Enumerative Combinatorics: class notes (temp version for in-class use)

Darij Grinberg

September 17, 2022 (unfinished!)

Preface

These are the class notes from my Math 5705 (Enumerative Combinatorics) class at the UMN in Fall 2018, revised (and digitized) for use in my Math 222 (Enumerative Combinatorics) class at Drexel University in Fall 2019. The websites of these two classes can be found at

`http://www.cip.ifi.lmu.de/~grinberg/t/18f` and
`http://www.cip.ifi.lmu.de/~grinberg/t/19fco`

and include some extra materials (such as homeworks, solutions and references).

This document is not a textbook, not even a set of lecture notes meant to be read on a standalone basis. It is mostly a digitalized (and slightly improved) version of the things I wrote down on the "blackboard" (actually, on papers I projected on a screen through a document camera) during classes. Thus, **proofs are mostly outlined** and hindsight is often missing.

What is this?

These notes cover the basics of enumerative combinatorics, with an emphasis on counting, identities and bijections. We assume that you (the reader) are well familiar with the basics of rigorous mathematics (such as proof methods, the constructions of integers and rationals, and basic properties of finite sets), as covered (for example) in [**?**, Chapters 1–5], [**?**] and [**?**]. We will not rely on any analysis, linear algebra or abstract algebra except for the little that we introduce ourselves.

References

[18f-hw1s] Darij Grinberg, *UMN Fall 2018 Math 5705 homework set #1 with solutions*, `http://www.cip.ifi.lmu.de/~grinberg/t/18f/hw1s.pdf`

[17f-hw1s] Darij Grinberg, *UMN Fall 2017 Math 4707 homework set #1 with solutions*, `http://www.cip.ifi.lmu.de/~grinberg/t/17f/hw1s.pdf`

[19f-hw0s] Darij Grinberg, *Drexel Fall 2019 Math 222 homework set #0 with solutions*, `http://www.cip.ifi.lmu.de/~grinberg/t/19fco/hw0s.pdf`

[19f-hw1s] Darij Grinberg, *Drexel Fall 2019 Math 222 homework set #1 with solutions*, `http://www.cip.ifi.lmu.de/~grinberg/t/19fco/hw1.pdf`

[19s-hw0s] Darij Grinberg, *UMN Spring 2019 Math 4281 homework set #0 with solutions*, `http://www.cip.ifi.lmu.de/~grinberg/t/19s/hw0s.pdf`

[Aigner07] Martin Aigner, *A Course in Enumeration*, Graduate Texts in Mathematics #238, Springer 2007.

[AigZie14] Martin Aigner, Günter M. Ziegler, *Proofs from the Book*, 6th edition, Springer 2018.

[AndFen04] Titu Andreescu, Zuming Feng, *A Path to Combinatorics for Undergraduates: Counting Strategies*, Springer 2004.

[ArdSta10] Federico Ardila, Richard P. Stanley, *Tilings*, The Mathematical Intelligencer **32** (2010), pp. 32–43. A preprint can be found at arXiv:math/0501170v3 .

[BenDre07] Arthur T. Benjamin and Gregory P. Dresden, *A Combinatorial Proof of Vandermonde's Determinant*, The American Mathematical Monthly, Vol. 114, No. 4 (Apr., 2007), pp. 338–341.
(Also available at `http://scholarship.claremont.edu/hmc_fac_pub/524/` .)

[BenQui03] Arthur T. Benjamin and Jennifer J. Quinn, *Proofs that Really Count: The Art of Combinatorial Proof*, The Mathematical Association of America, 2003.

[BenQui04] Arthur T. Benjamin and Jennifer J. Quinn, *Proofs that Really Count: The Magic of Fibonacci Numbers and More*, Mathematical Adventures for Students and Amateurs, (David F. Hayes and Tatiana Shubin, editors), Spectrum Series of MAA, pp. 83–98, 2004.

[BenQui08] Arthur T. Benjamin and Jennifer J. Quinn, *An Alternate Approach to Alternating Sums: A Method to DIE for*, The College Mathematics Journal, Volume 39, Number 3, May 2008, pp. 191-202(12).

[Bourba68] Nicolas Bourbaki, *Theory of Sets*, Springer 1968.

[Day16] Martin V. Day, *An Introduction to Proofs and the Mathematical Vernacular*, 7 December 2016.
https://www.math.vt.edu/people/day/ProofsBook/IPaMV.pdf .

[Grinbe15] Darij Grinberg, *Notes on the combinatorial fundamentals of algebra*, 10 January 2019.
http://www.cip.ifi.lmu.de/~grinberg/primes2015/sols.pdf
The numbering of theorems and formulas in this link might shift when the project gets updated; for a "frozen" version whose numbering is guaranteed to match that in the citations above, see https://github.com/darijgr/detnotes/releases/tag/2019-01-10 .

