k^{(m)} = e^{\frac{2k\pi i}{m}}.$$

Notice that $\omega_0^{(m)} = 1$; we will often use 1 in place of $\omega_0^{(m)}$ for this special root. Notice also that every $m \geq 2$ satisfies

$$\sum_{k=0}^{m-1} \omega_k^{(m)} = 0$$

because the left-hand side is the coefficient of z^{m-1} in the polynomial $z^m - 1$. Furthermore, with the addition in $k_1 + k_2$ being performed modulo m,

$$\omega_{k_1}^{(m)} \omega_{k_2}^{(m)} = \omega_{k_1+k_2}^{(m)}.$$

Given a sequence $(a_n)_{n\geq 0}$ with ordinary generating function $A(x)$ we have, for $m \geq 1$ and each $k \in \{0, \ldots, m-1\}$,

$$A(\omega_k^{(m)} x) = \sum_{n\geq 0} \omega_{nk}^{(m)} a_n x^n,$$

so

$$\begin{aligned}\sum_{k=0}^{m-1} A(\omega_k^{(m)}x) &= \sum_{k=0}^{m-1}\sum_{n\geq 0}\omega_{nk}^{(m)}a_n x^n \\ &= \sum_{n\geq 0}\left[\sum_{k=0}^{m-1}\omega_{nk}^{(m)}\right]a_n x^n.\end{aligned}$$

If n is a multiple of m, then $\sum_{k=0}^{m-1}\omega_{nk}^{(m)} = m$. If n is not a multiple of m, then

$$\begin{aligned}\sum_{k=0}^{m-1}\omega_{nk}^{(m)} &= \sum_{k=0}^{m-1} e^{\frac{2kn\pi i}{m}} \\ &= \frac{1-e^{2n\pi i}}{1-e^{\frac{2n\pi i}{m}}} \\ &= 0,\end{aligned}$$

using the sum of a finite geometric series in the second equality. So

$$\frac{1}{m}\sum_{k=0}^{m-1} A(\omega_k^{(m)}x) = \sum_{n\geq 0} a_{mn}x^{mn}$$

is the generating function of the sequence that agrees with $(a_n)_{n\geq 0}$ on all multiples of m, but is 0 everywhere else.

For example, we know that

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \ldots = e^x,$$

and that

$$1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots = \frac{e^x + e^{-x}}{2} = \cosh x.$$

But what about

$$1 + \frac{x^3}{3!} + \frac{x^6}{6!} + \ldots?$$

The three cube roots of 1 are 1, $\omega_1 = (-1+\sqrt{3}i)/2$ and $\omega_2 = (-1-\sqrt{3}i)/2$, and so

$$\begin{aligned}1 + \frac{x^3}{3!} + \frac{x^6}{6!} + \ldots &= \frac{e^x + e^{\omega_1 x} + e^{\omega_2 x}}{3} \\ &= \frac{1}{3}\left(e^x + 2e^{-\frac{x}{2}}\cos\left(\frac{\sqrt{3}x}{2}\right)\right),\end{aligned}$$

the second equality using DeMoivre's formula; and we could play a similar game with $\sum_{n\geq 0}\frac{x^{nm}}{(nm)!}$ for any $m \geq 1$.

Going back to sums of binomial coefficients, we now know that

$$\sum_{k\geq 0}\binom{n}{3k} = \frac{1}{3}\left(2^n + \left(\frac{1+\sqrt{3}i}{2}\right)^n + \left(\frac{1-\sqrt{3}i}{2}\right)^n\right),$$

and so, although it is not the case that $\sum_{k\geq 0}\binom{n}{3k} = 2^n/3$ (by divisibility conditions, this could not possibly be true), it is the case that

$$\lim_{n\to\infty}\frac{\sum_{k\geq 0}\binom{n}{3k}}{2^n} = \frac{1}{3},$$

the point being that both $1+\omega_1$ and $1+\omega_2$ have absolute values less than 2. More generally, since $|1+\omega_k^{(m)}| < 2$ for all $m \geq 1$ and $k \neq 0$, we get

$$\lim_{n\to\infty}\frac{\sum_{k\geq 0}\binom{n}{mk}}{2^n} = \frac{1}{m}.$$

It is also possible to use this "roots of unity" method to find the generating functions of sequences which agree with a given sequence at every mth term, not starting at the 0th but rather the first, or the second, etc., and are zero everywhere else. The details are a little messy. We'll just state the formula for the case $m = 3$, where they are easily verified by hand:

$$\sum_{n\geq 0} a_{3n+1}x^{3n+1} = \frac{A(x) + \omega_2 A(\omega_1 x) + \omega_1 A(\omega_2 x)}{3}$$

and

$$\sum_{n\geq 0} a_{3n+2}x^{3n+2} = \frac{A(x) + \omega_1 A(\omega_1 x) + \omega_2 A(\omega_2 x)}{3}.$$

42. Midterm exam with solutions

(1)
- What does it mean for a sequence $(a_n)_{n\geq 0}$ of positive terms to be *unimodal*?

  Solution: There exists $k \geq 0$ such that $a_0 \leq a_1 \leq \ldots \leq a_k \geq a_{k+1} \geq \ldots$.
- Verify *directly* that the sequence $\left(\binom{n}{k}\right)_{k=0}^{n}$ is log-concave.

  Solution: For each $1 \leq k \leq n-1$ we have

  $$\begin{array}{lll}\binom{n}{k}^2 \geq \binom{n}{k-1}\binom{n}{k+1} & \iff & \frac{n!n!}{k!(n-k)!k!(n-k)!} \geq \frac{n!n!}{(k-1)!(n-k+1)!(k+1)!(n-k-1)!} \\ & \iff & 1 \geq \frac{k(n-k)}{(k+1)(n-k+1)}\end{array}$$

  which is certainly true since $k \leq k+1$ and $n-k \leq n-k+1$, and all factors on the right are positive.

(2) A *hopscotch board with n squares* is a vector $(a_1, a_2, \ldots, a_\ell)$ (ℓ not fixed) with each $a_i \in \{1, 2\}$ and with $\sum_{i=1}^{\ell} a_i = n$. Let h_n be the number of hopscotch boards with n squares.
- Write down a linear depth two recurrence relation that the h_n's satisfy, and give initial conditions.

  Solution: $h_1 = 1$ (the one board being (1)), $h_2 = 2$ (the two boards being (2) and (1, 1)), and for $n \geq 2$

  $$h_n = h_{n-1} + h_{n-2}$$

  since the set of n-square boards decomposes into those that begin with a 1 and those that begin with a 2; there are h_{n-1} of the former, since these are in 1-1 correspondence with $(n-1)$-square boards, via deletion of the leading 1, and there are h_{n-2} of the latter, since these are in 1-1 correspondence with $(n-2)$-square boards, via deletion of the leading 2.

- Prove *combinatorially* that $h_n = \sum_{k\geq 0} \binom{n-k}{k}$.

  Solution: The set of n-square boards decomposes according to k, the number of 2's in the vector, with k ranging over non-negative integers. A board with k 2's has $n-2k$ 1's, so has $\ell = k + (n-2k) = n-k$. The board is fully specified by selecting the k indices among the $\ell = n-k$ indices in total which correspond to 2's; there are $\binom{n-k}{k}$ options for doing this.

(3)
- Give the definition of a *weak composition* of a number n into k parts.

  Solution: It is a vector $(a_1, \ldots, a_k)$ with all $a_i \geq 0$ and an integer, and with $\sum_{i=1}^{k} a_i = n$.

- Let $n \geq 0$ and $k \geq 3$. How many solutions are there to the equation $a_1 + a_2 + \ldots + a_k = n$, if each of a_1, a_2, a_3 are integers at least 0, and the rest of the a_i's are integers at least 1?

  Solution: Set $a_i' = a_i + 1$ for $i = 1, 2, 3$, and set $a_i' = a_i$ for $i > 3$. Solutions to $a_1 + a_2 + \ldots + a_k = n$ with each of a_1, a_2, a_3 integers at least 0, and the rest of the a_i's integers at least 1, are in 1-1 correspondence with solutions to $a_1' + a_2' + \ldots + a_k' = n + 3$ with each a_i' an integer at least 1, that is, with *compositions* of $n+3$ into k parts. There are

$$\binom{n+3-1}{k-1} = \binom{n+2}{k-1}$$

  such.

(4)
- Let $A_1, \ldots, A_n$ be n subsets of a set U. Write down the inclusion-exclusion formula for calculating $|U \setminus (A_1 \cup A_2 \cup \ldots \cup A_n)|$.

  Solution: $|U \setminus (A_1 \cup A_2 \cup \ldots \cup A_n)| = \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|} |\bigcap_{i \in I} A_i|$, where the empty intersection is taken to be U.

- A math department has n professors and $2n$ courses, repeated every semester, so each semester each professor teaches two of the courses. In how many ways can the courses be assigned to the professors in the spring semester, if no professor is to exactly repeat the pair of courses she taught in the fall (repeating one is fine)? [Your answer will most likely involve a summation.]

  Solution: Let U be the set of all possible assignments of the $2n$ courses to the n professors in the spring, and for each $i = 1, \ldots, n$ let A_i be the set of assignments in which the ith professor (in some predetermined ordering) teaches the same two courses she taught in the fall. We seek $|U \setminus (A_1 \cup A_2 \cup \ldots \cup A_n)|$.
  We have $|U| = \binom{2n}{2,2,\ldots,2}$ with (n 2's) and $|A_i| = \binom{2n-2}{2,2,\ldots,2}$ (with $n-1$ 2's), and more generally

$$|\bigcap_{i \in I} A_i| = \binom{2n-2|I|}{2,2,\ldots,2} \text{ (with } n - |I| \text{ 2's)}$$

for every $I \subseteq [n]$. Thus, by inclusion-exclusion,

$$\begin{aligned}
&|U \setminus (A_1 \cup A_2 \cup \ldots \cup A_n)| \\
&= \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|} \binom{2n-2|I|}{2,2,\ldots,2} \text{ (with } n-|I| \text{ 2's)} \\
&= \sum_{k=0}^{n} (-1)^k \binom{n}{k} \binom{2n-2k}{2,2,\ldots,2} \text{ (with } n-k \text{ 2's).}
\end{aligned}$$

(5) A *rooted* Stirling partition of the second kind of $[n]$ into k parts is an (unordered) partition of $[n]$ into k blocks, all non-empty, with each block having a distinguished element. For example, with the distinguished element identified in bold face, here are three different rooted Stirling partitions of [7] into 3 parts: **1**|25**6**7|3**4**, **1**|25**6**7|**3**4 and **1**|25**6**7|3**4**. Write $\left\{ {n \atop k} \right\}_{\mathrm{r}}$ for the number of rooted Stirling partitions of the second kind of $[n]$ into k parts.

- Justify the relation

$$\sum_{n,k \geq 0} \left\{ {n \atop k} \right\}_{\mathrm{r}} \frac{x^n}{n!} y^k = e^{xye^x}.$$

Solution: Rooted Stirling partitions of the second kind are produced from the component process, where the number of connected objects on a labelled set of size n is n (the root has to be chosen). The exponential generating function of the sequence $(c_1, c_2, \ldots)$ where $c_n = n$ is

$$\sum_{n \geq 1} n \frac{x^n}{n!} = x \sum_{n \geq 1} \frac{x^{n-1}}{(n-1)!} = xe^x.$$

From the exponential formula the claimed relation follows immediately.

- Justify the formula $\left\{ {n \atop k} \right\}_{\mathrm{r}} = \binom{n}{k} k^{n-k}$.

Solution: To specify a rooted Stirling partition of the second kind of $[n]$ into k parts, we may first specify the roots of the k blocks; there are $\binom{n}{k}$ options for this. We may then specify, for each of the remaining $n-k$ elements, which root that element is associated with (i.e., what is the root of the block in which it belongs); there are k^{n-k} options for this. This completes the specification of the partition, and evidently all partitions are obtained once and only once by this process.

43. Set systems

We have been focussing so far on enumerative questions. We now shift gears and turn our attention to more structural and extremal problems in combinatorics. For the rest of the semester, the basic object of study will be (some variant of) the *set system*.

Definition 43.1. *A* set system*, or* family of sets *on* ground set X*, is a set* $\mathcal{F} = \{A_i : i \in I\}$ *where* I *is some index set, and each* A_i *is a subset (possibly empty) of* X*.*

Almost always our ground set will be finite, in which case we typically take it to be $[n]$ $(= \{1, \ldots, n\})$ for some natural number n. In this case the set system $\mathcal{F}$ will also be finite,

and we will often write it as $\mathcal{F} = \{A_1, \ldots, A_m\}$ (so, to recap: n will typically denote the size of the ground set, and m the number of distinct subsets in our system).

Before diving into some problems, we establish some notation related to a few very basic set systems.

Notation 43.2. *The set system consisting of all 2^n subsets of $[n]$ is denoted $\mathcal{B}_n$, and referred to as the n-dimensional* Boolean cube *or* Boolean lattice.

- *We write $\binom{[n]}{k}$ for the set of subsets of $[n]$ of size exactly k; this is referred to as the k*th level *of the Boolean cube.*
- *We also write $\binom{[n]}{\leq k}$ for the set of subsets of $[n]$ of size at most k, and likewise, we write $\binom{[n]}{\geq k}$ for the set of subsets of $[n]$ of size at least k.*
- *For any $0 \leq k \leq \ell \leq n$, we write $\binom{[n]}{[k,\ell]}$ for the set of subsets of $[n]$ of size between k and ℓ inclusive.*

For a fixed element x of the ground set X, $\mathcal{F}^{(x)}$ denotes the star *on x: the set of subsets of X that include element x. More generally for $A \subseteq X$, $\mathcal{F}^{(A)}$ denotes the* star *on A: the set of subsets of X that include A as a subset.*

Some comments about the Boolean cube: first, there is a natural bijective correspondence between $\mathcal{B}_n$ and the set of all words of length n over alphabet $\{0, 1\}$ (equivalently the set of 0-1 vectors of length n), given by mapping $A \in \mathcal{B}_n$ to the vector χ_A whose ith co-ordinate is 1 if $i \in A$ and 0 otherwise; we refer to χ_A as the *indicator vector* of A. Second, note that $\mathcal{B}_n$ decomposes as $\bigcup_{k=0}^{n} \binom{[n]}{k}$. We will often represent the Boolean cube visually as a diamond ($\diamond$) with the bottom vertex representing the empty set ($\varnothing \in \binom{[n]}{0}$), the top vertex representing the set $[n]$.

Rather than try to broadly describe the kinds of questions that we will consider about set systems, we list a few representative examples:

- How large can a set system be, if no two sets in the family are allowed to be disjoint? What about if it is required that any two sets in the family intersect in at least t elements, for some fixed $t \geq 1$? What if our set systems are restricted to live inside $\binom{[n]}{k}$ for some k?
- A set system on $[n]$ has the property that the size of the intersection of any two distinct elements is independent of the particular choice of pair of elements (i.e., there is some absolute λ such that $|A \cap B| = \lambda$ for all $A \neq B$ in the set system). How large can the set system be?
- What is the greatest number of elements that one can select from the Boolean cube, with the property that no one element is contained in any other?
- A *distinguishing family* is a set system $\mathcal{D}$ on ground set $[n]$ with the property that for each $A, B \in \mathcal{B}_n$ with $A \neq B$, there is $D \in \mathcal{D}$ with $|A \cap D| \neq |B \cap D|$. What is the size of the smallest distinguishing family?

44. Intersecting set systems

A set system $\mathcal{F}$ is *intersecting* if $A \cap B \neq \emptyset$ for each $A, B \in \mathcal{F}$.

Question 44.1. *What is $\max |\mathcal{F}|$, with the maximum taken over all intersecting families on ground set $[n]$?*

One way to make sure that a set of sets pairwise intersects is to determine in advance their (common) intersection; this leads to an obvious candidate for an extremal intersecting family, namely $\mathcal{F}^{(x)}$, the set of sets containing a particular element x. This has size 2^{n-1}. Another way to ensure pairwise intersection is to choose only large sets for the family. This leads to an obvious candidate for an extremal intersecting family, at least in the case where n is odd:

$$\mathcal{F}_{\text{large}} := \binom{[n]}{\geq n/2}.$$

Since n is odd, any $A \in \mathcal{F}_{\text{large}}$ has $|A| > n/2$, so by the pigeon-hole principle $\mathcal{F}_{\text{large}}$ is intersecting. How big is it? Well, by the binomial symmetry $\binom{n}{k} = \binom{n}{n-k}$ we have

$$|\mathcal{F}_{\text{large}}| = \sum_{k>n/2} \binom{n}{k} = \frac{1}{2} \sum_{k=0}^{n} \binom{n}{k} = 2^{n-1}.$$

For even n, we get an intersecting family by taking all subsets of size at least $n/2+1$, but this can be augmented somewhat by adding any collection of sets of size exactly $n/2$ that themselves form an intersecting family; for example, we can add to $\binom{n}{\geq n/2+1}$ all subsets of n of size $n/2$ that include element 1. Let us set, for n even,

$$\mathcal{F}_{\text{large}} := \binom{[n]}{\geq n/2+1} \cup \{A \in \binom{[n]}{n/2} : 1 \in A\}.$$

This is an intersecting family, and again by a symmetry argument we have $|\mathcal{F}_{\text{large}}| = 2^{n-1}$.

Theorem 44.2. *Let $\mathcal{F}$ be an intersecting family on ground set $[n]$, where $n \geq 0$. We have $|\mathcal{F}| \leq 2^{n-1}$.*

Proof. The result is trivial for $n = 0$. For $n > 0$, partition the set $\mathcal{B}_n$ into 2^{n-1} pairs of the form $(A, [n] \setminus A)$. (Such a partition exists: the map $f : \mathcal{B}_n \to \mathcal{B}_n$ given by $f(A) = [n] \setminus A$ is a bijection with the property that f^2 is the identity, and (since $n > 0$) has no fixed points, so the permutation of $\mathcal{B}_n$ corresponding to f has 2^{n-1} cycles each of length 2. These cycles give the required partition.) An intersecting family can only include at most one element from each pair in this partition, and so must have size at most 2^{n-1}. □

Many results in the theory of set systems come with characterizations of cases of equality in inequalities. Typically there are only a very small number (up to isomorphism) of equality cases, but not for the bound we have just proven. The following theorem, which we will not prove, is due to Erdős and Hindman.

Theorem 44.3. *Let $n = 2m$ be even. There are at least $2^{\binom{2m-1}{m-1}}$ intersecting families on ground set $[n]$ of size 2^{n-1}. In the other direction, there is a constant $c > 0$ such that the number of intersecting families on ground set $[n]$ of size 2^{n-1} is at most $2^{(1+c(\log m)/m)\binom{2m-1}{m-1}}$.*

There is an analogous statement for odd n.

We now explore what happens if we demand that pairs of elements intersect more substantially. Say that a set system $\mathcal{F}$ is *t-intersecting* if $|A \cap B| \geq t$ for each $A, B \in \mathcal{F}$.

Question 44.4. *Fix n and t with $n \geq t \geq 1$. What is $\max |\mathcal{F}|$, with the maximum taken over all t-intersecting families on ground set $[n]$?*

In the case $t = 1$, we are asking simply about intersecting families. Both natural extremal examples for intersecting families extend to t-intersecting. For the first, if we fix a set A of size t, then $\mathcal{F}^{(A)}$, the set of sets containing A as a subset, is a t-intersecting family of size 2^{n-t}.

For the second, again we separate into two cases. If $n + t$ is even then set

$$\mathcal{F}_{\text{large}} := \binom{[n]}{\geq (n+t)/2}.$$

For $A, B \in \mathcal{F}_{\text{large}}$ we have $n \geq |A \cup B| = |A| + |B| - |A \cap B| \geq n + t - |A \cap B|$ so $|A \cap B| \geq t$.

For $n + t$ odd, we again want to include all sets in all levels that are high enough that large intersection is forced by simple cardinality/pigeon-hole considerations. To this end, we begin with $\binom{[n]}{\geq (n+t+1)/2}$ — any two sets in this family have intersection size at least t. But this family can be augmented somewhat by adding any collection of sets of size exactly $(n + t - 1)/2$ that itself forms a t-intersecting family. One way to form such a family is to take all subsets of level $(n + t - 1)/2$ that contain a fixed set of size t; this family has size $\binom{n-t}{(n-t-1)/2}$. Another way to form such a family is to simply take all subsets of size $(n+t-1)/2$ of $\{2, \ldots, n\}$ (by pigeon-hole principle this is t-intersecting); this family has size $\binom{n-1}{(n+t-1)/2}$.

When $t = 1$ both of these families have the same size, but for $t > 1$ (and $n + t$ odd) it turns out that $\binom{n-1}{(n+t-1)/2} > \binom{n-t}{(n-t-1)/2}$. Indeed, by symmetry $\binom{n-1}{(n+t-1)/2} = \binom{n-1}{(n-t-1)/2}$, so the inequality is equivalent to $\binom{n-1}{(n-t-1)/2} > \binom{n-t}{(n-t-1)/2}$, which follows from the more general inequality

$$\binom{m_1}{k} > \binom{m_2}{k}$$

for $m_1 > m_2 \geq k$; this is true because $(x)_k$ is evidently increasing in x for $x \geq k$.

It also turns out (as we will soon see) that the latter construction is optimal. So in the case $n + t$ odd we take

$$\mathcal{F}_{\text{large}} := \binom{[n]}{\geq (n+t+1)/2} \cup \{A \in \binom{[n]}{(n+t-1)/2} : 1 \notin A\}.$$

How big is $\mathcal{F}_{\text{large}}$, and how does its size compare to that of $\mathcal{F}^{(A)}$, which is 2^{n-t}? We have

$$|\mathcal{F}_{\text{large}}| = \begin{cases} \sum_{k \geq (n+t)/2} \binom{n}{k} & \text{if } n+t \text{ even,} \\ \binom{n-1}{(n+t-1)/2} + \sum_{k \geq (n+t+1)/2} \binom{n}{k} & \text{if } n+t \text{ odd.} \end{cases}$$

When $t = 1$ the two families have equal size for all n, but for $t > 1$ a little numerical computation suggests that $\mathcal{F}_{\text{large}}$ has significantly more elements than $\mathcal{F}^{(A)}$ has. To get a sense of what is going on, we take a look at the sizes of the levels of the Boolean cube.

45. The levels of the Boolean cube, and Stirling's formula

The size of the kth level of the Boolean cube is $\binom{n}{k}$. The sequence $(\binom{n}{k})_{k=0}^{n}$ is easily seen to be unimodal, with mode around $n/2$. Specifically, for even n we have

$$\binom{n}{0} < \binom{n}{1} < \ldots < \binom{n}{n/2-1} < \binom{n}{n/2} > \binom{n}{n/2+1} > \ldots > \binom{n}{n-1} > \binom{n}{n}$$

while for odd n we have

$$\binom{n}{0} < \binom{n}{1} < \ldots < \binom{n}{(n-3)/2} < \binom{n}{(n-1)/2} = \binom{n}{(n+1)/2} > \ldots > \binom{n}{n-1} > \binom{n}{n}.$$

The ratio of consecutive binomial coefficients is

$$\frac{\binom{n}{k}}{\binom{n}{k-1}} = \frac{n-k+1}{k}$$

which is greater than 1 and decreasing (rapidly at first) as k increases from 1 to around $n/2$, and then less than 1 and decreasing as k increases from around $n/2$ to $n-1$. This says that the binomial coefficients increase rapidly to begin with, then increase more slowly, until k gets to around $n/2$, at which point the ratio is very close to 1 and the coefficients are barely changing; then this pattern is mirrored beyond k around $n/2$, with a slow decrease followed by a rapid decrease.

We try to quantify how slowly the binomial coefficients are changing when k is close to $n/2$. For convenience, we consider only the case when $n = 2m$ is even. It will be convenient to re-parameterize slightly, and think about binomial coefficients of the form $\binom{2m}{m-k}$ for $k = 0, 1, 2, \ldots$. We have, for all k,

$$\binom{2m}{m-k} \leq \binom{2m}{m}.$$

On the other hand, we also have

$$\begin{aligned} \frac{\binom{2m}{m-k}}{\binom{2m}{m}} &= \frac{(m)_k}{(m+k)_k} \\ &\geq \left(\frac{m-k}{m}\right)^k \\ &= \left(1 - \frac{k}{m}\right)^k \\ &\geq 1 - \frac{k^2}{m}. \end{aligned}$$

In the first inequality we have used

$$\frac{a+c}{b+c} \geq \frac{a}{b}$$

for positive a, b, c with $b \geq a$ (the "batting average inequality": if Derek Jeter has had a hits in b at-bats up to some point in the season, and then gets a base hit in each of his next c at bats, his batting average cannot decrease). The second inequality is a special case of the more general, and quite useful, inequality

$$(1+x)^r \begin{cases} \geq 1 + rx & \text{if } x \geq -1,\ r \in \mathbb{R} \setminus (0,1) \\ \leq 1 + rx & \text{if } x \geq -1,\ r \in [0,1]. \end{cases}$$

We only need this for $r \in \mathbb{N}$ in which case the proof is an easy induction. It follows that for all $k \geq 0$,

$$\left(1 - \frac{k^2}{m}\right)\binom{2m}{m} \leq \binom{2m}{m-k} \leq \binom{2m}{m}.$$

Suppose now that $k = k(n)$ is any function of n with the property that $k(n) = o(\sqrt{n})$, that is, that $k(n)/\sqrt{n} \to 0$ as $n \to \infty$. Then $k^2/m \to 0$ as $n \to \infty$, and we have the following: as $n \to \infty$,

$$\binom{2m}{m} \sim \binom{2m}{m-k}$$

(that is, the limit of the ratios tends to 1 as n goes to infinity). This says that for large n, all binomial coefficients of the form $\binom{2m}{m-k}$ have the same size, roughly that of $\binom{2m}{m}$, as long as k is smaller than $\sqrt{n}$.

In particular, as long as $k \leq \sqrt{m/2}$ we have

$$\frac{1}{2}\binom{2m}{m} \leq \binom{2m}{m-k} \leq \binom{2m}{m},$$

so at the expense of a constant factor in our estimates we can treat all of these near-middle binomial coefficient as being the same.

We now use this to estimate $\binom{2m}{m}$. Of course, we have the very basic estimates

$$\frac{2^{2m}}{2m+1} \leq \binom{2m}{m} \leq 2^{2m}$$

obtained from noticing that $\binom{2m}{m}$ is the largest term in the $(2m+1)$-term binomial expansion $(1+1)^{2m}$. The lower bound here plays a key role in Erdős' proof of Bertrand's postulate[19] that there is always a prime between n and $2n$. For a better estimate: on the one hand, just looking at the coefficients of the binomial expansion of $(1+1)^n$ of the form $\binom{2m}{m\pm k}$ for $k = 1, 2, \ldots, \left\lfloor \sqrt{m/2} \right\rfloor$, we have

$$\begin{aligned} 2^n &= \sum_{\ell=0}^{2m} \binom{2m}{\ell} \\ &\geq 2\left\lfloor \sqrt{m/2} \right\rfloor \binom{2m}{m - \left\lfloor \sqrt{m/2} \right\rfloor} \\ &\geq \left\lfloor \sqrt{m/2} \right\rfloor \binom{2m}{m}, \end{aligned}$$

so

$$\binom{2m}{m} \leq \frac{2^{2m}}{\left\lfloor \sqrt{m/2} \right\rfloor} \leq 2\frac{2^{2m}}{\sqrt{m}}, \tag{44}$$

the last inequality valid for all large enough m. On the other hand, we have, for $k \geq 0$,

$$\frac{\binom{2m}{m-k}}{\binom{2m}{m-k-1}} = \frac{m+k+1}{m-k},$$

a quantity that is increasing as k increases. It follows that for any fixed $\ell \geq 0$ the sum $\binom{2m}{m-\ell} + \binom{2m}{m-\ell-1} + \ldots + \binom{2m}{0}$ is dominated by $\binom{2m}{m-\ell}$ times the sum of an infinite geometric

[19]For a treatment in the same spirit as this course, see `http://www3.nd.edu/~dgalvin1/pdf/bertrand.pdf`.

series with first term 1 and common ratio $(m-\ell)/(m+\ell+1)$; in other words,

$$\sum_{k=0}^{m-\ell}\binom{2m}{k} \leq \binom{2m}{m-\ell}\left(\frac{m+\ell+1}{2\ell+1}\right) \leq 2^{2m}\left(\frac{2}{\sqrt{m}}\right)\left(\frac{m+\ell+1}{2\ell+1}\right),$$

the second inequality using (44) and assuming m sufficiently large. Applying with $\ell = \lfloor 5\sqrt{m} \rfloor$ and using symmetry, and assuming m sufficiently large we get

$$\begin{aligned}(2\lfloor 5\sqrt{m}\rfloor + 1)\binom{2m}{m} &\geq \sum_{k=m-\lfloor 5\sqrt{m}\rfloor}^{m+\lfloor 5\sqrt{m}\rfloor}\binom{2m}{k} \\ &\geq \frac{1}{2}2^{2m},\end{aligned}$$

so that, again for all sufficiently large m, we get

$$\binom{2m}{m} \geq \frac{1}{21}\frac{2^{2m}}{\sqrt{m}}. \tag{45}$$

We have established, using only very elementary considerations, that there are constants $c, C > 0$ such that for all sufficiently large m we have

$$c\frac{2^{2m}}{\sqrt{m}} \leq \binom{2m}{m} \leq C\frac{2^{2m}}{\sqrt{m}}. \tag{46}$$

A similar argument establishes

$$c'\frac{2^{2m}}{\sqrt{m}} \leq \binom{2m+1}{m} = \binom{2m+1}{m+1} \leq C'\frac{2^{2m}}{\sqrt{m}} \tag{47}$$

for constants $c', C' > 0$. It follows that all binomial coefficients of the form $\binom{n}{\lfloor n/2\rfloor \pm k}$ for $k = o(\sqrt{n})$ have magnitude, up to a constant, $2^n/\sqrt{n}$.

A more precise estimate is provided by *Stirling's formula*.

Theorem 45.1.

$$\lim_{n\to\infty}\frac{n!}{n^n e^{-n}\sqrt{2\pi n}} = 1.$$

A little algebra gives as a corollary that

$$\binom{2m}{m} \sim \frac{2^{2m}}{\sqrt{\pi m}},$$

a much sharper result than we could obtain using our elementary methods.

46. Back to intersecting set systems

Recall that we had two candidates for an extremal t-intersecting family on ground set $[n]$: namely, $\mathcal{F}^{([t])}$, which has size 2^{n-t}, and

$$\mathcal{F}_{\text{large}} = \begin{cases} \binom{[n]}{\geq (n+t)/2} & \text{if } n+t \text{ even,} \\ \binom{[n]}{\geq (n+t+1)/2} \cup \{A \in \binom{[n]}{(n+t-1)/2} : 1 \notin A\} & \text{if } n+t \text{ odd.}\end{cases}$$

which has size

$$|\mathcal{F}_{\text{large}}| = \begin{cases} \sum_{k \geq (n+t)/2} \binom{n}{k} & \text{if } n+t \text{ even,} \\ \binom{n-1}{(n+t-1)/2} + \sum_{k \geq (n+t+1)/2} \binom{n}{k} & \text{if } n+t \text{ odd.} \end{cases}$$

For each fixed t, as n grows, it is clear from the estimates of the last section that $\mathcal{F}_{\text{large}}$ has essentially 2^{n-1} elements, a factor 2^{t-1} more than $\mathcal{F}^{([t])}$: we are just losing from the full top half of the cube (size 2^{n-1}) about $t/2$ levels near the middle (each of size about $2^{n-1}/\sqrt{n}$), so asymptotically the size remains 2^{n-1}. The following theorem is due to Katona.

Theorem 46.1. *For each $n \geq t \geq 1$, if $\mathcal{F}$ is a t-intersecting family on ground set $[n]$, then $|\mathcal{F}| \leq |\mathcal{F}_{\text{large}}|$.*

We're going to prove this using the powerful method of *compression*, which requires a little background.

47. Compression

Set systems have very little inherent structure; this is one of the things that makes many problems concerning set systems so difficult. The method of *compression*, or *shifting*, is a way to impose structure on a set system, exploiting an (arbitrary) linear order on the elements of the ground set. The idea is to take a set system with a certain property, and push it as much as possible "towards the left" (towards having smaller elements), without changing the size of the system, or losing the key property enjoyed by the system. Once no more pushing towards the left can be done, it is possible that the system now has some structure that can be used to prove things about its size.

We assume throughout that we are working with ground set $[n]$, and we use the usual $<$ order on $\{1, \ldots, n\}$.

Definition 47.1. *Fix $1 \leq i < j \leq n$. For $A \subseteq [n]$ with $j \in A$ and $i \notin A$, the ij-*shift *of A (or ij-*compression*) is the set $S_{ij}(A) := (A \setminus \{j\}) \cup \{i\}$.*

Notice that $|S_{ij}(A)| = |A|$. Notice also that S_{ij} is a bijection from the set of sets that include j but not i to the set of sets that include i but not j.

Definition 47.2. *Fix $1 \leq i < j \leq n$. Let $\mathcal{F}$ be a set system. Let $\mathcal{F}_{ij}$ be those $A \in \mathcal{F}$ that include j, don't include i, and have the property that $S_{ij}(A)$ is not in $\mathcal{F}$. The ij-*shift *of $\mathcal{F}$ (or ij-*compression*) is the set system*

$$S_{ij}(\mathcal{F}) = (\mathcal{F} \setminus \mathcal{F}_{ij}) \cup \{S_{ij}(A) : A \in \mathcal{F}_{ij}\}.$$

Notice that $|S_{ij}(\mathcal{F})| = |\mathcal{F}|$.

A set system $\mathcal{F}$ is said to be *stable* or *invariant* if for every $i < j$, $S_{ij}(\mathcal{F}) = \mathcal{F}$. Notice that being invariant does not mean that for every $i < j$, there is no $A \in \mathcal{F}$ with $j \in A$ and $i \notin A$; it means that for every $i < j$ and every $A \in \mathcal{F}$ satisfying $j \in A$ and $i \notin A$, it must be that $(A \setminus \{j\}) \cup \{i\} \in \mathcal{F}$.

Claim 47.3. *For every set system $\mathcal{F}$, the iterative operation of: "locate an $i < j$ such that $S_{ij}(\mathcal{F}) \neq \mathcal{F}$, perform the ij-shift on $\mathcal{F}$ (that is, replace $\mathcal{F}$ by $S_{ij}(\mathcal{F})$), repeat until no such $i < j$ exists" always terminates after a finite number of iterations with a stable set system.*

Proof. Given a set system $\mathcal{F}$, let $s(\mathcal{F}) = \sum_{A \in \mathcal{F}} \sum_{k \in A} k$. Suppose that $\mathcal{F}$ is not invariant. Then for any $i < j$ for which $S_{ij}(\mathcal{F}) \neq \mathcal{F}$, we have $s(S_{ij}(\mathcal{F})) < s(\mathcal{F})$. But s is a non-negative quantity, so its value can only be reduced finitely many times. $\square$

The operation of shifting preserves many properties of set-systems. For example:

Claim 47.4. *Let $\mathcal{F}$ be a set system on ground set $[n]$ that is t-intersecting. Fix $1 \leq i < j \leq n$. Let $\mathcal{F}' = S_{ij}(\mathcal{F})$. Then $\mathcal{F}'$ is also t-intersecting.*

Proof. Fix $A', B' \in \mathcal{F}'$. If $A', B' \in \mathcal{F}$, then immediately we have $|A' \cap B'| \geq t$. If $A', B' \notin \mathcal{F}$, then $A := (A' \setminus \{i\}) \cup \{j\}, B := (B' \setminus \{i\}) \cup \{j\} \in \mathcal{F}$. We have $|A \cap B| \geq t$ and $A' \cap B' = (A \cap B \setminus \{j\}) \cup \{i\}$; since $i \notin A \cap B$, we get $|A' \cap B'| = |A \cap B| \geq t$.

There remains the case $A' \in \mathcal{F}, B' \notin \mathcal{F}$. In this case we have $B \in \mathcal{F}$ for $B := (B' \setminus \{i\}) \cup \{j\}$. Thus, $i \notin B$ and $|A' \cap B| \geq t$. We treat three subcases.

- If $j \notin A' \cap B$, then by the construction of B' from B (removing j and adding i), $|A' \cap B'| \geq |A' \cap B| \geq t$.
- If $j \in A' \cap B$ and $i \in A'$, then by the construction of B' from B, $|A' \cap B'| = |A' \cap B| \geq t$.
- If $j \in A' \cap B$ and $i \notin A'$, then A' was a candidate for shifting; since it was not shifted, there is $A'' \in \mathcal{F}$ with $A'' = (A' \setminus \{j\}) \cup \{i\}$. We have $|A'' \cap B| \geq t$ and $|A' \cap B'| = |A'' \cap B|$ (i and j are not involved in either intersection), so $|A' \cap B'| \geq t$.

□

48. Proof of Theorem 46.1

We prove the theorem only for $n + t$ even; the case $n + t$ odd is very similar and left as an exercise. We proceed by induction on $n \geq 1$, the induction hypothesis $P(n)$ being "for all t satisfying $n \geq t \geq 1$ such that $n + t$ is even, if $\mathcal{F}$ is a t-intersecting set system on ground set $[n]$ then $|\mathcal{F}| \leq \sum_{k \geq (n+t)/2} \binom{n}{k}$". For $n = 1$, the only relevant value of t is $t = 1$, and the result is trivial. For $n = 2$, the only relevant value of t is $t = 2$, and again the result is trivial. For $n = 3$, the relevant values of t are $t = 3$ (for which the result is trivial), and $t = 1$, which is an instance of our previous result on intersecting families. So now we assume $n \geq 4$, $n \geq t \geq 1$, $n + t$ even, and that the result is true for all pairs (n', t') with $n' \geq t' \geq 1$ with $n' + t'$ even and with $n' < n$.

By Claim 47.4 we may assume that $\mathcal{F}$ is stable under shifting. To apply induction, we decompose the set $\mathcal{F}$ into two parts, both on a smaller ground set:

$$\mathcal{F}_1 := \{A \setminus \{1\} : A \in \mathcal{F},\ 1 \in A\}$$

and

$$\mathcal{F}_0 := \{A : A \in \mathcal{F},\ 1 \notin A\}.$$

Clearly $\mathcal{F}_1$ is $(t-1)$-intersecting on a ground set of size $n - 1$, so by induction

$$|\mathcal{F}_1| \leq \sum_{k \geq (n+t-2)/2} \binom{n-1}{k} = \sum_{k \geq (n+t)/2} \binom{n-1}{k-1}.$$

Equally clearly $\mathcal{F}_0$ is t-intersecting on ground set of size $n - 1$, but this does not allow us to apply induction, as $n + t - 1$ is odd. The main point of the proof is that (as we will prove below) $\mathcal{F}_0$ is $(t+1)$-intersecting! This allows us to say by induction that

$$|\mathcal{F}_0| \leq \sum_{k \geq (n+t)/2} \binom{n-1}{k}.$$

Since $|\mathcal{F}| = |\mathcal{F}_1| + |\mathcal{F}_0|$, Pascal's identity gives $|\mathcal{F}| \leq \sum_{k \geq (n+t)/2} \binom{n}{k}$, as required.

To prove the $(t+1)$-intersecting claim, fix $A, B \in \mathcal{F}_0$, and $j \in A \cap B$. Since $j \in A$ and $1 \notin A$, it follows that $A' := (A \setminus \{j\}) \cup \{1\} \in \mathcal{F}$ (otherwise $\mathcal{F}$ would not be stable). Now

$|A' \cap B| \geq t$; but since $1 \notin B$ and $j \in B$, $A \cap B = (A' \cap B) \cup \{j\}$, and so since $j \notin A'$, we therefore have $|A \cap B| = |A' \cap B| + 1 \geq t + 1$.

49. The Erdős-Ko-Rado theorem

We have been thinking about intersecting families where the elements are allowed to have any size. Now we restrict our attention to individual levels of the Boolean cube, where all elements of the set system are required to have the same size. This is sometimes called a *uniform* set system (or *uniform hypergraph*).

So: Let $\mathcal{F}$ be an intersecting set system in $\binom{[n]}{r}$ for some $0 \leq r \leq n$. How large can $\mathcal{F}$ be? (Notice that we are just considering *intersecting* families here, not t-intersecting). If $n < 2r$ then this is trivial: $|\mathcal{F}| \leq \binom{n}{r}$, and the set system $\binom{[n]}{r}$ itself shows that this bound can be achieved. If $n = 2r$ (so necessarily n even), things are more interesting. At most one of each pair $(A, [n] \setminus A)$ can be in $\mathcal{F}$, so $|\mathcal{F}| \leq \binom{n}{n/2}/2 = \binom{n-1}{n/2-1}$, and (for example) the family consisting of all subsets of size $n/2$ of $[n]$ that include element 1 gives an example of an intersecting family of that size.

More generally, for $n \geq 2r$ the family of elements of $\binom{[n]}{r}$ that all include a particular element gives rise to an r-uniform intersecting family of size $\binom{n-1}{r-1}$, and it is not clear how one may do better. That it is not possible to do better is the content of the following theorem, due to Erdős, Ko and Rado, one of the first theorems in extremal set theory.

Theorem 49.1. *Fix $n \geq 2r$, $r \geq 1$. Let $\mathcal{F}$ be an r-uniform intersecting family. Then $|\mathcal{F}| \leq \binom{n-1}{r-1}$.*

The centrality of this result in the theory of set systems is attested to by the fact that at least six substantially different proofs have been presented. We present two, the first a gem of double-counting, and the second using compression/shifting.

Proof. (Katona's proof of Theorem 49.1) The idea of this proof is to put structure on a set system by imposing a cyclic order on the ground set. We shall use permutations instead of cyclic orders, in order to deal with more familiar objects.

If σ is a permutation of $[n]$, then we shall denote the periodic sequence

$$(\sigma(1), \sigma(2), \ldots, \sigma(n), \sigma(1), \sigma(2), \ldots, \sigma(n), \sigma(1), \sigma(2), \ldots, \sigma(n), \ldots)$$

by $\overline{\sigma}$. For each permutation σ of $[n]$, we count the number of elements of $\mathcal{F}$ that occur as a consecutive block in the periodic sequence $\overline{\sigma}$. Specifically, consider the set $\mathcal{P}$ that consists of all pairs (A, σ), where A is an element of our r-uniform intersecting family $\mathcal{F}$, where σ is a permutation of $[n]$, and where A occurs as a consecutive block in $\overline{\sigma}$, meaning that there is some k, $1 \leq k \leq n$, such that $A = \{\sigma(k), \sigma(k+1), \ldots, \sigma(k+r-1)\}$, where addition is performed modulo n (so $n+1 = 1$, $n+2 = 2$, etc.).