[Grinbe16] Darij Grinberg, *Notes on linear algebra*, version of 13 December 2016.
https://github.com/darijgr/lina

[GrKnPa94] Ronald L. Graham, Donald E. Knuth, Oren Patashnik, *Concrete Mathematics, Second Edition*, Addison-Wesley 1994.
See https://www-cs-faculty.stanford.edu/~knuth/gkp.html for errata.

[Hammac15] Richard Hammack, *Book of Proof*, 3rd edition.
https://www.people.vcu.edu/~rhammack/BookOfProof/

[Joyner08] W. D. Joyner, *Mathematics of the Rubik's cube*, 19 August 2008.
https://web.archive.org/web/20160304122348/http://www.permutationpuzzles.org/rubik/webnotes/ (link to the PDF at the bottom).

[Kastel61] P. W. Kasteleyn, *The statistics of dimers on a lattice: I. The number of dimer arrangements on a quadratic lattice*, Physica **27** (1961), pp. 1209–1225.
https://doi.org/10.1016/0031-8914(61)90063-5

[LeLeMe16] Eric Lehman, F. Thomson Leighton, Albert R. Meyer, *Mathematics for Computer Science*, revised Tuesday 6th June 2018,
https://courses.csail.mit.edu/6.042/spring18/mcs.pdf .

[LLPT95] D. Laksov, A. Lascoux, P. Pragacz, and A. Thorup, *The LLPT Notes*, edited by A. Thorup, 1995,
http://www.math.ku.dk/~thorup/notes/sympol.pdf .

[Loehr11] Nicholas A. Loehr, *Bijective Combinatorics*, Chapman & Hall/CRC 2011.

[Read80] Ronald C. Read, *A Note on Tiling Rectangles with Dominoes*, Fibonacci Quarterly **18** (1980), pp. 24–27.
https://www.fq.math.ca/Scanned/18-1/read.pdf

[Strick13] Neil Strickland, *MAS201 Linear Mathematics for Applications*, lecture notes, 28 September 2013.
`http://neil-strickland.staff.shef.ac.uk/courses/MAS201/`

[Stucky15] Eric Stucky, *An Exposition of Kasteleyn's Solution of the Dimer Model*, senior thesis at Harvey Mudd College, 2015.
`https://scholarship.claremont.edu/hmc_theses/89/`

[Vorobi02] Nicolai N. Vorobiev, *Fibonacci Numbers*, Translated from the Russian by Mircea Martin, Springer 2002 (translation of the 6th Russian edition).

Class of 2019-09-23

1. Introduction

This is a class on *enumerative combinatorics*: the part of mathematics concerned with the sizes of finite sets, particularly their computation and the proof of equalities between them. More precisely, here are what I consider to be the three main threads of enumerative combinatorics:

- **Counting** – i.e., finding formulas for the sizes of certain finite sets. For example, we count the permutations of the set $\{1,2,\ldots,n\}$, or the k-subsets of $\{1,2,\ldots,n\}$ that contain no two consecutive elements. "Count" means finding a formula that expresses the number of such permutations or k-subsets in terms of n and k.
- **Proving polynomial identities** (such as the binomial formula $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$ or various deeper ones).
- **Finding and studying interesting maps between finite sets**. A basic example of such a map is the "bit-set encoding": the bijection from the set of all subsets of $\{1,2,\ldots,n\}$ (for a fixed positive integer n) to the set of all n-tuples $(i_1,i_2,\ldots,i_n) \in \{0,1\}^n$ (known as "length-n bitstrings") which sends each subset S of $\{1,2,\ldots,n\}$ to the n-tuple $(i_1,i_2,\ldots,i_n)$, where $i_k = \begin{cases} 1, & \text{if } k \in S; \\ 0, & \text{if } k \notin S \end{cases}$.
  We will care particularly about bijections, since they directly help in counting, but even non-bijective maps are fundamental to enumerative combinatorics.

You will see more examples of each of these three threads all over this class, starting with this introductory chapter.

We will also occasionally see some connections to linear algebra, abstract algebra, number theory and graph theory. There are other threads in enumerative combinatorics that we are not going to encounter (or only tangentially): applications (mostly), connections to

representation theory or geometry, asymptotics and many more. A one-semester course needs to have its limits!

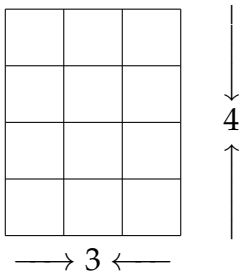
First, I will discuss some interesting (if you share my taste) questions, in no particular order. Not all of them will be answered right away.

1.1. Domino tilings

1.1.1. The problem

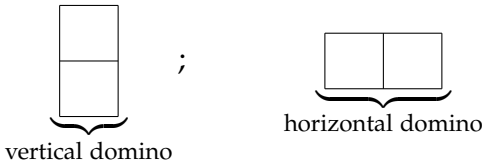
Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Here and in the following, $\mathbb{N}$ means the set $\{0, 1, 2, \ldots\}$.