For each of the $|\mathcal{F}|$ choices of $A \in \mathcal{F}$, there are exactly $n|A|!(n-|A|)! = nr!(n-r)!$ permutations σ such that A occurs as a consecutive block in $\overline{\sigma}$ (we may choose an arbitrary $k \in [n]$ to satisfy $A = \{\sigma(k), \sigma(k+1), \ldots, \sigma(k+r-1)\}$, and we may order the elements of A and the elements of $[n] \setminus A$ arbitrarily), and so

$$|\mathcal{P}| = |\mathcal{F}|nr!(n-r)!.$$

Now we count in the other direction. Fix a permutation σ. A priori, there can be at most n different $A \in \mathcal{F}$ such that A occurs as a consecutive block in $\overline{\sigma}$ (one for each of the consecutive blocks of size r), so $|\mathcal{P}| \leq n(n-1)!$. This leads to the trivial bound $|\mathcal{F}| \leq \binom{n}{r}$. So we need a

better bound on the number of different $A \in \mathcal{F}$ such that A occurs as a consecutive block in $\overline{\sigma}$. Suppose that $\{\sigma(k), \sigma(k+1), \ldots, \sigma(k+r-1)\} \in \mathcal{F}$. There are $2r-1$ consecutive blocks of length r in $\overline{\sigma}$ that intersect this block: the blocks $B_{-r+1}, B_{-r+2}, \ldots, B_{r-1}$, where $B_j = \{\sigma(k+j), \sigma(k+j+1), \ldots, \sigma(k+j+r-1)\}$. (Notice that $B_0 = \{\sigma(k), \sigma(k+1), \ldots, \sigma(k+r-1)\}$.) Because $n \geq 2r > 2r-1$, these $2r-1$ blocks are all distinct (if we imagine the n distinct numbers $\sigma(1), \sigma(2), \ldots, \sigma(n)$ drawn counterclockwise around a circle, then the $2r-1$ blocks $B_{-r+1}, B_{-r+2}, \ldots, B_{r-1}$ correspond to $2r-1$ arcs on this circle, and the inequality $n > 2r-1$ ensures that no two of these arcs cover the same set of numbers). Now we can partition these $2r-1$ blocks into a singleton part (the block B_0) together with $r-1$ pairs (the pair (B_{-r+1}, B_1), the pair (B_{-r+2}, B_2), etc., the pair (B_{-1}, B_{r-1})), and, again because $n \geq 2r$, this partition has the property that pairs of blocks in a single class are disjoint (i.e., the blocks B_{-r+j} and B_j are disjoint whenever $j \in [r-1]$). Hence at most one of them in each pair can be in $\mathcal{F}$ (since $\mathcal{F}$ is intersecting), and so the number of different $A \in \mathcal{F}$ such that A occurs as a consecutive block in $\overline{\sigma}$ is at most $1 + (r-1) = r$. (All this was on the supposition that there is at least one such A; but if there is not, the bound we have established still holds, since $0 \leq r$). It follows that

$$|\mathcal{P}| \leq rn!.$$

Comparing this with $|\mathcal{P}| = |\mathcal{F}|nr!(n-r)!$, we find

$$|\mathcal{F}| \leq \frac{rn!}{nr!(n-r)!} = \frac{(n-1)!}{(r-1)!(n-r)!} = \binom{n-1}{r-1}.$$

□

Proof. (Compression proof of Theorem 49.1) For a proof based on compression, we proceed by induction on $n \geq 2$. Our inductive hypothesis $P(n)$ will be "for all r satisfying $n \geq 2r$, $r \geq 1$, if $\mathcal{F}$ is an r-uniform intersecting family on ground set $[n]$ then $|\mathcal{F}| \leq \binom{n-1}{r-1}$". The base case $n = 2$ is easy, so assume $n > 2$. The case $n = 2r$ is easy (we have already established it in the discussion preceding the statement of the theorem), so we may assume $n > 2r$. Let $\mathcal{F}$ be an intersecting r-uniform family, which we may assume to be stable under compression. The key observation is that for $A, B \in \mathcal{F}$, not only is $A \cap B \neq \emptyset$, but also $A \cap B \cap [2r] \neq \emptyset$.

The proof of this uses that fact that $\mathcal{F}$ is stable. Suppose for contradiction that there are $A, B \in \mathcal{F}$ with $|A \cap B \cap [2r]| = 0$, and with $|A \cap B|$ minimal subject to this condition. Since $|A \cap B| \neq 0$ (and $|A \cap B \cap [2r]| = 0$) there is some $j > 2r$ such that $j \in A \cap B$. From $|A \cap B| \neq 0$, we also obtain $|A \cup B| < |A| + |B| = r + r = 2r$; hence, $[2r]$ cannot be a subset of $A \cup B$. In other words, there is some $i \in [2r]$ with $i \notin A \cup B$. Clearly, $j > 2r \geq i$. Since $i \notin A$, $j \in A$, $j > i$ and $\mathcal{F}$ is stable, we have that $A' := A \setminus \{j\} \cup \{i\} \in \mathcal{F}$. But now $A' \cap B = (A \cap B) \setminus \{j\}$. Therefore, $|A' \cap B| < |A \cap B|$ and $|A' \cap B \cap [2r]| = 0$, contradicting the minimality of (A, B).

Decompose the set $\mathcal{F}$ according to the sizes of the intersection of elements of $\mathcal{F}$ with $[2r]$; meaning, set, for $k = 0, \ldots, r$,

$$\mathcal{F}_k = \{A \in \mathcal{F} : |A \cap [2r]| = k\}.$$

Notice that since $A \cap B \cap [2r] \neq \emptyset$ for $A, B \in \mathcal{F}$, it follows that $A \cap [2r] \neq \emptyset$, so $\mathcal{F}_0 = \emptyset$, and $|\mathcal{F}| = \sum_{k=1}^{r} |\mathcal{F}_k|$. For $k = 1, \ldots, r$ set

$$\mathcal{F}'_k = \{A \cap [2r] : A \in \mathcal{F}_k\}.$$

Each element of $\mathcal{F}'_k$ extends to[20] at most $\binom{n-2r}{r-k}$ elements of $\mathcal{F}_k$, so

$$|\mathcal{F}| \leq \sum_{k=1}^{r} |\mathcal{F}'_k| \binom{n-2r}{r-k}. \tag{48}$$

But we know that $\mathcal{F}'_k$ is a k-uniform intersecting family on ground set $[2r]$, and $n > 2r \geq 2k \geq k \geq 1$, and so by induction $|\mathcal{F}'_k| \leq \binom{2r-1}{k-1}$, and therefore (48) becomes

$$|\mathcal{F}| \leq \sum_{k=1}^{r} \binom{2r-1}{k-1}\binom{n-2r}{r-k} = \sum_{k=0}^{r-1} \binom{2r-1}{k}\binom{n-2r}{r-1-k} = \binom{n-1}{r-1}$$

by the Vandermonde identity. □

50. A quick statistical application of Erdős-Ko-Rado

Suppose we have access to n observations, $Y_1, \ldots, Y_n$, of a random variable Y that takes value 1 with some (unknown) probability $p \geq 1/2$, and value 0 with probability $1-p$. We want to use the observations to predict the value of an $(n+1)$-st observation. Here's a possible scheme: fix a vector $(\alpha_1, \ldots, \alpha_n)$ of non-negative reals that sum to 1, with the further property that $\sum_{i\in A} \alpha_i \neq 1/2$ for any $A \subseteq [n]$. We then predict that the $(n+1)$-st observation will be

$$\begin{array}{ll} 1 & \text{if } \sum_{i=1}^n \alpha_i Y_i > 1/2, \\ 0 & \text{if } \sum_{i=1}^n \alpha_i Y_i < 1/2. \end{array}$$

(Notice that by imposing the condition $\sum_{i\in A} \alpha_i \neq 1/2$, we ensure that we never have to break a tie when the α-linear combination of observations is exactly $1/2$).

The probability that we are in error using our estimation rule is

$$\begin{aligned} &\Pr(Y=1)(1-\Pr(\sum_{i=1}^n \alpha_i Y_i > 1/2)) + \Pr(Y=0)\Pr(\sum_{i=1}^n \alpha_i Y_i > 1/2) \\ &= p - (2p-1)\Pr(\sum_{i=1}^n \alpha_i Y_i > 1/2). \end{aligned}$$

Let's look at the special case $n=1$. Here, the only valid prediction rule of the kind described is given by $\alpha_1 = 1$, so $\Pr(\sum_{i=1}^n \alpha_i Y_i > 1/2) = p$. Liggett observed the following corollary of the Erdős-Ko-Rado theorem:

Corollary 50.1. *With the notation as above,*

$$\Pr(\sum_{i=1}^n \alpha_i Y_i > 1/2) \geq p.$$

The interpretation of this result is that any prediction scheme that we might construct (of the kind described above) will certainly have no greater probability of error than the trivial scheme when $n=1$ (assuming, it must be re-iterated, that we have the information that $p \geq 1/2$).

[20] "extends to" means "is a subset of".

Proof. (of Corollary 50.1) For each $1 \leq r \leq n$ let $\mathcal{F}_r$ be the family of subsets $A \in \binom{[n]}{r}$ such that $\sum_{i \in A} \alpha_i > 1/2$. Each $\mathcal{F}_r$ is evidently an r-uniform intersecting family, and so by Theorem 49.1,

$$|\mathcal{F}_r| \leq \binom{n-1}{r-1}$$

for each r with $n \geq 2r \geq 1$.

What about $|\mathcal{F}_r|$ for larger r? The key observation here is that for any r,

$$|\mathcal{F}_r| + |\mathcal{F}_{n-r}| = \binom{n}{r}, \tag{49}$$

since for each $A \in \binom{[n]}{r}$ exactly one of A, $[n] \setminus A$ has the property that the sum of the corresponding α_i's exceeds $1/2$.

Now we have

$$\Pr(\sum_{i=1}^{n} \alpha_i Y_i > 1/2) = \sum_{r=0}^{n} |\mathcal{F}_r| p^r (1-p)^{n-r}.$$

In order to easily compare this with p, we write

$$p = p \sum_{r=1}^{n} \binom{n-1}{r-1} p^{r-1}(1-p)^{n-r} = \sum_{r=0}^{n} \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

(adopting the usual convention that $\binom{n}{-1} = 0$). Now we have

$$\Pr(\sum_{i=1}^{n} \alpha_i Y_i > 1/2) - p = \sum_{r=0}^{n} \left[|\mathcal{F}_r| - \binom{n-1}{r-1} \right] p^r (1-p)^{n-r}.$$

The net contribution to this sum from summation indices r and $n-r$ ($r < n/2$) is

$$\left[|\mathcal{F}_r| - \binom{n-1}{r-1} \right] p^r (1-p)^{n-r} + \left[|\mathcal{F}_{n-r}| - \binom{n-1}{n-r-1} \right] p^{n-r} (1-p)^r$$

which, after setting $\binom{n-1}{n-r-1} = \binom{n-1}{r}$ and using (49), and applying Pascal's identity, simplifies to

$$\left[|\mathcal{F}_r| - \binom{n-1}{r-1} \right] \left(p^r (1-p)^{n-r} - p^{n-r} (1-p)^r \right).$$

If n is even and $r = n/2$, then (49) says that $|\mathcal{F}_{n/2}| = (1/2)\binom{n}{r} = \binom{n-1}{r-1}$. It follows that

$$\Pr(\sum_{i=1}^{n} \alpha_i Y_i > 1/2) - p = \sum_{r < n/2} \left[|\mathcal{F}_r| - \binom{n-1}{r-1} \right] \left(p^r (1-p)^{n-r} - p^{n-r} (1-p)^r \right).$$

For $p \geq 1/2$ we have $p^{n-r}(1-p)^r \geq p^r (1-p)^{n-r}$, and by Erdős-Ko-Rado $|\mathcal{F}_r| \leq \binom{n-1}{r-1}$, so the sum is non-negative, as required. □

A final note on Erdős-Ko-Rado: any intersecting family in the Boolean cube can be extended to one of maximum possible cardinality, namely 2^{n-1}; in other words, for intersecting set systems the notions of maximal and maximum coincide (this is exercise (1) in the next section). This is not the case for uniform intersecting set systems. Perhaps the simplest example here is the *Fano plane*. This is the 3-uniform set system on ground set $\{a, b, c, m_{ab}, m_{ac}, m_{bc}, o\}$ consisting of the seven sets $\{a, m_{ab}, b\}$, $\{b, m_{bc}, c\}$, $\{c, m_{ac}, a\}$,

$\{a, o, m_{bc}\}$, $\{b, o, m_{ac}\}$, $\{c, o, m_{ab}\}$ and $\{m_{ab}, m_{bc}, m_{ac}\}$. This is easily checked to be intersecting, and maximal (every other subset of size 3 of the ground set is disjoint from at least one of these seven sets); but it is not maximum, since $7 < 15 = \binom{6}{2}$. The Fano plane is the finite projective plane of order 2; see, for example, `http://en.wikipedia.org/wiki/Fano_plane`. In the picture on that page a, b and c are the vertices of the triangle, the m's are the midpoints, and o is the center.

51. Some problems

(1) Let $\mathcal{F}$ be an intersecting set system on ground set $[n]$. Show that there exists an intersecting set system $\mathcal{F}'$ on $[n]$ with $|\mathcal{F}'| = 2^{n-1}$ and with $\mathcal{F} \subseteq \mathcal{F}'$. (This shows that the notions of maximal and maximum coincide for intersecting set systems.)

Solution: Among all the intersecting set systems on ground set $[n]$ that include all the elements of $\mathcal{F}$ (note that this set is non-empty, since it includes $\mathcal{F}$ itself, and finite), let $\mathcal{F}'$ be a maximal one with respect to inclusion. We claim that if $A \subseteq [n]$ is such that $A \notin \mathcal{F}'$, then $[n] \setminus A \in \mathcal{F}'$; this would show that from each of the 2^{n-1} pairs $\{A, [n] \setminus A\}$ exactly one is in $\mathcal{F}'$, so that indeed $|\mathcal{F}'| = 2^{n-1}$.

If $A \subseteq [n]$ is such that $A \notin \mathcal{F}'$ then by maximality of $\mathcal{F}'$ there is $B \in \mathcal{F}'$ that is disjoint from A. Suppose also that $[n] \setminus A \notin \mathcal{F}'$. Then again by maximality of $\mathcal{F}'$ there is $C \in \mathcal{F}'$ that is disjoint from $[n] \setminus A$. But then B and C are disjoint (since $B \subseteq [n] \setminus A$ and $C \subseteq A$), a contradiction, so indeed $[n] \setminus A \in \mathcal{F}'$.

(2) Let $n = 2m$ be even. Let $\mathcal{F} \subseteq \binom{[2m]}{m}$ be the set of all subsets of $[2m]$ of size m that include element 1. Let $\mathcal{F}'$ be an arbitrary subset (possibly empty) of $\mathcal{F}$. Let

$$\mathcal{F}'' = \mathcal{F}' \cup \{[2m] \setminus A : A \in \mathcal{F} \setminus \mathcal{F}'\}.$$

Show that $\mathcal{F}''$ is an intersecting family.

Solution: Note that $\{[2m] \setminus A : A \in \mathcal{F} \setminus \mathcal{F}'\}$ is a collection of m-element subsets, none of which contain 1, so is disjoint from $\mathcal{F}'$. Note also that since the map $A \mapsto [2m] \setminus A$ is a bijection, $|\mathcal{F}''| = |\mathcal{F}|$.

Let E_1 and E_2 be two elements of $\mathcal{F}''$.

- If $E_1, E_2 \in \mathcal{F}'$ then $E_1 \cap E_2 \neq \emptyset$ since $1 \in E_1 \cap E_2$.
- If one of E_1, E_2 is in $\mathcal{F}'$, say, without loss of generality, E_1, and the other, E_2, is not in $\mathcal{F}'$, then there is $A_2 \in \mathcal{F} \setminus \mathcal{F}'$ with $E_2 = [2m] \setminus A_2$. Since $A_2 \neq E_1$ (and both contain 1 and are of size m) there is $j \in E_1$ with $j \notin A_2$, so $j \in E_2$ and $j \in E_1 \cap E_2$.
- If $E_1, E_2 \notin \mathcal{F}'$ then there are $A_1, A_2 \in \mathcal{F} \setminus \mathcal{F}'$, distinct, with $E_i = [2m] \setminus A_i$ for $i = 1, 2$. Since A_1, A_2 are both of size m and have an element in common, there must be $j \in [2m]$ with $j \notin A_i$ for $i = 1, 2$; so $j \in E_1 \cap E_2$.

(3) Use the results of the last two questions to show that if $n = 2m$ is even then there are at least $2^{\binom{2m-1}{m-1}}$ intersecting families on ground set $[n]$ that have size 2^{n-1}. [Note that this is a large number, approximately $2^{n^{-1/2}2^n}$.]

Solution: The set system $\mathcal{F}$ from the last question has size $\binom{2m-1}{m-1}$. As observed in the solution to that question, for each $\mathcal{F}' \subseteq \mathcal{F}$ we get a distinct set system $\mathcal{F}''$ also of size $\binom{2m-1}{m-1}$. All of these set systems are intersecting and m-uniform (by the last question). There are also all maximal as m-uniform intersecting set systems (from Erdős-Ko-Rado). Thus we have a collection of $2^{\binom{2m-1}{m-1}}$ maximal m-uniform intersecting families. By the first question, each of these is contained in a maximal

intersecting set system (one of size 2^{n-1}), and these maximal set systems are still distinct, being distinguished by their collections of sets of size m.

(4) An *ideal* is a set system $\mathcal{I}$ on ground set $[n]$ that is closed under taking subsets: if $A \in \mathcal{I}$ and $B \subseteq A$ then $B \in \mathcal{I}$. Prove that if $\mathcal{I}$ is an ideal, then each $i \in [n]$ appears in at most half the members of $\mathcal{I}$.

Solution: Here is a prosaic approach: an ideal is determined completely by its maximal elements, say $A_1, \ldots, A_\ell$ (which, incidentally, form an antichain). Fix an element i; without loss of generality assume that $i \in A_1 \cap A_2 \cap \ldots \cap A_k$, but $i \notin A_j$ for $j > k$. We claim that i is in exactly half of the sets in the ideal that are subsets of one of the A_j's, $1 \leq j \leq k$; since i cannot be in any other members of the ideal, this gives the result we require.

To prove the claim, we use inclusion-exclusion: For each j, $1 \leq j \leq k$, let $\mathcal{B}_j$ be the set of subsets of A_j that include element i. Notice that for any $J \subseteq [k]$, the set system $\bigcap_{j \in J} \mathcal{B}_j$ is the set of sets that are subsets of each of A_j, $j \in J$, and include i; this is the same as the set of subsets of $\bigcap_{j \in J} A_j$ that include i (note that $i \in \bigcap_{j \in J} A_j$). Now exactly half the subsets of a set contain a given element of the set, so $|\bigcap_{j \in J} \mathcal{B}_j|$ is half the size of the set of subsets of $\bigcap_{j \in J} A_j$, that is

$$\left| \bigcap_{j \in J} \mathcal{B}_j \right| = \frac{1}{2} \left| \bigcap_{j \in J} \mathcal{A}_j \right|,$$

where $\mathcal{A}_j$ is the set of subsets of A_j. Now inclusion-exclusion gives

$$\begin{aligned} \left| \bigcup_{j=1}^{k} \mathcal{B}_j \right| &= \sum_{J \subseteq [k],\ J \neq \emptyset} (-1)^{|J|-1} \left| \bigcap_{j \in J} \mathcal{B}_j \right| \\ &= \frac{1}{2} \sum_{J \subseteq [k],\ J \neq \emptyset} (-1)^{|J|-1} \left| \bigcap_{j \in J} \mathcal{A}_j \right| \\ &= \frac{1}{2} \left| \bigcup_{j=1}^{k} \mathcal{A}_j \right|; \end{aligned}$$

in other words, i is in exactly half of the sets in the ideal that are subsets of one of the A_j's, $1 \leq j \leq k$, as claimed.

Here is something rather more slick: fix $i \in [n]$. Let $\mathcal{I}_1$ be the set of sets in the ideal $\mathcal{I}$ that include element i, and let $\mathcal{I}_2$ be the set of sets that do not. By definition of ideal, if $A \in \mathcal{I}_1$ then $A \setminus \{i\} \in \mathcal{I}_2$, so the map that sends A to $A \setminus \{i\}$ is a map from $\mathcal{I}_1$ to $\mathcal{I}_2$. Moreover, it is clearly an injection, so $|\mathcal{I}_1| \leq |\mathcal{I}_2|$ and (adding $|\mathcal{I}_1|$ to both sides) $|\mathcal{I}_1| \leq (1/2)|\mathcal{I}|$, as required.

(5) Let $\mathcal{I}$ be an ideal on groundset $[n]$, and let $\mathcal{I}'$ be the set of complements of members of $\mathcal{I}$ (formally,

$$\mathcal{I}' = \{[n] \setminus A : A \in \mathcal{I}\}).$$

Prove that there is a bijection $f : \mathcal{I} \to \mathcal{I}'$ satisfying $A \subseteq f(A)$ for all $A \in \mathcal{I}$. [**Hint**: Induction on $|\mathcal{I}|$; decompose $\mathcal{I}$ into those sets that contain a fixed element, and those that don't.]

Solution: We proceed by induction on n, with the base cases $n = 1, 2$ trivial. For $n \geq 3$, if $\mathcal{I} = \{\emptyset\}$ then the result is trivial, so assume $\mathcal{I} \neq \{\emptyset\}$. Fix an i such that $\mathcal{I}_1$, the set of elements of $\mathcal{I}$ that contain element i, is non-empty; without loss of generality, $i = n$, and set $\mathcal{I}_2 = \mathcal{I} \setminus \mathcal{I}_1$. Notice that by the result of the last question, $|\mathcal{I}_1| \leq |\mathcal{I}_2|$.

The set $\mathcal{I}_2$ is an ideal on $[n-1]$, so by induction there is a bijection f_2 from $\mathcal{I}_2$ to $\mathcal{I}_2'$ (complement taken inside $[n-1]$) with $A \subseteq f_2(A)$ always. We can extend this to a bijection $\hat{f}_2$ from $\mathcal{I}_2$ to $\mathcal{I}_2'$ (complement taken inside $[n]$) with $A \subseteq \hat{f}_2(A)$ by $\hat{f}_2(A) = f_2(A) \cup \{n\}$.

Now consider $A \in \mathcal{I}_1$. If we extend $\hat{f}_2$ to $\mathcal{I}_1$ by sending A to $\hat{f}_2(A \setminus \{n\})$, then we have a map from $\mathcal{I}$ to $\mathcal{I}'$ with $A \subseteq f(A)$ for all $A \in \mathcal{I}$. But it is not a bijection, since for each $A \in \mathcal{I}_1$, we have $\hat{f}_2(A) = \hat{f}_2(A \setminus \{n\})$. We can remedy this, and turn $\hat{f}_2$ into a bijection, by, for each such A, redefining $\hat{f}_2(A \setminus \{n\})$ to be $f_2(A)$.

52. Systems of disjoint representatives, and Hall's marriage theorem

Joe Donnelly is the junior US senator from Indiana. Some of his constituents are women, some are hispanic, some are army veterans, some are farmers, and some drive Hondas. Joe wants to gather together five constituents, with each one belonging to a designated one of the five mentioned groups. (A designate of one of the groups is allowed to belong to many of the groups, but may only be designated as the representative of one group). Can he succeed? Presumably yes, but it is not certain. For example, in the unlikely event that the number of people in Indiana who are either women, or hispanic, or army veterans, or farmers, or Honda drivers, is at most four, then Joe is clearly out of luck.

What Joe is looking for is a system of distinct representatives, and in this section we establish necessary and sufficient conditions for such a thing to exist.

Definition 52.1. *Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be a set system on ground set $[n]$. A* system of distinct representatives *(SDR) for $\mathcal{F}$ is a vector $(a_1, \ldots, a_m)$ of distinct elements from $[n]$ with the property that $a_i \in A_i$ for each i, $i = 1, \ldots, m$.*

It is evident that if there is an $I \subseteq [m]$ such that $|\bigcup_{i \in I} A_i| < |I|$, then a system of distinct representatives cannot exist for $\mathcal{F}$ (and so in particular if $m > n$ then there can be no SDR). Surprisingly, this is the only obstacle to the existence of an SDR. The following theorem is due to Philip Hall, but was discovered independently by many researchers in the 1930's.

Theorem 52.2. *Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be a set system on ground set $[n]$. There is an SDR for $\mathcal{F}$ if and only if it holds that for each $I \subseteq [m]$,*

$$|\bigcup_{i \in I} A_i| \geq |I|. \tag{50}$$

We refer to (50) as *Hall's SDR condition* for I.

Before proving Theorem 52.2, we reformulate it in graph language. A *bipartite graph* $G = (X, Y, E)$ with *color classes* X and Y consists of sets $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_n\}$ ($X \cup Y$ is the set of *vertices* of the graph), together with a set E whose elements are pairs (x_i, y_j) for some $1 \leq i \leq m$, $1 \leq j \leq n$ (E is the set of *edges* of the graph). A *matching* in G is a set $M = \{(x_{i_1}, y_{i_1}), \ldots, (x_{i_k}, y_{i_k})\}$ of edges with the property that the x_{i_j}'s are distinct and the y_{i_j}'s are distinct. The matching *saturates* X if it has size m (involves m edges); in this case necessarily each of the x_i's appears once and only once among the x_{i_j}'s. The problem

of finding saturating matchings is often accompanied by the following story: there are girls and boys at a dance. Each girl likes some of the boys but not others. Is it possible for all the girls to simultaneously find a boy that she likes, if each boy can dance at any moment with at most one girl? Naming the girls $x_1, \ldots, x_m$ and the boys $y_1, \ldots, y_n$, and declaring that (x_i, y_j) is an edge if girl x_i likes boy y_j, the question becomes, does the bipartite graph so constructed have a matching that saturates X?

As with the problem of SDR's, a trivial necessary condition for such a matching to exist is that for each $I \subseteq [m]$,

$$|\{y_j : (x_i, y_j) \in E \text{ for some } i \in I\}| \geq |I|. \tag{51}$$

We refer to (51) as *Hall's marriage condition* for I. And as with for SDR's, this necessary condition is sufficient. The following, again due to Philip Hall, is commonly known as *Hall's marriage theorem*.

Theorem 52.3. *Let $G = (X, Y, E)$ be a finite bipartite graph with the vertices of X indexed by $[m]$. There is a matching that saturates X if and only if Hall's marriage condition holds for all $I \subseteq [m]$.*

Before proving the theorem, we note that it implies Theorem 52.2. Indeed, let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be a set system on ground set $[n]$. Construct from $\mathcal{F}$ a bipartite graph $G = (X, Y, E)$ as follows: $X = \{1, \ldots, m\}$, $Y = \{1, \ldots, n\}$, and $(i, j) \in E$ exactly where $j \in A_i$. Hall's SDR condition (in $\mathcal{F}$) for an $I \subseteq [m]$ translates exactly to his marriage condition (in G) for the same I, so if the SDR condition is satisfied for all I, then by Theorem 52.3 G has a perfect matching. If that matching consists of edges $(1, j_1), \ldots, (m, j_m)$, then $(j_1, \ldots, j_m)$ forms an SDR for $\mathcal{F}$.

There are many proofs of Hall's marriage theorem. Here is perhaps the shortest.

Proof. (Theorem 52.3) That it is necessary for Hall's condition to hold for all $I \subseteq [m]$ in order for there to be a matching saturating X is clear. So now, let G be such that Hall's marriage condition is satisfied for all $I \subseteq [m]$. We show that G has a matching saturating X, by induction on m, with the base case $m = 0$ trivial.

For $m > 0$, suppose first that it is the case that something stronger than Hall's condition holds: namely, for all $I \subseteq [m]$, $I \neq \emptyset, [m]$,

$$|\{y_j : (x_i, y_j) \in E \text{ for some } i \in I\}| \geq |I| + 1. \tag{52}$$

Pick j arbitrarily with $(x_m, y_j) \in E$, and consider the bipartite graph $G' = (X', Y', E')$ obtained from G by $X' = X \setminus \{x_m\}$, $Y' = Y \setminus \{y_j\}$, and E' obtained from E by deleting all edges involving either x_m or y_j. Using (52) it is easily seen that G' satisfies Hall's marriage condition for all $I \subseteq [m-1]$, so by induction there is a matching in G' that saturates X'; adding (x_m, y_j) then gives a matching in G saturating X.

Otherwise there is a $J \subseteq [m]$, $J \neq \emptyset, [m]$, such that $|N(J)| = |J|$, where $N(J) := \{y_j : (x_i, y_j) \in E \text{ for some } i \in J\}$. Consider the following two bipartite graphs: first, G_J obtained from G by restricting X to J, restricting Y to $N(J)$, and only keeping those edges in E that begin with an element of J and end with an element of $N(J)$; and second, G'_J obtained from G by restricting X to $[m] \setminus J$, restricting Y to $[n] \setminus N(J)$, and only keeping those edges in E that go between these two sets. It's clear that G_J satisfies Hall's marriage condition for all $I \subseteq J$, and so G_J has (by induction) a matching M saturating J. But also, G'_J satisfies Hall's marriage condition for all $I \subseteq [m] \setminus J$ (we will justify this in a moment), and so G'_J

has (again by induction) a matching M' saturating $[m] \setminus J$; and $M \cup M'$ is a matching in G saturating X.

To see that G'_J satisfies Hall's marriage condition for all $I \subseteq [m] \setminus J$, suppose, for a contradiction, that there is $I' \subseteq [m] \setminus J$ such that

$$|\{y_j : (x_i, y_j) \in E'_J \text{ for some } i \in I'\}| < |I'|$$

(where E'_J denotes the set of edges of the graph G'_J). Then

$$|\{y_j : (x_i, y_j) \in E \text{ for some } i \in I' \cup J\}| < |I'| + |N(J)| = |I'| + |J| = |I' \cup J|,$$

the required contradiction. □

As a quick application of Hall's theorem, we prove the following, which will be useful later.

Proposition 52.4. *Fix $n \geq 1$. Let $k \geq 0$ be such that $k + 1 \leq \lceil n/2 \rceil$. Construct a bipartite graph $G = G^n_{k,k+1}$ as follows: set $X = \binom{[n]}{k}$ and $Y = \binom{[n]}{k+1}$, and put $(A, B) \in E$ if $A \subseteq B$. There there is a matching in G that saturates X.*

Proof. We just need to show that for each k-uniform set system $\mathcal{F} \subseteq \binom{[n]}{k}$, the system $\mathcal{F}'$ defined by

$$\mathcal{F}' = \{B \in \binom{[n]}{k+1} : B \supseteq A \text{ for some } A \in \mathcal{F}\}$$

satisfies $|\mathcal{F}'| \geq |\mathcal{F}|$; then the result is immediate from Theorem 52.3.

To obtain the required inequality, note that the number of edges in G is $|\mathcal{F}|(n-k)$ (since for each $A \in \mathcal{F}$ there are exactly $n - k$ sets at level $k + 1$ that contain A), and it is also at most $|\mathcal{F}'|(k + 1)$ (since for each $B \in \mathcal{F}'$ there are exactly $k + 1$ sets at level k that are contained in B, but not all of these need be in $\mathcal{F}$). It follows that

$$|\mathcal{F}'| \geq \left(\frac{n-k}{k+1}\right)|\mathcal{F}| \geq |\mathcal{F}|,$$

the last inequality following from the bound on k. □

There exist other proofs of Proposition 52.4, some of which give an explicit construction of the matching[21].

53. Antichains and Sperner's theorem

We now ask a different type of extremal question. How many subsets can we select from a set of size $[n]$, if we are not allowed any containments between the subsets?

Definition 53.1. *A* chain *in the Boolean cube $\mathcal{B}_n$ is a sequence $(A_1, \ldots, A_k)$ of distinct sets, all subsets of $[n]$, with*

$$A_1 \subseteq A_2 \subseteq \ldots \subseteq A_k.$$

The chain is maximal *if there is no set A distinct from the A_i's such that $\{A_1, \ldots, A_k, A\}$ (in some order) forms a chain.*

An antichain *is a set $\{A_1, \ldots, A_k\}$ of subsets of $[n]$ with neither $A_i \subseteq A_j$ nor $A_j \subseteq A_i$ for any $1 \leq i < j \leq k$. The antichain is maximal if there is no set A distinct from the A_i's such that $\{A_1, \ldots, A_k, A\}$ forms an antichain.*

[21]For such constructions, see: Martin Aigner, Lexicographic Matching in Boolean Algebras, *Journal of Combinatorial Theory B* **14** (1973), 187–194.

Since a chain can have at most one set of each possible size (for 0 to n), it is easy to check that every chain can be extended to a chain with $n+1$ elements, one of each size, and there can be no larger chains. So for chains, the notions of maximum and maximal coincide. For antichains the situation is different; for example, each of $\binom{[n]}{k}$, $k = 0, \ldots, n$, forms a maximal antichain, so many different maximal sizes occur.

Among these $n+1$ examples of maximal antichains, one of them is as least as large as all the others: the middle layer of the cube, $\binom{[n]}{\lfloor n/2 \rfloor}$ (here we use the notation $\lfloor x \rfloor$ for the largest integer that does not exceed x, rather than the usual $[x]$, to avoid confusion with the notation $[n]$ for $\{1, \ldots, n\}$). If n is even, this is the unique layer of maximal size, while if n is odd it is one of two, the other being $\binom{[n]}{\lceil n/2 \rceil}$ (here we use $\lceil x \rceil$ for the smallest integer that is not smaller than x). It is tempting to think that it might be possible to get a larger antichain by taking "most" of the middle layer and augmenting with some sets from the layers immediately above and/or below the middle layer, or indeed by some other rather different construction, but this turns out not to be the case, as was first proved by Sperner in 1928.

Theorem 53.2. *Let $\mathcal{A}$ be an antichain in $\mathcal{B}_n$. Then $|\mathcal{A}| \leq \binom{n}{\lfloor n/2 \rfloor}$.*

We present two proofs, both using chains. The first is more prosaic; the second is one of the gems of combinatorics, and gives a stronger result.

Proof. (Theorem 53.2) We will be done if we can decompose $\mathcal{B}_n$ into $\binom{n}{\lfloor n/2 \rfloor}$ sets, each of which forms a chain, because no antichain can intersect a chain in more than one element. Proposition 52.4 fairly quickly gives such a decomposition. If n is even, fix, for each $0 \leq k \leq n/2 - 1$, a matching from $\binom{[n]}{k}$ to $\binom{[n]}{k+1}$ (in the containment graph defined in Proposition 52.4) that saturates $\binom{[n]}{k}$. Also fix, for each $n/2 + 1 \leq k \leq n$, a matching from $\binom{[n]}{k}$ to $\binom{[n]}{k-1}$ that saturates $\binom{[n]}{k}$ (by symmetry such matchings exist). These matchings decompose $\mathcal{B}_n$ into chains indexed by elements of $\binom{[n]}{n/2}$, so $\binom{n}{n/2}$ chains in all, as required. For odd n we do the same, except now we match up to level $\lfloor n/2 \rfloor$, down to level $\lceil n/2 \rceil$, and then add a matching between the two middle levels. □

Here is a second proof, due to Lubell, a genuine "proof from the book".

Proof. (Theorem 53.2) The Boolean cube has $n!$ maximal chains, indexed naturally by the $n!$ permutations of $[n]$ (the permutation σ of $[n]$ written in one-line notation as $\sigma(1) \ldots \sigma(n)$ corresponds to chain

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

Fix an antichain $\mathcal{F}$, and let $\mathcal{P}$ be the set of pairs $(A, \mathcal{M})$ with $A \in \mathcal{F}$, $\mathcal{M}$ a maximal chain, and $A \in \mathcal{M}$. Each $A \in \mathcal{F}$ appears in exactly $|A|!(n - |A|)!$ of the maximal chains, so

$$|\mathcal{P}| = \sum_{A \in \mathcal{F}} |A|!(n - |A|)! = \sum_{A \in \mathcal{F}} \frac{n!}{\binom{n}{|A|}}.$$

On the other hand, for each maximal chain $\mathcal{M}$ there is at most one $A \in \mathcal{F}$ with $A \in \mathcal{M}$, so

$$|\mathcal{P}| \leq n!.$$

It follows that $\sum_{A\in\mathcal{F}} \frac{n!}{\binom{n}{|A|}} = |\mathcal{P}| \leq n!$. Dividing this by $n!$, we obtain

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

On the other hand, since $\binom{n}{|A|} \leq \binom{n}{\lfloor n/2\rfloor}$, we have

$$\sum_{A\in\mathcal{F}} \frac{1}{\binom{n}{|A|}} \geq \frac{|\mathcal{F}|}{\binom{n}{\lfloor n/2\rfloor}}.$$

Combining these last two displayed equations gives the result. □

The chain decomposition given by the first proof above may be quite arbitrary in its appearance. For a subsequent application, it is useful to have a quite ordered chain decomposition of $\mathcal{B}_n$.

Definition 53.3. *A chain* $\{A_1, \ldots, A_\ell\}$ *is* symmetric *if* $\{|A_1|, \ldots, |A_\ell|\} = \{k, k+1, \ldots, n-k-1, n-k\}$ *for some* k, $0 \leq k \leq \lfloor n/2\rfloor$. *A* symmetric chain decomposition *of* $\mathcal{B}_n$ *is a decomposition into symmetric chains.*

Lemma 53.4. *For every* n, $\mathcal{B}_n$ *has a symmetric chain decomposition with* $\binom{n}{\lfloor n/2\rfloor}$ *non-empty chains.*

Proof. We proceed by induction on n, with (for example) $n = 1, 2$ trivial. For $n \geq 2$, let $\mathcal{B}_n = \bigcup_{i=1}^{\binom{n}{\lfloor n/2\rfloor}} \mathcal{C}_i$ be a symmetric chain decomposition with $\binom{n}{\lfloor n/2\rfloor}$ chains. For each chain $\mathcal{C}_i = \{A_1, \ldots, A_\ell\}$, with A_ℓ being the largest element in the chain, consider the two chains $\mathcal{C}_i' = \{A_1 \cup \{n+1\}, \ldots, A_{\ell-1} \cup \{n+1\}\}$ and $\mathcal{C}_i'' = \{A_1, \ldots, A_\ell, A_\ell \cup \{n+1\}\}$. These new chains are symmetric in $\mathcal{B}_{n+1}$, and moreover the collection of chains $\{\mathcal{C}_i', \mathcal{C}_i'' : i = 1, \ldots, \binom{n}{\lfloor n/2\rfloor}\}$ is pairwise disjoint and covers $\mathcal{B}_{n+1}$, so (the collection of non-empty chains in this set of chains) forms a symmetric chain decomposition.

When n is odd, say $n = 2m+1$, the symmetric chain decomposition of $\mathcal{B}_n$ has $\binom{2m+1}{m}$ chains, none of which have length 1, so that the number of non-empty chains in the symmetric chain decomposition of $\mathcal{B}_{n+1}$ is

$$2\binom{2m+1}{m} = \binom{2m+2}{m+1} = \binom{n+1}{\lfloor (n+1)/2\rfloor},$$

as required.

When n is even, say $n = 2m$, the symmetric chain decomposition of $\mathcal{B}_n$ has $\binom{2m}{m}$ chains, with $\binom{2m}{m} - \binom{2m}{m-1}$ of these having length 1 (because the chains that do not have length 1 are in one-to-one correspondence with the $m-1$-element subsets of $[n]$, of which there are $\binom{2m}{m-1}$). Each of these length 1 chains will contribute one empty chain in the construction, so that the number of non-empty chains in the symmetric chain decomposition of $\mathcal{B}_{n+1}$ is

$$2\binom{2m}{m} - \left(\binom{2m}{m} - \binom{2m}{m-1}\right) = \binom{2m}{m} + \binom{2m}{m-1} = \binom{2m+1}{m} = \binom{n+1}{\lfloor (n+1)/2\rfloor},$$

as required. □

54. The Littlewood-Offord problem

While studying zeros of random polynomials, Littlewood and Offord encountered the following problem: given a collection $\{z_1, \ldots, z_n\}$, all of modulus at least 1, determine the maximum number of the 2^n sums $\sum_{i=1}^n \varepsilon_i z_i$ (with each $\varepsilon_i \in \{+1, -1\}$) that can lie in the open disc of radius one around the origin. (A note on the requirement that the z_i's have absolute value at least one: some kind of lower bound on $|z_i|$ is needed, else the problem is trivial).

The question has an interpretation in terms of random walks: suppose we take an n step random walk in the complex plane, as follows: at the ith step, we toss a fair coin, and if it comes up heads we move from our current position z to $z + z_i$, while if it comes up tails we move to $z - z_i$. The probability the final position after n steps is within 1 of the origin is exactly the number of sums of the kind described above, scaled by 2^n.

Littlewood and Offord proved that there is a constant C such that the number of small sums is never more than $C2^n \log n/\sqrt{n}$ (and so the probability of being within 1 of the origin is no larger than order $(\log n)/\sqrt{n}$). This is close to best possible; if n is even and all the z_i's are equal, then the only way to end up within unit distance of the origin is by getting an equal number of heads and tails; so no general upper bound better than $\binom{n}{n/2} \approx 2^n/\sqrt{n}$ can be found.

Erdős improved the Littlewood-Offord bound to the best possible $\binom{n}{\lfloor n/2 \rfloor}$ in the case when all the z_i's are real, and later Kleitman and Katona extended this to the general (complex) case. Erdős' proof is a direct corollary of Theorem 53.2, but the Kleitman-Katona proof requires some further development. We present the real case first.

Theorem 54.1. *Let $x_1, \ldots, x_n$ be a collection on n real numbers, with $|x_i| \geq 1$ for each i. There are at most $\binom{n}{\lfloor n/2 \rfloor}$ vectors $(\varepsilon_i : i = 1, \ldots, n) \in \{+1, -1\}^n$ with $|\sum_{i=1}^n \varepsilon_i x_i| < 1$.*

Proof. We may assume that each x_i is non-negative (this is clear from the random walk interpretation of the problem: if there is a negative x_i, then we get a completely equivalent problem by replacing the instruction to step to $x + x_i$ if the ith coin toss is heads, and to $x - x_i$ if the ith toss is tails, by the instruction to step to $x + (-x_i)$ if the ith coin toss is heads, and to $x - (-x_i)$ if the ith toss is tails).

Now for each $\bar{\varepsilon} = (\varepsilon_i : i = 1, \ldots, n) \in \{+1, -1\}^n$ with $|\sum_{i=1}^n \varepsilon_i x_i| < 1$, associate the set $A_{\bar{\varepsilon}} = \{i : \varepsilon_i = +1\}$, and let $\mathcal{A}$ be the collection of all such $A_{\bar{\varepsilon}}$'s. We claim that $\mathcal{A}$ is an antichain on ground set $[n]$. Indeed, if $A, B \in \mathcal{A}$ satisfy $A \subseteq B$ (with $A \neq B$) then the difference between the corresponding sums is $2\sum_{i \in B \setminus A} x_i$, which is at least 2; but the difference between two numbers with absolute values less than 1 is strictly less than 2, a contradiction. It follows from Theorem 53.2 that $|\mathcal{A}|$, and so the number of small sums, is at most $\binom{n}{\lfloor n/2 \rfloor}$. □

To generalize this result to the complex plane, we need something that is in some sense a "two-dimensional" version of Sperner's Theorem. The right formulation turns out to be in terms of a partition of the ground set.

Definition 54.2. *Let $\mathcal{P} = X_1 \cup X_2 \cup \ldots \cup X_\ell$ be a partition of $[n]$ into non-empty parts. Say that a set system $\mathcal{A}$ on ground set $[n]$ is* Sperner relative to $\mathcal{P}$ *if, whenever $A_1, A_2 \in \mathcal{A}$ satisfy $A_1 \subseteq A_2$ with $A_1 \neq A_2$, then $A_2 \setminus A_1$ is not contained in any of the X_i.*

Notice that being Sperner relative to the trivial partition $\mathcal{P} = [n]$ is the same as being an antichain. On the other hand, while all antichains are Sperner relative to all non-trivial

partitions, for every such partition there are set systems which are Sperner relative to the partition, but are not antichains. It turns out, however, that at least in the case $\ell = 2$ this extra latitude does not increase the size of a maximum-sized family.

Theorem 54.3. *Let $\mathcal{P} = X_1 \cup X_2$ be a partition of $[n]$ into non-empty parts. If a set system $\mathcal{A}$ on ground set $[n]$ is Sperner relative to $\mathcal{P}$, then $|\mathcal{A}| \leq \binom{n}{\lfloor n/2 \rfloor}$.*

Proof. Form symmetric chain decompositions of both the power set of X_1 and the power set of X_2. Given $\mathcal{E} = (E_k, E_{k+1}, \ldots, E_{|X_1|-k})$, a chain in the decomposition of the power set of X_1, and $\mathcal{F} = (F_{k'}, F_{k'+1}, \ldots, F_{|X_2|-k'})$, a chain in the decomposition of the power set of X_2, denote by $\mathcal{EF}$ the set of sets of the form $E_i \cup F_j$ with $k \leq i \leq |X_1| - k$, $k' \leq j \leq |X_2| - k'$. Each element of the power set of $[n]$ can be expressed (uniquely) as a union of a subset of X_1 and a subset of X_2, and so the sets of the form $\mathcal{EF}$ form a partition of the power set of $[n]$.

Thinking of $\mathcal{EF}$ as a rectangular array whose ij entry is $E_i \cup F_j$ ($k \leq i \leq |X_1| - k$, $k' \leq j \leq |X_2| - k'$), it is evident that each row of the array can have at most one element of $\mathcal{A}$, and similarly each column can have at most one element of $\mathcal{A}$, and so the number of elements of $\mathcal{A}$ that can appear in the entire array is the minimum of the number of rows of the array, and the number of columns, that is, the minimum of the length of the chain $\mathcal{E}$ and the length of the chain $\mathcal{F}$. But (and here is where we use that the chains are symmetric) this minimum is exactly the number of entries in the array that have cardinality $\lfloor n/2 \rfloor$. It follows that $|\mathcal{A}| \leq \binom{n}{\lfloor n/2 \rfloor}$, as required. □

We can now prove Kleitman and Katona's generalization of Theorem 54.1:

Theorem 54.4. *Let $z_1, \ldots, z_n$ be a collection of n complex numbers, with $|z_i| \geq 1$ for each i. There are at most $\binom{n}{\lfloor n/2 \rfloor}$ vectors $(\varepsilon_i : i = 1, \ldots, n) \in \{+1, -1\}^n$ with $|\sum_{i=1}^n \varepsilon_i z_i| < 1$.*

Proof. As in the real case, we may assume that each z_i has non-negative real part. Let Z_1 be the set of indices i such that z_i is in the first quadrant (imaginary part at least 0), and let Z_2 be the set of indices i such that z_i is in the fourth quadrant (imaginary part negative); note that $[n] = Z_1 \cup Z_2$ is a decomposition of Z.

Now for each $\bar{\varepsilon} = (\varepsilon_i : i = 1, \ldots, n) \in \{+1, -1\}^n$ with $|\sum_{i=1}^n \varepsilon_i z_i| < 1$, associate the set $A_{\bar{\varepsilon}} = \{i : \varepsilon_i = +1\}$, and let $\mathcal{A}$ be the collection of all such $A_{\bar{\varepsilon}}$'s. We claim that $\mathcal{A}$ is Sperner relative to the partition $[n] = Z_1 \cup Z_2$ (or to the partition $[n] = [n]$, if one of Z_1 and Z_2 is empty). Indeed, suppose that $A, B \in \mathcal{A}$ satisfy $A \subseteq B$ (with $A \neq B$), and $B \setminus A \subseteq Z_1$ (or Z_2). Write A as $A = A_{\bar{\varepsilon}}$ for some $\bar{\varepsilon} = (\varepsilon_i : i = 1, \ldots, n)$, and write B as $B = A_{\bar{\delta}}$ for some $\bar{\delta} = (\delta_i : i = 1, \ldots, n)$. Then the difference $\sum_{i=1}^n \delta_i z_i - \sum_{i=1}^n \varepsilon_i z_i$ between the corresponding sums is $2\sum_{i \in B \setminus A} z_i$. This is a sum of complex numbers, all in the first (or fourth) quadrant, and all of length at least 2, and so the two original sums cannot both lie inside the unit circle. It follows from Theorem 54.3 that $|\mathcal{A}|$, and so the number of small sums, is at most $\binom{n}{\lfloor n/2 \rfloor}$. □

We have proved an extension of Theorem 53.2 to set systems which are Sperner relative to a partition of $[n]$ into two non-empty parts. No such clean extension exists to set systems which are Sperner relative to a partition of $[n]$ into three or more non-empty parts. Here's an example: the set system $\{\{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$ on ground set [3] is Sperner relative to the partition $[3] = \{1\} \cup \{2\} \cup \{3\}$, but has size 4, which is greater than $\binom{3}{\lfloor 3/2 \rfloor}$. In fact, it is known[22] that the largest size of a set system on ground set $[n]$ that is Sperner relative

[22] a result of Griggs, Odlyzko and Shearer, cited in Anderson's book "Combinatorics of finite sets"

to some partition of $[n]$ into k non-empty parts is asymptotic to

$$\left(\frac{k}{2\log k}\right)^{1/2}\frac{2^n}{\sqrt{n}}.$$

Despite the absence of such a generalization of Theorem 53.2, Kleitman was able to extend Theorems 54.1 and 54.4 to arbitrary dimensions. We omit the proof of the following result[23].

Theorem 54.5. *Let $x_1,\ldots,x_n$ be a collection of n vectors in $\mathbb{R}^n$, with $|x_i|\geq 1$ for each i. There are at most $\binom{n}{\lfloor n/2\rfloor}$ vectors $(\varepsilon_i : i=1,\ldots,n)\in\{+1,-1\}^n$ with $|\sum_{i=1}^n \varepsilon_i x_i|<1$.*

55. Graphs

A *graph* G is a pair (V,E) where V is a (finite, non-empty) set of *vertices* (singular: vertex), and E is a (possibly empty) set of *edges*, that is, unordered pairs of vertices. A graph may be represented graphically by a set of points in the plane labelled with the names of the vertices, with an arc or curve joining two vertices if those two vertices comprise an edge, but the precise location of the points and nature of the arcs is immaterial: all that determines a graph is the names of the vertices and the set of pairs that form edges.

The edge set of a graph can be thought of as a 2-uniform set system on ground set V. We may also think of it as a *point* in the $\binom{n}{2}$-dimensional Boolean cube on ground set $\binom{V}{2}$ (the set of two-element subsets of V); this latter point of view is useful later when thinking about families of graphs.

We will mostly use n for the number of vertices, and typically take $V=\{1,\ldots,n\}$ or $\{v_1,\ldots,v_n\}$, and we will use m for the number of edges. Rather than writing edges as sets, for notational simplicity will write them as juxtaposed pairs of vertices: that is, we will tend to write "$ij\in E$" rather than "$\{i,j\}\in E$", with the convention that ij and ji are interchangeable. If $e=ij\in E$ we say that

- i and j are *adjacent*;
- i and j are *joined by an edge*, or *joined by* e;
- i and j are *neighbors* (and i is a *neighbor* of j);
- i and j are *endvertices of* e.

The *neighborhood* of a vertex i, denoted $N(i)$, is the set of j such that j is a neighbor of i. The *degree* of a vertex i, denoted $d(i)$ (or $d_G(i)$ if we want to make clear which graph is under discussion), is $|N(i)|$.

If $i\in V$, then the graph $G-i$, the graph obtained from G *by removing* i, is the graph on vertex set $V\setminus\{i\}$ whose edge set consists of all edges of G which do not have i as an endvertex (in other words, $G-i$ is obtained from G by deleting i and all edges involving i). If $e=ij\in E$, then the graph $G-e=G-ij$, the graph obtained from G *by removing* e (or ij), is the graph on vertex set V with edge set $E\setminus\{e\}=E\setminus\{ij\}$.

A *subgraph* of a graph G is a graph obtained from G by the process of removing some edges and/or some vertices; in other words, $H=(V',E')$ is a subgraph of the graph $G=(V,E)$ if H is a graph and if $V'\subseteq V$ and $E'\subseteq E$; we write $H\leq G$ and say that G *contains* H.

Two graphs $G=(V,E)$ and $G'=(V',E')$ are *isomorphic* if they can be made identical by a relabelling, that is, if there is a bijection $f:V\to V'$ satisfying $xy\in E(G)$ if and only if $f(x)f(y)\in E(G')$. Sometimes when we talk about "a graph", what we will actually mean is "an isomorphism class of graphs". For example, if we say "a triangle is a graph on

[23]It can be found in Anderson's book "Combinatorics of finite sets" (page 183)

three vertices with all pairs of vertices joined by an edge", we are in fact referring to the isomorphism class of graphs on three vertices x, y, z with edge set $\{xy, xz, yz\}$, as $\{x, y, z\}$ varies over all sets of size 3; and when we say that "G contains a triangle" we will mean that G has a subgraph H that belongs to this isomorphism class. This informality should not lead to confusion, as it should always be clear from the context what is being considered.

Here are some common isomorphism classes of graphs that we come up in subsequent sections. In each case we will just give one representative example of the class.

- The *complete graph* on n vertices is the graph K_n on vertex set $\{1, \dots, n\}$ with edge set $\binom{[n]}{2}$ (all possible edges present).
- The *empty graph* on n vertices is the graph E_n on vertex set $\{1, \dots, n\}$ with empty edge set.
- The *complete bipartite graph with parameters a, b* is the graph $K_{a,b}$ on vertex set $\{x_1, \dots, x_a\} \cup \{y_1, \dots, y_b\}$ with edge set consisting of all pairs $x_i y_j$, $1 \leq i \leq a$, $1 \leq j \leq b$. The sets $\{x_1, \dots, x_a\}$ and $\{y_1, \dots, y_b\}$ are referred to as the *partition classes* of $K_{a,b}$.
- The *path graph* on n vertices is the graph P_n on vertex set $\{1, \dots, n\}$ with edge set $\{12, 23, \dots, (n-1)n\}$.
- The *cycle graph* on n vertices is the graph C_n on vertex set $\{1, \dots, n\}$ with edge set $\{12, 23, \dots, (n-1)n, n1\}$.
- The *star graph* on n vertices is the graph S_n on vertex set $\{1, \dots, n\}$ with edge set $\{12, 13, \dots, 1n\}$.

Notice that there is some overlap: for example, the star graph S_n is the same as the complete bipartite graph $K_{1,n-1}$, and also that K_3 and C_3 are the same. This latter class of graphs is referred to as the *triangle*.

56. Mantel's theorem and Turán's theorem

In this section we begin a consideration of the follow fundamental extremal question:

Question 56.1. *Fix a graph F. At most how many edges can a graph on n vertices have, if it does not contain F as a subgraph?*

It will be helpful to fix some notation. We set

$$\mathrm{ex}(n, F) = \max\{m : \text{there is } n\text{-vertex, } m\text{-edge graph that does not contain } F\}$$

and

$$\mathrm{Ex}(n, F) = \{G : G \text{ has } n \text{ vertices, } \mathrm{ex}(n, F) \text{ edges, and does not contain } F\}.$$

When F has 1 or 2 vertices the question is very easy, as it is when F has three vertices and is not the triangle. But already when $F = K_3$ the question is non-trivial.

One way to avoid having a triangle in a graph is to decompose the vertex set into two classes, and only put edges between the two classes; to maximize the number of edges in this construction, we should put all possible edges crossing between the two classes, an we should make the two classes be of nearly equal size. The resulting graph is $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$, which is certainly a *maximal* example of a graph without a triangle. The following theorem, due to Mantel, says that this construction is also *maximum*, and is in fact the unique maximum.

Theorem 56.2. *For all $n \geq 2$, we have* $\mathrm{Ex}(n, K_3) = \{K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}\}$ *and*

$$\mathrm{ex}(n, K_3) = \lfloor n/2 \rfloor \lceil n/2 \rceil = \begin{cases} \frac{n^2}{4} & \text{if } n \text{ is even} \\ \frac{n^2-1}{4} & \text{if } n \text{ is odd} \end{cases}$$

Proof. Let $G = (V, E)$ be a graph on n vertices without a triangle. For each $e = xy \in E$ we have that the two subsets $N(x) \setminus \{y\}$ and $N(y) \setminus \{x\}$ of $V \setminus \{x, y\}$ are disjoint (since otherwise G would have a triangle), so $(d(x) - 1) + (d(y) - 1) \leq n - 2$, or, equivalently, $d(x) + d(y) \leq n$. Summing this inequality over all edges, we get

$$\sum_{e=xy \in E} (d(x) + d(y)) \leq mn, \tag{53}$$

where $m = |E|$. Now, for each $x \in V$, the term $d(x)$ occurs exactly $d(x)$ times on the left-hand side of (53). Hence, the left-hand side of (53) rewrites as follows:

$$\sum_{e=xy \in E} (d(x) + d(y)) = \sum_{x \in V} d(x)d(x) = \sum_{x \in V} d^2(x),$$

where $d^2(x)$ is short for $(d(x))^2$. Hence, (53) becomes

$$\sum_{x \in V} d^2(x) \leq mn. \tag{54}$$

Now we apply the Cauchy-Schwarz inequality (also known as Cauchy-Schwarz-Bunyakovsky inequality), which says that for reals $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$,

$$\left(\sum_{i=1}^{n} x_i y_i \right)^2 \leq \left(\sum_{i=1}^{n} x_i^2 \right) \left(\sum_{i=1}^{n} y_i^2 \right).$$

(Here's a quick proof: the quadratic polynomial in z, $\sum_{i=1}^{n} (x_i z + y_i)^2$, is non-negative for all real z, so must have non-positive discriminant. An alternative expression for the polynomial is

$$\left(\sum_{i=1}^{n} x_i^2 \right) z^2 + 2 \left(\sum_{i=1}^{n} x_i y_i \right) z + \left(\sum_{i=1}^{n} y_i^2 \right).$$

The required inequality is now exactly the statement that the discriminant is non-positive.) We apply with the x_i's being the degrees of the vertices of G and the y_i's all being 1; this yields

$$\left(\sum_{x \in V} d(x) \right)^2 \leq n \sum_{x \in V} d^2(x),$$

so, since $\sum_{x \in V} d(x) = 2m$,

$$\sum_{x \in V} d^2(x) \geq \frac{4m^2}{n}. \tag{55}$$

Combining (54) and (55) yields $m \leq n^2/4$, which can be strengthened to $m \leq (n^2 - 1)/4$ when n is odd, since m is an integer. Since the right-hand side of these two inequalities is the number of edges in $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$, we have established everything claimed in the theorem except the uniqueness of $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ as an extremal example.

For uniqueness, suppose that triangle-free G with the maximum possible number of edges has an edge xy with $d(x) + d(y) < n$. The graph $G - \{x, y\}$ is triangle-free with $n - 2$ vertices,

and so has at most $\lfloor (n-2)/2 \rfloor \lceil (n-2)/2 \rceil$ edges (by the part of Mantel that we have already proven[24]). Restoring x and y adds back $(d(x)-1)+(d(y)-1)+1 < n-1$ edges, so that G has fewer than $\lfloor (n-2)/2 \rfloor \lceil (n-2)/2 \rceil + n - 1$ edges. In other words, G has fewer than $\lfloor n/2 \rfloor \lceil n/2 \rceil$ edges (since $\lfloor (n-2)/2 \rfloor \lceil (n-2)/2 \rceil + n - 1 = \lfloor n/2 \rfloor \lceil n/2 \rceil$), a contradiction.

We may assume, then, that G satisfies $d(x)+d(y) = n$ for all $xy \in E$. Fix $xy \in E$ and set $X = N(x)$ and $Y = N(y)$; note that $X \cup Y = V$, that X, Y are disjoint, and that there are no edges joining pairs of vertices in X or pairs of vertices in Y (all this by triangle-freeness). By maximality of G we may assume that G is the complete bipartite graph with classes X and Y, and so has $|X||Y| = k(n-k)$ edges for some $k \in \{1, \ldots, n\}$ with (without loss of generality) $k \leq n/2$. This quantity is maximized uniquely when $k = \lfloor n/2 \rfloor$, so that indeed $G = K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$. □

We now ask a more general question:

Question 56.3. *Fix $r \geq 3$. What are $\mathrm{ex}(n, K_r)$ and $\mathrm{Ex}(n, K_r)$?*

There is a natural candidate. We take $V = [n]$ for this discussion. Decompose V into $r-1$ blocks $A_1, \ldots, A_{r-1}$ (not necessarily non-empty), and put an edge between i and j exactly when i and j are in different blocks. The resulting graph does not contain K_r, and has

$$\binom{n}{2} - \sum_{i=1}^{r-1} \binom{|A_i|}{2}$$

edges. To maximize this quantity, we need the following simple observation, whose quick proof is left as an exercise (if follows from the convexity of $x^2 - x$). Here for real x we write $\binom{x}{2}$ for $x(x-1)/2$.

Claim 56.4. *For $x < y$*

$$2\binom{(x+y)/2}{2} < \binom{x}{2} + \binom{y}{2}.$$

A corollary of this claim is that if our decomposition has two blocks whose sizes differ by 2 or more, then by moving a vertex from the smaller block to the larger block we can increase the number of edges in the graph. It follows that we get the greatest number of edges in a construction of this kind by taking a decomposition in which the $|A_i|$'s are as near equal in size as possible. If we order the blocks by size from smallest to largest this decomposition satisfies $|A_1| \leq |A_2| \leq \ldots \leq |A_{r-1}| \leq |A_1| + 1$, and, subject to this condition, the vector $(|A_i| : i = 1, \ldots, r-1)$ is unique.

The preceding discussion motivates the following definition.

Definition 56.5. *Fix $n, r \geq 1$. The* Turán graph *$T_r(n)$ is the graph on vertex set $[n]$ constructed as follows: where $A_1 \cup \ldots \cup A_r$ is a decomposition of $[n]$ (with the blocks not necessarily non-empty) with $|A_1| \leq |A_2| \leq \ldots \leq |A_r| \leq |A_1| + 1$, there is an edge from i to j exactly when i and j are in different blocks. The number of edges in the Turán graph is denoted by $t_r(n)$.*

The following result, due to Turán, generalizes Mantel's theorem.

Theorem 56.6. *For all $n \geq 1$ and $r \geq 2$, $\mathrm{ex}(n, K_r) = t_{r-1}(n)$ and $\mathrm{Ex}(n, K_r) = \{T_{r-1}(n)\}$.*

[24]Strictly speaking, we need $n - 2 \geq 2$ in order to apply Mantel's theorem. But the cases where $n - 2 < 2$ are easily checked by the reader.

In other words, the Turán graph on n vertices with $r-1$ parts is the unique graph on n vertices with the largest number of edges that does not contain r mutually adjacent vertices.

Before proving Theorem 56.6 we establish a weaker, but still very useful, version of it, which is easier to prove. We begin with an estimate on $t_r(n)$. Using Jensen's inequality for the convex function $\binom{x}{2}$ (see Claim 56.4 for the convexity), we have

$$\begin{aligned} t_r(n) &= \binom{n}{2} - \sum_{i=1}^{r} \binom{|A_i|}{2} \\ &\leq \binom{n}{2} - r\binom{n/r}{2} \\ &= \frac{n^2}{2}\left(1 - \frac{1}{r}\right). \end{aligned}$$

The following result is a slight weakening of Theorem 56.6 as it only gives a tight upper bound on $\mathrm{ex}(n, K_r)$ in the case when $r-1$ divides n, and does not address equality in that case. Nonetheless it gives a substantial result, especially for large n. Noting that the greatest number of edges in a graph on n vertices is $\binom{n}{2} \sim n^2/2$ (as $n \to \infty$), it says that for large n if a graph has more than a proportion $1-(1/(r-1))$ of potential edges present then it must contain K_r.

Theorem 56.7. *For all $n \geq 1$ and $r \geq 2$,*

$$\mathrm{ex}(n, K_r) \leq \frac{n^2}{2}\left(1 - \frac{1}{r-1}\right).$$

Proof. The case $r = 2$ is trivial, so we assume $r \geq 3$. For each fixed r we proceed by induction on n. For $n < r$ the result is implied by

$$\binom{n}{2} \leq \frac{n^2}{2}\left(1 - \frac{1}{r-1}\right)$$

which is indeed true (it reduces to $n \leq r-1$). For $n = r$ the result is implied by

$$\binom{r}{2} - 1 \leq \frac{r^2}{2}\left(1 - \frac{1}{r-1}\right),$$

which is true for all $r \geq 2$. So now fix $n > r$, and let G on vertex set $[n]$ have the greatest number of edges among n-vertex graphs that do not contain K_r. Since in a graph that does not contain K_{r-1} we cannot create a K_r by adding a single edge, it must be that G contains K_{r-1}. Let A be a set of $r-1$ vertices in G that are mutually adjacent, and let $B = [n] \setminus A$. The number of edges completely within A is $\binom{r-1}{2}$. No vertex in B can be adjacent to everything in A (or else G would contain K_r) so the number of edges that go between A and B is at most $|B|(|A|-1) = (n-(r-1))(r-2)$. Since there is no K_r contained completely within B, by induction (applied to the subgraph on vertex set B) there are at most

$$\frac{(n-(r-1))^2}{2}\left(1 - \frac{1}{r-1}\right)$$

edges completely within B. It follows that

$$\begin{aligned} |E(G)| &\leq \binom{r-1}{2} + (n-(r-1))(r-2) + \frac{(n-(r-1))^2}{2}\left(1-\frac{1}{r-1}\right) \\ &= \frac{n^2}{2}\left(1-\frac{1}{r-1}\right). \end{aligned}$$

□

Now we turn to the proof of Theorem 56.6.

Proof. (Theorem 56.6) We proceed by induction on n, with the base cases $n \leq 3$ all easy. Fix $n \geq 4$. The case $r = 2$ is easy, so we assume $r \geq 3$.

Let G be a graph on n vertices, without K_r, that has the largest possible number of edges. Let v_1 be a vertex of maximum degree in G (this must be at least 1, since $r \geq 3$), and let $S = N(v_1)$ and set $T = V \setminus S$ (where V denotes the vertex set of G). Notice that there is no K_{r-1} inside S, since otherwise it would form a K_r together with v_1.

Form a graph H on the same vertex set as G as follows: within S, leave the edge structure unchanged; inside T, remove all edges; and between S and T put in all possible edges. Since there is no K_{r-1} within S, and no edges within T, H has no K_r. Also, notice that every vertex in H has degree at least as large as the corresponding vertex in G (for vertices in T this follows from the choice of v_1 as vertex with maximum degree; for vertices in S, this is clear from the construction). It follows that H has as least as many edges as G, so, by maximality of G, the graph H has the same number of edges as G.

In H there are $|S||T|$ edges that are not inside S, so in G there must have been this many edges not inside S, too. In other words, the number of edges in G not inside S is $|S||T|$.

Let $v_1, \ldots, v_k$ be the vertices of T, and for each i, $1 \leq i \leq k$, let e_i be the number of edges (in G) leaving v_i that go to another vertex of T, and d_i the number of edges (in G) leaving v_i that go to S. Since v_1 has maximum degree in G, we have $d_i + e_i \leq |S|$ for each i. Hence, $\sum_i (d_i + e_i) \leq |S||T|$.

Now, the number of edges in G not inside S is $\sum_i d_i + \frac{1}{2}\sum_i e_i$ (by the definition of d_i and e_i). But the same number is $|S||T|$ (as we have seen before). Hence,

$$|S||T| = \sum_i d_i + \frac{1}{2}\sum_i e_i = \sum_i (d_i + e_i) - \frac{1}{2}\sum_i e_i \leq |S||T| - \frac{1}{2}\sum_i e_i,$$

from which it follows that $e_i = 0$ and $d_i = |S|$ for all i, and that $G = H$.

Now focus on S. We have $1 \leq |S| \leq n-1$, and within S there is no K_{r-1}. Moreover, by maximality of G, the number of edges inside S is maximal subject to the condition that S has no K_{r-1} (since any graph with vertex set S having no K_{r-1} but having more edges than the restriction of G to S could be transplanted into G and create a graph on vertex set V having no K_r but more edges than G; this would contradict the maximality of G). By induction the restriction of G to S is a Turán graph with $r-2$ blocks, and so, since everything in T is adjacent to everything in S and there are no edges inside T, we conclude that G has the following form: its vertex set is decomposed into $r-1$ blocks $A_1, \ldots, A_{r-1}$ (not necessarily all non-empty), with an edge from i to j if and only if i and j are in different blocks. Since, as we have observed before, among all such graphs the unique one (up to isomorphism) with the greatest number of edges is $T_{r-1}(n)$, this is enough to establish $\mathrm{Ex}(n, K_r) = \{T_{r-1}(n)\}$ and $\mathrm{ex}(n, K_r) = t_{r-1}(n)$. □

57. The chromatic number of a graph

We have answered the extremal question, Question 56.1, whenever F is a complete graph, and the answer we have obtained is quite exact. We now consider more general F, where the answers will not be quite as exact. In general for non-complete F the best that we will be able to get is information about the proportion of the potential edges whose presence ensures that a graph contains F for all sufficiently large n. In other words, we will be examining the quantity

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n,F)}{\binom{n}{2}}.$$

Although a priori we have no reason to suppose that this limit exists (except in the case $F = K_r$, where Turán's Theorem 56.6 gives that the limit exists and equal $1 - 1/(r-1)$ for all $r \geq 2$), it turns out that the limit does exist for all F, and is easily expressible in terms of a natural parameter of F. This parameter is the *chromatic number*, which we now define.

Definition 57.1. *Fix an integer $q \geq 0$. A* proper q-coloring *of a graph $G = (V, E)$ is a function $f : V \to [q]$ (recall $[q] = \{1, \ldots, q\}$) with the property that if $xy \in E$ then $f(x) \neq f(y)$. We visualize a proper q-coloring f of G by pretending that the elements of $[q]$ are colors, and each vertex $v \in V$ is colored with the color $f(v)$. The* chromatic number *of G, denoted $\chi(G)$, is the minimum q for which a proper q-coloring of G exists. Equivalently, $\chi(G)$ is the least q for which there exists a decomposition of the vertex set $V = A_1 \cup \ldots \cup A_q$ such that all edges of G join vertices in different blocks of the decomposition.*

Notice that if G has n vertices, then it trivially has a proper n-coloring, so $\chi(G)$ always exists and is finite.

As an example, the chromatic number of K_r is r; the chromatic number of the cycle graph C_n is 2 if n is even and 3 if n is odd, and the chromatic number of the wheel graph W_n — a cycle C_{n-1} together with one more vertex, joined to all the vertices of the cycle — is 3 if n is odd and 4 if n is even.

The notion of chromatic number first came up recreationally, in the famous *four-color problem*, first posed in the 1850's. This problem asks whether it is always possible to color the regions of a map using only four colours, in such a way that no two regions that share a boundary line receive the same color. Translated into graph language, the problem asks whether every *planar* graph — a graph which can be drawn in the plane without any of the edges meeting, except possible at their endpoints — has chromatic number at most 4. This translation into graph language is obtained by representing the map as a graph with the regions being the vertices of the graph, and with an edge between two regions exactly if those two regions share a boundary line. The long history of the four color problem, involving false proofs, the creation of much of modern graph theory in the twentieth century, and the controversy surrounding the use of computers to assist in its resolution, is well documented, see for example the book by Wilson[25]; for a more mathematical discussion of the history, see for example the survey article by Thomas[26].

The chromatic number of a graph has many practical applications in scheduling and assignment. If, for example, the vertices of a graph are radio towers, and there is an edge between two towers if they are close enough that they would interfere with each other if they

[25]R. Wilson, Four Colors Suffice, Penguin Books, London, 2002.

[26]R. Thomas, An Update on the Four-Color Theorem, *Notices of the American Mathematical Society* **45** (issue 7) (1998), 848–859.

broadcasted at the same frequency, then the chromatic number of the graph is the smallest number of distinct frequencies needed so that one can be assigned to each tower in such a way that no two towers interfere with each other. If the vertices are committees, and there is an edge between two committees if they share a member, then the chromatic number is the smallest number of meeting times needed so that all committees can meet without someone being forced to be in two meetings at the same time. If the vertices are zoo animals, and there is an edge between two animals if they would attack each other when confined to a cage together, then the chromatic number is the smallest number of cages needed so that the animals can be caged harmoniously.

Given the wide range of applications, it would be nice if it was relatively straightforward to compute the chromatic number of a graph. Unfortunately, although is quite straightforward to answer the question "is $\chi(G) = 2$?", for all $k \geq 3$ there is no known algorithm that always (for all possible input graphs G) correctly answers the question "is $\chi(G) = k$?", and which runs in a time that is polynomial in the number of vertices of G. For example, the fastest known algorithm to answer the question "is $\chi(G) = 3$?" takes roughly 1.33^n steps, where n is the number of vertices of G.

Part of the difficulty of coloring is that although it seems like a "local" property (proper q-coloring only puts restrictions on colors given to neighboring vertices), it is actually quite "global" — a decision made about the color of a vertex in one part of a graph can have a significant influence on the choice of colors available at a vertex very far away. We digress a little now from the main theme (the greatest number of edges in a graph that avoids containing some smaller graph) to give one concrete result that demonstrates the "non-locality" of graph colouring. First we need a definition.

Definition 57.2. *Fix an integer $k \geq 0$. A* clique of size k *in a graph $G = (V, E)$ is a subset K of V of size k with the property that if $x, y \in K$ satisfy $x \neq y$, then $xy \in E$. The* clique number *of G, denoted $\omega(G)$, is the maximum k for which a clique of size k exists. Equivalently, $\omega(G)$ is the size of the largest set of mutually adjacent vertices in G.*

It is obvious that $\chi(G) \geq \omega(G)$, and while it is equally obvious that $\chi(G) \neq \omega(G)$ in general (C_5, with chromatic number 3 and clique number 2, is an example), it seems reasonable to think that the chromatic number cannot in general be much larger than the clique number; but this is not the case.

Theorem 57.3. *For each $k \geq 2$, there are graphs G with $\chi(G) = k$ and $\omega(G) = 2$.*

The proof is constructive. We will describe an operation on graphs which preserves the property of having clique number equal to 2, but also increases chromatic number. The construction is due to Mycielski.

Definition 57.4. *Let $G = (V, E)$ be any graph, with $V = \{v_1, \ldots, v_n\}$. The* Mycielskian *$M(G)$ of G is the graph on vertex set $V \cup W \cup \{z\}$ where $W = \{w_1, \ldots, w_n\}$ with the following edges: $v_i v_j$ for each i, j with $v_i v_j \in E(G)$; $v_i w_j$ for each i, j with $v_i v_j \in E(G)$; and $w_i z$ for each i.*

For example, $M(K_2) = C_5$. The graph $M(C_5)$, on 11 vertices, is known as the *Grötzsch graph.*

We will prove the following two lemmas concerning the Mycielskian, after which the proof of Theorem 57.3 is immediate: defining a sequence of graphs by $G_2 = K_2$ and, for $k \geq 3$, $G_k = M(G_{k-1})$, the graph G_k has chromatic number k and clique number 2; for example,

the Grötzsch graph has no triangles but has chromatic number 4 (and it turns out to be the unique graph with the least number of vertices satisfying this property).

Lemma 57.5. *If G is a graph without any triangles then $M(G)$ also has no triangles.*

Lemma 57.6. *If $\chi(G) = k$ then $\chi(M(G)) = k + 1$.*

Proof. (Lemma 57.5) Let $G = (V, E)$ be a triangle-free graph. By construction it is evident that in $M(G)$ there is no triangle involving z, none involving three or two vertices from W, and none involving three vertices from V. So if $M(G)$ has a triangle, its vertex set must be of the form $\{w_i, v_j, v_k\}$ with $j \neq k$ and (by construction) $i \neq j, k$. But if $\{w_i, v_j, v_k\}$ forms a triangle in $M(G)$ then, by construction, $\{v_i, v_j, v_k\}$ forms a triangle in G, a contradiction. □

Proof. (Lemma 57.6) Let $G = (V, E)$ satisfy $\chi(G) = k$, and let $f : V \to [k]$ be a proper k-coloring. Consider the function $f' : V \cup W \cup \{z\} \to [k+1]$ defined by $f'(v_i) = f'(w_i) = f(v_i)$ for each i, and $f'(z) = k+1$. This is readily verified to be a proper $(k+1)$-coloring of $M(G)$, so $\chi(M(G)) \leq k + 1$.

Suppose that $g : V \cup W \cup \{z\} \to [k]$ was a proper k-coloring of $M(G)$. We will construct from this a proper $(k-1)$-coloring of G, for a contradiction, that will establish that in fact $\chi(M(G)) = k + 1$. Assume without loss of generality that $g(z) = k$. There is no i with $g(w_i) = k$, but there may be some i with $g(v_i) = k$ (indeed, there *must* be such i, otherwise the restriction of f to V would furnish a proper $(k-1)$-coloring of G). Let K be the set of those i such that $g(v_i) = k$. Define a function $g' : V \to [k-1]$ as follows:

$$g'(v_i) = \begin{cases} g(w_i) & \text{if } i \in K \\ g(v_i) & \text{otherwise;} \end{cases}$$

note that this is indeed a function to $[k-1]$. We claim that it is a proper $(k-1)$-coloring of G, for which we need to show that there is no i, j with $v_i v_j \in E(G)$ satisfying $g'(v_i) = g'(v_j)$. To see this, first note that there are no pairs $i, j \in K$ with $v_i v_j \in E(G)$, or else g would not have been a proper coloring of $M(G)$. Next, note that if $i, j \notin K$ satisfy $v_i v_j \in E(G)$ then $g'(v_i) \neq g'(v_j)$ since on $\{v_i, v_j\}$ the function g' agrees with the proper coloring g. It remains to consider $i \in K$, $j \notin K$ with $v_i v_j \in E(G)$. We have $g'(v_i) = g(w_i)$ and $g'(v_j) = g(v_j)$; but since $w_i v_j \in E(M(G))$, $g(w_i) \neq g(v_j)$ and so $g'(v_i) \neq g'(v_j)$. This concludes the proof. □

We have constructed a graph G_k that has clique number 2 but chromatic number k; this graph has about 2^k vertices. Is such a large graph needed? For each $k \geq 2$, define $f(k)$ to be the least n such that there exists a graph on n vertices with clique number 2 and chromatic number k. Notice that $f(2) = 2$ and $f(3) = 5$; it can also be shown that $f(4) = 11$, so that for the first three values, the sequence $(G_k)_{k \geq 2}$ gives examples of graphs whose numbers of vertices achieve $f(k)$. It turns out, however, that for large k the number of vertices in G_k is far larger than $f(k)$; rather than needing exponentially many in k vertices to construct a triangle-free graph with chromatic number k, we need just around quadratically many.

We begin by establishing the necessity of roughly quadratically many vertices.

Claim 57.7. *There is a constant $C > 0$ such that if G is a triangle-free graph on n vertices with chromatic number k, then $n \geq Ck^2$.*

Part of the proof will involve the following simple upper bound on the chromatic number.

Claim 57.8. *If G is a graph with maximum degree Δ, then $\chi(G) \leq \Delta + 1$.*

Proof. Let the vertex set be $\{v_1, \ldots, v_n\}$. Assign colors to the vertices sequentially (from v_1 to v_n), from palette $\{1, \ldots, \Delta + 1\}$, according to the following rule: assign to v_i the least color that has not already been assigned to a neighbor v_j of v_i (with, of course, $j < i$). Since the maximum degree in G is Δ, there will be no v_i for which a color cannot be assigned, and so the final result of this process will be a proper $(\Delta + 1)$-coloring of G. □

The bound in Claim 57.8 is tight; consider, for example, odd cycles or complete graphs. It turns out that these are essentially the only examples of tightness. We won't prove the following, which is known as *Brooks' Theorem*.

Theorem 57.9. *If G is a connected graph with maximum degree Δ that is not a complete graph or an odd cycle, then $\chi(G) \leq \Delta$.*

The bound in Claim 57.8 can also be very far from the truth; consider for example the complete bipartite graph $K_{\Delta,\Delta}$ which has chromatic number 2 but maximum degree Δ.

The coloring described in the proof of Claim 57.8 is "greedy": it colors each new vertex using the best available color, without regard for what impact this might have on future decisions. Sometimes the greedy approach can yield a very good coloring, and sometimes it can give a coloring which uses far more than the optimal number of colors. Consider, for example, the graph on vertex set $\{x_1, \ldots, x_n, y_1, \ldots, y_n\}$ with no edges between x_i and x_j or between y_i and y_j for any $i \neq j$, and an edge between x_i and y_j if and only if $i \neq j$ (this is $K_{n,n}$ with a matching removed). It is easy to check that if the vertices are ordered $(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)$ then the greedy coloring from the proof of Claim 57.8 yields a coloring with 2 colors (the optimal number); but if the vertices are ordered $(x_1, y_1, x_2, y_2, \ldots, x_n, y_n)$ then we get a coloring with n colors.

Proof. (Claim 57.7) We assign colors to the vertices of G, in such a way that adjacent vertices do not receive the same color, in the following "greedy" manner. If at any time there is a vertex v in the graph with the property that it has at least $\lfloor\sqrt{n}\rfloor$ neighbors that have not yet been assigned a color, then choose a color that has not yet been used before, and assign it to each of the uncolored neighbors of v (notice that this does not create any edge both of whose endvertices receive the same color, since G has no triangles). This part of the process uses at most $n/\lfloor\sqrt{n}\rfloor$ colors. Once this process can no longer be continued, the remaining uncolored graph has maximum degree at most $\lfloor\sqrt{n}\rfloor$, and so can be colored (by Claim 57.8) using at most $1 + \lfloor\sqrt{n}\rfloor$ further colours. It follows that $k = \chi(G) \leq n/\lfloor\sqrt{n}\rfloor + \lfloor\sqrt{n}\rfloor + 1$. So $\chi(G) = k \leq O(\sqrt{n})$, $n \geq \Omega(k^2)$. □

Without building up a significant amount of background material we cannot prove that roughly quadratically many vertices are sufficient to force high chromatic number in a triangle-free graph, but we can fairly quickly prove the following.

Claim 57.10. *There is a constant $C > 0$ such that $f(k) \leq Ck^3 \log^3 k$ for all $k \geq 2$.*

The proof will use *Markov's inequality*: if X is a non-negative random variable with finite expectation $E(X) > 0$, then for all $a > 0$

$$\Pr(X \geq aE(X)) \leq \frac{1}{a}.$$

(To verify this, consider Y that takes value 0 whenever X takes value less than $aE(X)$, and value a otherwise. Then $X \geq Y$ so $E(X) \geq E(Y) = aE(X)\Pr(X \geq aE(X))$, from which the

result follows). An important consequence of Markov is that if X only takes values $0, 1, 2, \ldots$ then

$$\Pr(X > 0) = \Pr(X \geq 1) \leq E(X)$$

and $\Pr(X = 0) \geq 1 - E(X)$.

Proof. (Claim 57.10) You would expect that this involves an explicit construction, but it is completely non-constructive. We'll show that for large N there is a triangle-free graph on $4N/5$ vertices that has no independent sets of size greater than $t := 8N^{2/3} \log N$, (recall that an independent set is a set of vertices with no two in the set adjacent) and therefore has chromatic number at least $(4N/5)/(8N^{2/3} \log N) = N^{1/3}/(10 \log N)$. Taking $k = N^{1/3}/(10 \log N)$, we get $f(k) \leq 51200k^3 \log^3 k$, for suitable large k (here using $k \geq N^{1/4}$ for large enough N); may have to fix the constant to deal with smaller k.

We find such a graph by choosing randomly. Start with a set of N vertices, and for each pair of vertices decide (independently) with probability $p = 1/N^{2/3}$ to put edge between those two vertices, and with probability $1 - p$ to leave edge out.

What's the expected number of triangles in this graph? It's $\binom{N}{3}p^3 \leq N^3p^3/6 = N/6$. From Markov's inequality it follows that the probability of having more than $N/5$ triangles is at most $5/6$, and so the probability of having no more than $N/5$ triangles is at least $1/6$.

What's the expected number of independent sets of size t? It's $\binom{N}{t}(1-p)^{\binom{t}{2}} \leq N^t e^{-pt^2/4} = e^{t \log N - pt^2/4} = e^{-8N^{2/3} \log^2 N} < 1/100$ (using $1 - p \leq e^{-p}$, and assuming N large). So, again by Markov, the probability of having no independent set of size t (or greater) is at least $99/100$.

Since $1/6 + 99/100 > 1$, there's at least one graph with no more than $N/5$ triangles and no independent set of size t (or greater). On deleting no more than $N/5$ vertices from this graph, we get the graph we claimed exists. □

One of the highpoints of graph theory around the end of the last century was the following result, proved with very sophisticated tools from probability.

Theorem 57.11. *There are constants $c, C > 0$ such that for all $k \geq 2$,*

$$ck^2 \log k \leq f(k) \leq Ck^2 \log k.$$

To end this digression into the "non-locality" of the chromatic number, we mention two more results. The *girth* of a graph is the length of a shortest cycle. If G has girth g then it is easy to see that, starting from any vertex v, it is possible to properly color the ball of radius $\lceil g/2 \rceil - 1$ around v (the set of vertices at distance at most $\lceil g/2 \rceil - 1$ from v, distance measured by the length of the shortest path of edges connecting the vertices), using only two colors. One would then imagine that if a graph had large girth, it should have small chromatic number: one starts with some set of vertices with the property that the balls of radius roughly $g/2$ around those vertices cover the entire graph. One starts coloring from these vertices, using only two colors in total to get out to the edges of each of the balls, and then using a few more colors to properly "merge" the balls. In a seminal result that established the use of tools from probability in graph theory, Erdős proved the following.

Theorem 57.12. *Fix $g \geq 3$, $k \geq 1$. There exists a graph whose girth is at least g and whose chromatic number is at least k.*

Erdős also obtained the following result, in a similar vein.

Theorem 57.13. *For every $k \geq 2$ there exists an $\varepsilon = \varepsilon(k) > 0$ such that for all sufficiently large n there exist graphs on n vertices with chromatic number at least k, but for which the chromatic number of the graph restricted to any subset of vertices of size at most εn is at most 3.*

58. Number of edges needed to force the appearance of any graph

Fix a graph F. Erdős, Stone and Simonovits settled the question of how many edges in a graph on n vertices force a copy of F to appear. The following theorem was proved by Simonovits in the 1970s; it is called the "Erdős-Stone-Simonovits Theorem" because Simonovits realized that it was an easy corollary of a theorem of Erdős and Stone from the 1940s.

Theorem 58.1. *If F has at least one edge then*

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n,F)}{\binom{n}{2}} = 1 - \frac{1}{\chi(F)-1}$$

where $\chi(F)$ is the chromatic number of F.

(More correctly, the limit above exists and equals the claimed value.) In other words, if a graph on n vertices has a proportion (slightly more than) $1-1/(\chi(F)-1)$ of potential edges present, then for all sufficiently large n it contains F as a subgraph; and for all large n there are graphs with a proportion (slightly less than) $1-1/(\chi(F)-1)$ of potential edges present, that do not contain F as a subgraph.

The main point in the proof is the Erdős-Stone Theorem, which says that if G has just a few more edges that $T_{r-1}(n)$, the extremal graph for excluding K_r as a subgraph, then not only does it have K_r, it has a very rich structure that contains many copies of K_r. In what follows, $K_r(s)$ is used to denote a graph on rs vertices $A_1 \cup \ldots \cup A_r$, with $|A_i| = s$ for each i, with an edge from u to v if and only if u and v are in different A_i's. In other words, $K_r(s)$ is the Turán graph $T_r(rs)$.

Theorem 58.2. *Fix $r \geq 2$, $s \geq 1$ and $\varepsilon > 0$. There is $n_0 = n_0(r,s,\varepsilon)$ such that for all $n \geq n_0$, if G is a graph on n vertices with*

$$|E(G)| \geq \left(1 - \frac{1}{r-1} + \varepsilon\right)\frac{n^2}{2}$$

then G has $K_r(s)$ as a subgraph.

We do not at the moment present a proof of Theorem 58.2, but we will give the derivation of Theorem 58.1 from it.

Proof. (Theorem 58.1) Let $\chi(F) = r$. First note that $T_{r-1}(n)$ does not contain F as a subgraph; if it did, then by coloring a vertex of F with color i if the vertex is in A_i (where $A_1, \ldots, A_{r-1}$ are the blocks of the decomposition of the vertex set of $T_{r-1}(n)$ between which the edges go), we would get an $(r-1)$-coloring of F with no edge connecting a pair of vertices with the same color, contradicting $\chi(F) = r$. It follows that for all n, $\mathrm{ex}(n,F) \geq t_{r-1}(n)$. Now since

$$t_{r-1}(n) \geq \left\lfloor \frac{n}{r-1} \right\rfloor^2 \binom{r-1}{2}$$

and (an easy but annoying calculation) for fixed $r \geq 2$

$$\lim_{n\to\infty} \frac{\left\lfloor \frac{n}{r-1} \right\rfloor}{\frac{n}{r-1}} = 1,$$

we get

$$\liminf_{n\to\infty} \frac{\mathrm{ex}(n,F)}{\binom{n}{2}} \geq 1 - \frac{1}{r-1}. \tag{56}$$

In the other direction, fix $\varepsilon > 0$. Let $f : V(F) \to [r]$ be a proper r-coloring of F, and let s be the size of the largest color class in the coloring (i.e., the number of times that the most frequently occurring color occurs). It is evident that F is a subgraph of $K_r(s)$. For all n large enough, by Theorem 58.2 if G on n vertices has more than $(1 - 1/(r-1) + \varepsilon)(n^2/2)$ edges, then G has $K_r(s)$ and so F as a subgraph, and so

$$\limsup_{n\to\infty} \frac{\mathrm{ex}(n,F)}{\binom{n}{2}} \leq 1 - \frac{1}{r-1} + \varepsilon.$$

Since this is true for all $\varepsilon > 0$ it is also true at $\varepsilon = 0$, and combining with (56) gives the theorem. □

Theorem 58.1 essentially settles the question of the extremal number of edges in a graph on n vertices that avoids a graph F, for all F with $\chi(F) = r \geq 3$: if the edge density of G (number of edges divided by number of vertices) is below $1 - (1/(r-1))$ then G might not have F as a subgraph, but if it is above $1 - (1/(r-1))$ then G must have F as a subgraph. When $r = 2$, however, this says very little; only that a positive edge density ensures a copy of F. There are many possible ways for the edge density to be zero (G may have around $n = |V(G)|$ edges, or around $n^{3/2}$, or around $\sqrt{n}\log n$, etc.), and so many possible behaviors for $\mathrm{ex}(n,F)$ with $\chi(F) = 2$. In fact, the problem of determining $\mathrm{ex}(n,F)$ in general in this case is one of the major sources of research problems in extremal graph theory today. We'll now explore some results.

Theorem 58.3. *Let S_{k+1} be the star with $k+1$ vertices (so k edges), and let P_{k+1} be the path with $k+1$ vertices (so also k edges). For each $k \geq 1$,*

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n,S_{k+1})}{n} = \lim_{n\to\infty} \frac{\mathrm{ex}(n,P_{k+1})}{n} = \frac{k-1}{2}.$$

Proof. We will only prove the statement for P_{k+1}, with S_{k+1} being left as an exercise.

If $k|n$, then the graph consisting of n/k copies of K_k, the complete graph on k vertices, does not have P_{k+1} as a subgraph (being connected, it must be a subgraph of one the K_k's, which do not have enough vertices). This shows that

$$\liminf_{n\to\infty,\ k|n} \frac{\mathrm{ex}(n,P_{k+1})}{n} \geq \frac{k-1}{2};$$

that this construction can be modified for general n to show that

$$\liminf_{n\to\infty} \frac{\mathrm{ex}(n,P_{k+1})}{n} \geq \frac{k-1}{2}$$

is left as an exercise.

For the other direction, we will show that if G is any graph on n vertices with m edges, satisfying $m > (k-1)n/2$ then $P_{k+1} \leq G$; this says that

$$\limsup_{n\to\infty} \frac{\mathrm{ex}(n, P_{k+1})}{n} \leq \frac{k-1}{2},$$

and completes the proof of the theorem.

Notice that such a G has average vertex degree greater than $k-1$. It will be helpful to work with a graph that has large *minimum* degree, not just large average degree; to this end, construct G' from G by iterating the process of deleting vertices of degree at most $(k-1)/2$, as long as there are such vertices. Suppose that the resulting graph has n' vertices and m' edges. Since no more that $(k-1)/2$ edges were deleted at each iteration, we have

$$m' > \left(\frac{k-1}{2}\right) n - \left(\frac{k-1}{2}\right)(n-n') = \left(\frac{k-1}{2}\right) n'.$$

So $m' > 0$, and therefore $n' > 0$, and the resulting graph is not empty.

From now on, then, we assume that G satisfies $m > (k-1)n/2$, and also that all vertex degrees in G are greater than $(k-1)/2$. We may also assume that G is connected; for if not, any component of G that maximizes the edge density (ratio of edges to vertices) satisfies both the required edge density and minimum degree conditions, and we may restrict ourselves to considering that component.

Let $v_0, v_1, \ldots, v_t$ be the vertices of a longest path in G (with $v_0v_1, \ldots, v_{t-1}v_t \in E(G)$). If $t \geq k$ then $P_{k+1} \leq G$, so assume for a contradiction that $t \leq k-1$. Because the path we have identified is maximal, all neighbours of v_0 must be among $\{v_1, \ldots, v_t\}$. We claim that $v_0v_t \notin E(G)$. If it was, then for each $i = 1, \ldots, t$ we could use the cycle $v_0v_1 \ldots v_tv_0$ to create a maximal path, using the vertices $v_0, \ldots, v_t$, that has v_i as an endvertex, so by maximality, v_i has all of its neighbours among $\{v_0, \ldots, v_t\} \setminus \{v_i\}$. Connectivity of G then lets us conclude that $V(G) = \{v_0, \ldots, v_t\}$, and so $m \leq \binom{t+1}{2} \leq (k-1)n/2$, contradicting $m > (k-1)n/2$.

Notice that as well as showing $v_0v_t \notin E(G)$, this same argument shows that no cycle of length $t+1$ can be created using the vertices $v_0, \ldots, v_t$.

Now in the index-set $\{2, \ldots, t-1\}$ there are more than $(k-3)/2$ indices i such that $v_0v_i \in E(G)$, and more than $(k-3)/2$ indices i such that $v_tv_{i-1} \in E(G)$. By pigeon-hole principle (and using $t \leq k-1$) it follows that there is an index $i \in \{2, \ldots, t-1\}$ such that $v_0v_i, v_tv_{i-1} \in E(G)$. But then $v_0v_iv_{i+1} \ldots v_tv_{i-1}v_{i-2} \ldots v_0$ is a cycle using all of $v_0, \ldots, v_t$, a contradiction. So in fact $t \geq k$ and $P_{k+1} \leq G$. □

It is a general paradigm that in the family of trees, if a statement is true for paths (the trees with the greatest diameter) and stars (those with the smallest diameter) then it should be true for all trees. Following this intuition, Erdős and Sós conjectured the following.

Conjecture 58.4. *Fix $k \geq 1$ and a tree T with k edges. If graph G on n vertices with m edges satisfies $m > (k-1)n/2$, then G has T as a subgraph.*

The construction given in the proof of Theorem 58.3 also shows that if $k|n$ and T is a tree with k edges then $\mathrm{ex}(n, T) \leq (k-1)n/2$, an an easy modification of the construction (exercise) shows that there is a function $f_k(n)$ with the property $f_k(n) \to 0$ as $n \to \infty$ satisfying

$$\mathrm{ex}(n, T) \leq \left(\frac{k-1}{2} + f_k(n)\right) n.$$

Combined with the Erdős-Sós conjecture, this shows that for any tree T on k edges,

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n,T)}{n} = \frac{k-1}{2}.$$

Conjecture 58.4 has recently been established for all sufficiently large k (and sufficiently large $n = n(k)$). For all k and n, the following weaker statement is fairly easy to obtain, and is left as an exercise.

Proposition 58.5. *Fix $k \geq 1$ and a tree T with k edges. If graph G on n vertices with m edges satisfies $m > (k-1)n$, then G has T as a subgraph.*

This at least shows that for every fixed tree the growth rate of $\mathrm{ex}(n,T)$ is linear in n.

We end with an example of a graph F for which the growth rate of $\mathrm{ex}(n,F)$ is between linear and quadratic in n.

Theorem 58.6. *For every n, if G is a graph on n vertices with m edges that does not contain C_4 (the cycle on 4 vertices) as a subgraph, then*

$$m \leq \frac{n + n\sqrt{4n-3}}{4}.$$

On the other hand, there are infinitely many n for which there is a graph on n vertices with m edges that does not contain C_4 as a subgraph, and that satisfies

$$m \geq \frac{-n + n\sqrt{4n-3}}{4}.$$

Futhermore,

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n,C_4)}{n^{3/2}} = \frac{1}{2}.$$

Proof. Let G be an n-vertex C_4-free graph, and set

$$S = \{(z,\{u,v\}) : z,u,v \in V(G), uz \in E(G), vz \in E(G)\}.$$

Because a graph is C_4-free if and only if for every two vertices their neighbourhoods have no more than one vertex in common, we have $|S| \leq \binom{n}{2}$.

On the other hand,

$$|S| = \sum_{z\in V(G)} \binom{d(z)}{2} = \frac{1}{2}\sum_{z\in V(G)} d^2(z) - \frac{1}{2}\sum_{z\in V(G)} d(z) \geq \frac{2m^2}{n} - m,$$

the inequality by Cauchy-Schwarz-Bunyakovsky. This leads to $2m^2/n - m \leq \binom{n}{2}$ or

$$m \leq \frac{n + n\sqrt{4n-3}}{4},$$

so that $\lim_{n\to\infty} \mathrm{ex}(n,C_4)/n^{3/2} \leq 1/2$ (if it exists).

For the lower bound let p be a prime and define a graph G_p whose vertices are the lines in the (three-dimensional) projective plane over the finite field $\mathbb{F}_p$ with p elements; that is, the vertices are the equivalence classes of $\mathbb{F}_p^3 \setminus \{(0,0,0)\}$ where (a,b,c) is equivalent to (x,y,z) if there is a non-zero λ with $(a,b,c) = \lambda(x,y,z)$; this graph has $(p^3-1)/(p-1) = p^2+p+1$ vertices.

We declare (a,b,c) and (x,y,z) to be adjacent if $ax+by+cz = 0$; observe that this respects the equivalence relation. For each non-zero (x,y,z) the space of solutions to $ax+by+cz = 0$ over $\mathbb{F}_p$ has dimension 2, so each vertex is adjacent to $(p^2-1)/(p-1) = p+1$ vertices.

We now argue that G_p has no C_4's. Indeed, if non-zero (a_1, b_1, c_1) and (a_2, b_2, c_2) are not linear multiples of one another (so they are distinct vertices) then the space of (simultaneous) solutions to

$$\begin{aligned} a_1x + b_1y + c_1z &= 0 \\ a_2x + b_2y + c_2z &= 0 \end{aligned}$$

is at most one-dimensions, so there can be at most one vertex (x, y, z) simultaneously adjacent to distinct (a_1, b_1, c_1) and (a_2, b_2, c_2).

It seems like we have constructed a graph on $n = p^2+p+1$ vertices with $(p+1)(p^2+p+1)/2$ edges that does not contain C_4 as a subgraph. For these values of m and n, we have $m = (n+n\sqrt{4n-3})/4$, so it might appear that this shows that the bound obtained from the first part of the argument is tight for infinitely many n; but we must allow for the possibility that a vertex in the constructed graph is adjacent to itself (this happens for any $(a, b, c) \in \mathbb{F}_p^3$ with $a^2 + b^2 + c^2 = 0$ (in $\mathbb{F}_p$)); this means that we can only be certain that degrees in our graph are between p and $p + 1$, leading to a bound $m \geq (-n + n\sqrt{4n-3})/4$.

All this is based on the assumption that n is of the form $p^2 + p + 1$ for some prime p. We extend to all n via a density argument. Fix $\varepsilon > 0$. For arbitrary n let p be any prime that lies between $(1-\varepsilon)\sqrt{n}$ and $\sqrt{n}$ (such a prime exists, for all sufficiently large n, by the prime number theorem). Add to the graph G_p some isolated vertices so that it has n vertices in total. The result is a C_4-free graph with

$$\frac{p(p^2+p+1)}{2} \geq \frac{p^3}{2} \geq \frac{(1-\varepsilon)^3 n^{3/2}}{2}$$

edges. Since $\varepsilon > 0$ was arbitrary, this combines with the upper bound to show that $\lim_{n\to\infty} \mathrm{ex}(n, C_4)/n^{3/2}$ exists and equals $1/2$. □

As an example of how quickly one reaches the frontiers of research when studying the numbers $\mathrm{ex}(n, F)$ for F with $\chi(F) = 2$, we end this section by noting that the polynomial growth-rate of $\mathrm{ex}(n, C_{2k})$ (which we have determined in Theorem 58.6 to be $n^{3/2}$ for $k = 2$) is only known for $k = 2, 3$ and 5. For example, we do not know if there exists constants c, C and $\gamma > 0$ such that for all sufficiently large n,

$$cn^\gamma \leq \mathrm{ex}(n, C_8) \leq Cn^\gamma.$$

59. Ramsey Theory

Ramsey theory is the study of situations where "total disorder is impossible". Some examples of Ramsey-type theorems include:

- Van der Waerden's theorem: whenever the natural numbers are partitioned into finitely many classes, one of the classes contains arbitrarily long arithmetic progressions.
- Erdős-Szekeres theorem: a sequence of $n^2 + 1$ reals contains a monotone (increasing or decreasing) subsequence of length $n + 1$.
- Party problem: among six people at a party, either some three all mutually know each other, or some three all mutually don't know each other. Here's a quick proof: fix one person. Either she knows at least three people among the other five, or she doesn't know at least three. If she knows three, and any two of those three know each other, those two together with her form a triple who mutually know each other. If she knows three, but no two of those three know each other, those three form a triple who mutually don't know each other. The argument when she doesn't know three is

identical. Note that "six" here can't be replaced by "five": suppose the five people are A, B, C, D and E, and that A and B know each other, and B and C, and C and D, and D and E, and E and A, and no other pairs; it's an easy check then that there are no triples who mutually know each other, or mutually don't know each other.

We generalize this last example: define, for $k \geq 1$, $R(k)$ (the *kth diagonal Ramsey number*) to be the minimum n such that for every coloring $f : E(K_n) \to \{R, B\}$ of the edges of K_n with two colors (red and blue), there is a subset A of the vertices of size k with the property that either $f(e) = R$ for all edges e that join two vertices of A ("there is a red K_k"), or $f(e) = B$ for all such edges ("there is a blue K_k"). The last example shows that $R(3) = 6$. It is also known that $R(4) = 18$, but no other diagonal Ramsey number is known exactly.

In order to prove (using essentially the same proof as already given) that $R(k)$ always exists and is finite, we define an "asymmetric" version of $R(k)$: for $k, \ell \geq 1$, $R(k, \ell)$ is the minimum n such that for every two-coloring of the edges of K_n, there is a either a red K_k or a blue K_ℓ.

Theorem 59.1. *For all $k, \ell \geq 1$, $R(k, \ell)$ exists and satisfies*

$$R(k, 1) = R(1, \ell) = 1$$

for all $k, \ell \geq 1$, and

$$R(k, \ell) \leq R(k - 1, \ell) + R(k, \ell - 1)$$

for all $k, \ell \geq 2$. Moreover,

$$R(k, \ell) \leq \binom{k + \ell - 2}{k - 1}.$$

Proof. We prove the first part by induction on $k + \ell$, with the only non-trivial part being the induction step when $k, \ell \geq 2$. Fix $n = R(k - 1, \ell) + R(k, \ell - 1)$, and fix a two-colouring of the edges of K_n. Pick a vertex v. Either it has at least $R(k - 1, \ell)$ neighbors to which it is joined by a red edge, or $R(k, \ell - 1)$ neighbors to which it is joined by a blue edge. In the former case: either inside those $R(k - 1, \ell)$ neighbors there is a red K_{k-1}; together with v, this gives a red K_k, or inside those $R(k - 1, \ell)$ neighbors there is a blue K_ℓ. The latter case is dealt with similarly.

The second part is an easy induction on $k + \ell$, using Pascal's identity. □

A corollary of this is that $R(k) \leq \binom{2k-2}{k-1}$; using Stirling's approximation to the binomial this is seen to be asymptotically $c(4^k)/\sqrt{k}$ for some constant c. This bound was proved in 1935, and has not been improved a great deal since. The best known upper bound for $R(k)$ is from 2009, due to Conlon: there is $C > 0$ such that

$$R(k) \leq \frac{4^k}{k^{\frac{C \log k}{\log \log k}}}.$$

Notice that although the denominator has gone from polynomial to superpolynomial, we still do not know if there is a $\varepsilon > 0$ such that the "4" in the numerator can be replaced by $4 - \varepsilon$.

Now we think about lower bounds for $R(k)$. Turán observed in the 1930's that $R(k) > (k - 1)^2$: Namely, we can two-color the edges of $K_{(k-1)^2}$ by decomposing the vertex set into $k - 1$ blocks each of size $k - 1$, and coloring edges red if they go between vertices that are in the same block of the decomposition, and blue otherwise. This two-colouring has no red K_k or blue K_k. This bound was substantially improved by Erdős in a landmark 1947 paper.

Theorem 59.2. *If $n, k \geq 1$ satisfy*

$$\binom{n}{k} 2^{1-\binom{k}{2}} < 1$$

then $R(k) > n$. In particular, for $k \geq 3$, $R(k) > 2^{k/2}$.

Proof. Fix such an n and k. Two-color the edges of K_n randomly as follows: each edge is given color "red" with probability $1/2$, and "blue" with probability $1/2$, all choices independent of each other. The probability that this coloring has a red K_k is at most the sum, over all subsets A of size k of the vertex set, of the probability that all edges between vertices of A are red. This is $\binom{n}{k}$ (for the choice of A) times $2^{-\binom{k}{2}}$ (for the $\binom{k}{2}$ edges that must be colored red); by hypothesis this is less than $1/2$. Repeating this argument for a blue K_k, we find that the probability that the coloring has either a red K_k or a blue K_k is less than 1, so there must exist at least one coloring of the edges of K_n that has neither a red K_k nor a blue K_k; i.e., $R(k) > n$.

Using the bound $\binom{n}{k} < n^k/k!$, we find that if $n = 2^{k/2}$ then the required condition reduces to $2^{k/2+1} < k!$, which is true for all $k \geq 3$. □

Some remarks are in order.

- The proof is non-constructive. There is no $\varepsilon > 0$ for which we know of an explicit construction of a two-colouring of the edges of $K_{(1+\varepsilon)^k}$ that avoids a red K_k or a blue K_k, for infinitely many k.
- A more careful analysis of the inequality $\binom{n}{k} 2^{1-\binom{k}{2}} < 1$ leads to the bound

  $$R(k) \geq (1 + o(1)) \frac{k 2^{k/2}}{e\sqrt{2}}.$$

  The best known lower bound for $R(k)$ is from 1977, due to Lovász:

  $$R(k) \geq (1 + o(1)) \frac{\sqrt{2} k 2^{k/2}}{e}.$$

  This is just a factor 2 better than the 1947 bound! We still do not know if there is an $\varepsilon > 0$ such that the "$2^{1/2}$" in the numerator can be replaced by $2^{1/2} + \varepsilon$.
- Since $2^{k/2+1}/k!$ goes to zero as k goes to infinity, the proof actually shows that for large k, *almost all* two-colourings of the edges of $K_{2^{k/2}}$ avoid having a red K_k or a blue K_k, which makes the lack of a $(1 + \varepsilon)^k$ construction very surprising!

60. Restricted intersection theorems

The constructive lower bounds for $R(k)$ that we will discuss all revolve around "restricted intersection theorems". Here is an example, that will be needed for the first construction.

Theorem 60.1. *Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be a set system on ground set $[n]$ satisfying that there are numbers $b > a \geq 0$ with*

- *$|A_i| = b$ for each i, and*
- *$|A_i \cap A_j| = a$ for each $i \neq j$.*

Then $m \leq n$.

This is *Fisher's inequality*, with applications in the design of experiments. Specifically, a *balanced incomplete block design* (BIBD) is a set of n subsets (called *blocks*) of an m-element set X (of *points*) with the property that each

- each block has the same size,
- each point is in the same number of b of blocks, and
- each pair of distinct points is in the same number a of blocks.

Associate to a BIBD $\{B_1, \ldots, B_n\}$ a set system $\{A_1, \ldots, A_m\}$ on $\{1, \ldots, n\}$ by letting A_i be the set of all indices j such that point i is in block B_j. Then each A_i has size b, and for each $i \neq j$ the intersection of A_i and A_j has size a. If $b > a$, then from Theorem 60.1 it follows that $m \leq n$, that is, that at least as many blocks are needed in the BIBD as there are points.

Here is a second example, that will also be needed for the first construction.

Theorem 60.2. *Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be a set system on ground set $[n]$ satisfying*

- *$|A_i|$ is odd for each i, and*
- *$|A_i \cap A_j|$ is even for each $i \neq j$.*

Then $m \leq n$.

Both results are proved using the *linear algebra method*: to upper bound the size of a family $\mathcal{F}$, map each $A \in \mathcal{F}$ to a vector v_A in an appropriately chosen vector space V, in such a way that the v_A's form a linearly independent set of vectors. Then any upper bound on the dimension $\dim(V)$ of V yields an upper bound on $|\mathcal{F}|$.

Proof. (Theorem 60.1) This is linear algebra over $\mathbb{R}^n$ (an n-dimensional vector space over $\mathbb{R}$). To each $A_i \in \mathcal{F}$ associate its *incidence vector* v_i, whose ℓth component is 1 if $\ell \in A_i$ and 0 otherwise; note that $v_i \in \mathbb{R}^n$. Note that $v_i \cdot v_j = |A_i \cap A_j|$ (where $\cdot$ indicates usual dot product); in particular,

$$v_i \cdot v_j = \begin{cases} b & \text{if } i = j \\ a & \text{otherwise.} \end{cases}$$

We claim that the v_i's form a linearly independent set of vectors. Indeed, consider the linear relation

$$\lambda_1 v_1 + \ldots + \lambda_m v_m = 0$$

in $\mathbb{R}^n$ (over $\mathbb{R}$). Taking the dot product of both sides with v_i, we get

$$a\lambda_1 + \ldots + b\lambda_i + \ldots + a\lambda_m = 0$$

and so

$$\lambda_i = \frac{a\lambda_1 + \ldots + a\lambda_i + \ldots + a\lambda_m}{a - b}$$

(note we use here $b > a$). This is independent of i, so all λ_i's share a common value λ, which satisfies $(ma + (b - a))\lambda = 0$, so $\lambda = 0$. Hence the v_i's are linearly independent, and since $\dim(\mathbb{R}^n) = n$ it follows that $m \leq n$. □

Proof. (Theorem 60.2) This is linear algebra over $\mathbb{F}_2^n$ (an n-dimensional vector space over $\mathbb{F}_2$, the field with 2 elements). To each $A_i \in \mathcal{F}$ associate its incidence vector, as before. Note $v_i \cdot v_j = |A_i \cap A_j| \pmod 2$ so

$$v_i \cdot v_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

This immediately shows that the v_i's form a linearly independent set of vectors — indeed, they are pairwise orthogonal! So again $m \leq n$. □

More complicated restricted intersection theorems require slightly more complicated vector spaces. Spaces of polynomials are particularly fruitful. We start with a simple example. A set of vectors $\{c_1, \ldots, c_m\}$ in $\mathbb{R}^n$ is a *one-distance set* if there is some number c such that $d(c_i, c_j) = c$ for all $i \neq j$, where $d(c_i, c_j)$ is the usual Euclidean distance: if $c_i = (c_{i1}, \ldots, c_{in})$, then $d(c_i, c_j) = \sqrt{\sum_{\ell=1}^{n}(c_{i\ell} - c_{j\ell})^2}$.

Theorem 60.3. *Let $\{c_1, \ldots, c_m\}$ be a one-distance set on the surface of the n-dimensional unit sphere in $\mathbb{R}^n$. Then $m \leq n+1$, and this bound is tight.*

Proof. Let c be the common distance. Assume WLOG that $c > 0$ (else, $m \leq 1 \leq n+1$). To each c_i associate the polynomial $p_i(x_1, \ldots, x_n)$ in variables x_1 through x_n given by

$$p_i(x_1, \ldots, x_n) = c^2 - d(x, c_i)^2,$$

where $x = (x_1, \ldots, x_n)$. Note that p_i is an element in the (infinite-dimensional) vector space over $\mathbb{R}$ of polynomials in variables x_1 through x_n. Note also that we have the relations

$$p_i(c_j) = \begin{cases} c^2 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

We claim that the p_i's are linearly independent. Indeed, consider a linear combination $\sum_{i=1}^{m} \lambda_i p_i(x_1, \ldots, x_n) = 0$. Evaluating at c_i, we get $\lambda_i c^2 = 0$, so $\lambda_i = 0$.

Write each c_i as $c_i = (c_{i1}, c_{i2}, \ldots, c_{in})$. Expanding out, we find that

$$\begin{aligned} p_i(x_1, \ldots, x_n) &= c^2 - \sum_{\ell=1}^{n}(x_\ell - c_{i\ell})^2 \\ &= \left(c^2 - \sum_{\ell=1}^{n} c_{i\ell}^2 - \sum_{\ell=1}^{n} x_\ell^2\right) + 2c_{i1}x_1 + \ldots + 2c_{in}x_n. \end{aligned}$$

Noting that $\sum_{\ell=1}^{n} c_{i\ell}^2 = 1$ (since c_i is a point on the surface of the unit sphere), it follows that p_i is in the span of the $n+1$ polynomials $(c^2 - 1 - \sum_{\ell=1}^{n} x_\ell^2), x_1, \ldots, x_n$, so the dimension of the span is at most $n+1$, and $m \leq n+1$, as claimed.

To show that this bound can be obtained, consider the n standard basis vectors in $\mathbb{R}^n$, $e_1, \ldots, e_n$ (where e_i has a 1 in the ith coordinate, and a zero in all other coordinates), together with the point $p_0 = (1+\sqrt{n+1}, \ldots, 1+\sqrt{n+1})$. These $n+1$ points form a one-distance set, and they are all equidistant from the point $(\mu_n, \ldots, \mu_n)$ where $\mu_n = (1+\sqrt{n+1})/(n\sqrt{n+1})$. By a translation followed by a dilation, these $n+1$ points can therefore by embedded in the unit sphere while preserving one-distanceness. □

The main restricted-intersection theorem that we will need for constructive Ramsey bounds is the following, known as the *non-uniform modular Ray-Chaudhuri-Wilson theorem.*

Theorem 60.4. *Fix a prime p, and let L be a subset of $\{0, \ldots, p-1\}$. If $\mathcal{F} = \{A_1, \ldots, A_m\}$ is a set system on ground set $[n]$ satisfying*

- *for all i, $|A_i| \neq \ell \pmod p$ for any $\ell \in L$ and*
- *for all $i \neq j$, $|A_i \cap A_j| = \ell \pmod p$ for some $\ell \in L$,*

then

$$|\mathcal{F}| \leq \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{|L|}.$$

Notice that when $p = 2$ and $L = \{0\}$ we recover Theorem 60.2 with the weaker bound $m \leq n+1$.

Proof. (Theorem 60.4) For each $A_i \in \mathcal{F}$, let v_i be its incidence vector, and associate with A_i the following polynomial in variables $x = (x_1, \ldots, x_n)$:

$$p_i(x) = \prod_{\ell \in L} (x \cdot v_i - \ell),$$

where $\cdot$ indicates the usual dot product. Notice that if $i \neq j$ then $p_i(v_j) = 0 \pmod{p}$, but $p_i(v_i) \neq 0 \pmod{p}$. It follows that the collection of polynomials $p_1, \ldots, p_m$ is linearly independent in the vector space of polynomials in variables $x = (x_1, \ldots, x_n)$ over $\mathbb{F}_p$, the field of p elements. Unfortunately, it is easily seen that the dimension of the span of the p_i is $\binom{n+|L|}{|L|}$ (it is a count of weak compositions), which exceeds the claimed upper bound on m.

The remedy for this problem is as follows. Form polynomial $\tilde{p}_i$ from p_i as follows: expand p_i out as a sum of monomials (over $\mathbb{F}_p$). Obtain $\tilde{p}_i$ by, in every monomial and for every i, replacing every occurrence of $x_i^2, x_i^3, \ldots$, with x_i. Since for all $k \geq 1$, $x_i^k = x_i$ if $x_i = 0$ or 1 (in $\mathbb{F}_p$), it follows that $\tilde{p}_i(v_j) = p_i(v_j)$ for all i and j, so that the collection of polynomials $\tilde{p}_1, \ldots, \tilde{p}_m$ is linearly independent in the vector space of polynomials in variables $x = (x_1, \ldots, x_n)$ over $\mathbb{F}_p$. The $\tilde{p}_i$'s live in the span of the multilinear monomials in variables $x_1, \ldots, x_n$ of degree at most $|L|$ ("multilinear" meaning linear in each variable), and there are exactly $\sum_{i=0}^{|L|} \binom{n}{i}$ such monomials. □

61. Constructive lower bounds for $R(k)$

Z. Nagy gave the following construction of a two-colouring of the edges of K_n, $n = \binom{k-1}{3}$, that contains neither a red K_k nor a blue K_k. Identify the vertices of K_n with subsets of size 3 of $\{1, \ldots, k-1\}$. Color the edge joining A and B "red" if $|A \cap B| = 1$, and "blue" if $|A \cap B| = 0$ or 2 (this covers all possibilities). A collection of vertices mutually joined by red edges corresponds to a set system on ground set $[k-1]$ in which all sets have the same size (3 elements), and all pairwise intersections have the same size (1 element); by Theorem 60.1, there can be at most $k-1$ such vertices. A collection of vertices mutually joined by blue edges corresponds to a set system on ground set $[k-1]$ in which all sets have odd size (3), and all pairwise intersections have even size (0 or 2); by Theorem 60.2, there can be at most $k-1$ such vertices. So this coloring fails to have a red K_k or a blue K_k.

Here is a generalization of Nagy's construction, due to Frankl and Wilson. Let p be a prime number, and for some large n set $N = \binom{n}{p^2-1}$. Identify the vertices of K_N with subsets of size $p^2 - 1$ of $[n]$. Two-color the edges of K_N as follows: if $|A \cap B| \neq p - 1 \pmod{p}$, then color the edge from A to B "blue", and otherwise color it "red". [Notice that when $p = 2$, this is exactly Nagy's construction].

Suppose that this coloring has a blue K_k. This corresponds to a (p^2-1)-uniform set-system $A_1, \ldots, A_m$ on ground set $[n]$ that satisfies the conditions of Theorem 60.4 with $L = \{0, \ldots, p-2\}$ (note that $|A_i| = p^2 - 1 = p - 1 \pmod{p}$), so that

$$k \leq \sum_{i=0}^{p-1} \binom{n}{i}.$$

What if this coloring has a red K_k? This corresponds to a (p^2-1)-uniform set-system $A_1, \ldots, A_m$ on ground set $[n]$, in which $|A_i \cap A_j| \in \{p-1, 2p-1, \ldots, p^2-p-1\}$ for each $i \neq j$, a set of $p-1$ possibilities. Letting p' be any prime greater than $p^2 - 1$, we can again

apply Theorem 60.4 to conclude

$$k \leq \sum_{i=0}^{p-1} \binom{n}{i}.$$

As long as n is large enough ($n \geq 4p$ would be enough), we have

$$\sum_{i=0}^{p-1} \binom{n}{i} < 2\binom{n}{p-1}.$$

We have proven the following.

Claim 61.1. *For every prime p and every large enough n there is a constructive two-coloring of the edges of the complete graph on $\binom{n}{p^2-1}$ vertices, that contains neither a red nor a blue complete graph on $2\binom{n}{p-1}$ vertices.*

For example, when $p = 2$ we find that there is a constructive two-coloring of the edges of the complete graph on $\binom{n}{3}$ vertices, that contains neither a red nor a blue complete graph on $2n$ vertices, showing (constructively) that the number $R(k)$ grows at least cubically. For $p = 3$ we get a constructive two-coloring of the edges of the complete graph on $\binom{n}{8}$ vertices, that contains neither a red nor a blue complete graph on $2\binom{n}{2}$ vertices. Notice that for n sufficiently large we have $\binom{n}{8} \geq n^7$ and $2\binom{n}{2} \leq n^2$, showing (constructively) that there are infinitely many k for which $R(k) > k^{3.5}$.

More generally, for each fixed prime $p > 2$, for n sufficiently large we have $\binom{n}{p^2-1} \geq n^{p^2-2}$ and $2\binom{n}{p-1} \leq n^{p-1}$, showing (constructively) that there are infinitely many k for which

$$R(k) > k^{\frac{p^2-2}{p-1}}.$$

A more careful analysis (and optimization), taking $n = p^3$ and using the Prime Number Theorem, shows that for all k one can show via this construction that

$$R(k) > k^{\frac{(1+o(1))\log k}{4\log\log k}}$$

where $o(1) \to 0$ as $k \to \infty$. While super-polynomial, this falls far short of $2^{k/2}$, and it remains an open problem to find a constructive proof of a bound of the form $R(k) > (1+\varepsilon)^k$ for some $\varepsilon > 0$.

62. Danzer-Grunbaum and "almost" one-distant sets

Theorem 60.3 shows that at most $n+1$ points can be found on the surface of the unit sphere with the property that the distance between any pair is a constant independent of the pair. If we drop the condition that the points lie on the surface of the unit sphere, then the same proof shows that at most $n+2$ points can be selected (because the length of the vectors is no longer constrained to be 1, we need to replace the single polynomial $c^2 - 1 - (x_1^2 + \ldots + x_n^2)$ with the pair $1, x_1^2 + \ldots + x_n^2$ when estimating the dimension of the span). But in fact, it can be shown that in this case, too, at most $n + 1$ points can be selected.

There's a natural relaxation of the problem: at most how many points can we select in $\mathbb{R}^n$, if we require that distance between any pair of the points is close to the same constant? Formally, define, for $\varepsilon \geq 0$, a 1_ε-*distance set* in $\mathbb{R}^n$ to be a set of points $\{c_1, \ldots, c_m\}$ in $\mathbb{R}^n$ with the property that there is some c such that for all $i \neq j$, $(1-\varepsilon)c \leq d(c_i, c_j) \leq (1+\varepsilon)c$, and denote by $f_\varepsilon(n)$ the cardinality of the largest 1_ε-distance set in $\mathbb{R}^n$.

Question 62.1. *Determine, for $\varepsilon \geq 0$ and $n \geq 1$, the quantity $f_\varepsilon(n)$.*

Note that when $\varepsilon = 0$ we have used linear algebra to determine $f_0(n) \leq n + 2$, and we have a construction showing $f_0(n) \geq n + 1$; and as observed above it can actually be shown that $f_0(n) = n + 1$.

In 1962, Danzer and Grunbaum asked the question: at most how many points can we select in $\mathbb{R}^n$, if all triples from among the points form an acute-angled triangle (a triangle all of whose angles are strictly less than $\pi/2$)? They conjectured that the answer is $2n - 1$, and proved this for $n = 2$ (easy) and 3 (less so). Danzer and Grunbaum's conjecture would imply that $f_\varepsilon(n) \leq 2n - 1$ for all sufficiently small ε. In 1983, however, Erdős and Füredi disproved the conjecture, showing that one can select exponentially many points in $\mathbb{R}^n$ such that all triples from among the points form an acute-angled triangle (specifically, they showed that one could select at least $\lfloor (2/\sqrt{3})^n/2 \rfloor$ such points). We modify their proof here to show that for all fixed $\varepsilon > 0$, $f_\varepsilon(n)$ grows exponentially in n. The proof is probabilistic.

Theorem 62.2. *For all $\varepsilon > 0$ there is $\delta > 0$ such that for all sufficiently large n, $f_\varepsilon(n) \geq (1 + \delta)^n$.*

Proof. Fix $\varepsilon > 0$. For a value of m to be determined later in the proof, select m points $c_1, \ldots, c_m$ in $\mathbb{R}^n$ randomly, by for each point choosing each coordinate to be 1 with probability $1/2$ and 0 with probability $1/2$, all selections made independently. Note that this is not necessarily a *set* of m points, as the same point may be chosen many times.

Consider the two points c_1, c_2. The quantity $d(c_1, c_2)^2$, the square of the distance between the two points, is exactly the number of coordinates positions j, $j = 1, \ldots, n$, on which c_1 and c_2 differ. The probability that c_1 and c_2 differ on a particular coordinate is $1/2$, and coordinates are independent, so the number of coordinates on which c_1 and c_2 differ is distributed binomially with parameters n and $1/2$; call this distribution X_{12}.

The expected value of $d(c_1, c_2)^2$ is $n/2$. The probability that $d(c_1, c_2)^2$ is at most $(1 - \varepsilon)^2(n/2)$ is

$$P(X_{12} \leq (1 - \varepsilon)^2(n/2)) = \frac{1}{2^n} \sum_{k=0}^{(1-\varepsilon)^2(n/2)} \binom{n}{k},$$

and by symmetry, the probability that the square of the distance is at least $(1 + \varepsilon)^2(n/2)$ is the same.

Now we need a binomial coefficient estimate. For every $\gamma < 1/2$, there is $\eta > 0$ such that for all large enough n,

$$\sum_{k=0}^{\gamma n} \binom{n}{k} \leq (2 - \eta)^n.$$

To prove this, note that the sum is at most $n\binom{n}{\gamma n}$; the result follows after estimating this binomial coefficient using Stirling's formula and doing some calculus.

It follows that there is a $\delta' > 0$ such that the probability that $d(c_1, c_2)$ is not between $(1-\varepsilon)\sqrt{n/2}$ and $(1+\varepsilon)\sqrt{n/2}$ is at most $(1-\delta')^n$; so the expected number of pairs $i \neq j$ such that $d(c_i, c_j)$ is not between $(1 - \varepsilon)\sqrt{n/2}$ and $(1 + \varepsilon)\sqrt{n/2}$ is at most $\binom{m}{2}(1 - \delta')^n$, which in turn is at most $m^2(1 - \delta')^n$.

We conclude that there is a collection of m points in $\mathbb{R}^n$ with the property that all but at most $m^2(1 - \delta')^n$ pairs from among the m are not at distance between $(1 - \varepsilon)\sqrt{n/2}$ and $(1 + \varepsilon)\sqrt{n/2}$ from each other. By removing at most $m^2(1 - \delta')^n$ points from the collection

(at most one from each of the "bad" pairs), we get a set of at least $m - m^2(1-\delta')$ points, with the property that any pairs from among the set are at distance between $(1-\varepsilon)\sqrt{n/2}$ and $(1+\varepsilon)\sqrt{n/2}$ of each other.

Optimizing by choosing $m = (1/2)(1-\delta')^{-n}$ we get a set of size $(1/4)(1-\delta')^{-n}$. For any $\delta < \delta'$ this is at least $(1+\delta)^n$. □