Let $R_{n,m}$ denote an $n \times m$-*rectangle*, i.e., a rectangle with width n and height m. (We imagine a specific such rectangle drawn somewhere in the plane.) For example, $R_{3,4}$ looks like this:[1]



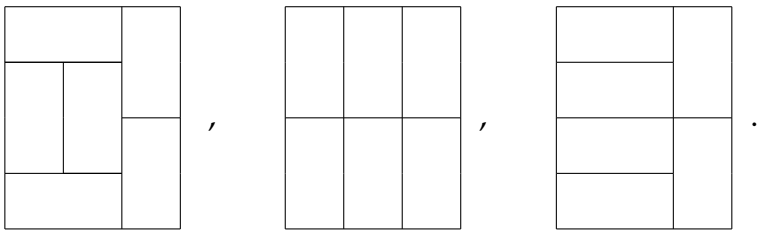
(1)

A *domino* shall mean a 1×2-rectangle or a 2×1-rectangle. More specifically: A *vertical domino* shall mean a 1×2-rectangle; a *horizontal domino* shall mean a 2×1-rectangle. Here is how they look like:



A *domino tiling* of $R_{n,m}$ is a way to cover the rectangle $R_{n,m}$ with non-overlapping dominos.

For example, here are three domino tilings of the rectangle $R_{3,4}$ (which rectangle you have seen in (**??**)):



(Now, of course, we are no longer drawing the grid lines, but only the outlines of the dominos.)

[1] We subdivide it with grid lines just to show its dimensions.

We can now state our first enumeration (i.e., counting) problem: How many domino tilings does $R_{n,m}$ have?

As we just saw, $R_{3,4}$ has at least 3 domino tilings, but in fact you can find several more. Counting them all is at the very least an unpleasant exercise in carefulness. Let us try a simpler example:

Example 1.1.1. Here are all domino tilings of $R_{3,2}$:

If we are to solve the above problem in general, our first step should be making it rigorous. We said that a domino tiling should be a way to cover the rectangle $R_{n,m}$ with non-overlapping dominos. What does "cover" mean, and what does "non-overlapping" mean? Visually, it is pretty clear, but we do not have bulletproof mathematical definitions yet. There are two ways to create such definitions:

- **The geometric way:** We really define $R_{n,m}$ as a rectangle of width n and height m in the Euclidean plane; for example, let us pick the rectangle with vertices $(0,0)$, $(n,0)$, (n,m) and $(0,m)$ (where we model the Euclidean plane through a Cartesian coordinate system as usual). We say that a set of dominos *covers* $R_{n,m}$ if their union (as sets) is $R_{n,m}$. It is harder to define what it means for a set of dominos to be *non-overlapping*; clearly, this is not quite the same as them being disjoint as sets (because they are allowed to have edges or vertices in common). There are several good ways to define non-overlappingness[2]. Unfortunately, once all these definitions are made, it is still far from clear how to **reason** about them rigorously! For example, it may seem obvious, but why exactly must all the dominos in a domino tiling of $R_{n,m}$ be "snapped to the grid" (i.e., why must their corners be grid points[3])? This is indeed true, but proving this would take nontrivial amounts of work. Thus, even our previous observation that $R_{3,2}$ has three domino tilings (shown visually in Example **??**) would become a nontrivial theorem. Thus, we leave this geometric model of domino tilings aside, and instead define things in ...

- **The combinatorial way:** We redefine $R_{n,m}$ as the set $[n] \times [m]$, where we set $[k] = \{1, 2, \ldots, k\}$ for each $k \in \mathbb{N}$.

  Its elements thus are the pairs (i,j) with $i \in [n]$ and $j \in [m]$; we call these pairs "*squares*". Thus, $R_{n,m}$ is a **finite** set of size $|R_{n,m}| = nm$.

[2]For example, you can say that two dominos are *non-overlapping* if their intersection is either the empty set or a point or a line segment. Then you can say that a set of dominos is *non-overlapping* if any two distinct dominos in it are non-overlapping.

[3]A *grid point* means a point with integer coordinates.

A *vertical domino* shall mean a set of the form $\{(i,j),\ (i,j+1)\}$ for some $i,j \in \mathbb{Z}$.

A *horizontal domino* shall mean a set of the form $\{(i,j),\ (i+1,j)\}$ for some $i,j \in \mathbb{Z}$.

A *domino* shall mean a set that is either a vertical domino or a horizontal domino.

If S is a set of squares (for example, $R_{n,m}$), then a *domino tiling* of S shall mean a set $\{S_1, S_2, \ldots, S_k\}$ of **disjoint** dominos whose union is S (that is, $S_1 \cup S_2 \cup \cdots \cup S_k = S$).

A few words are in order about what this has to do with our visual concept of domino tilings.

First of all, why are we suddenly considering the finite set $R_{n,m}$ to be a "rectangle"? Because we are no longer thinking in terms of all points in the plane, but rather thinking in terms of *grid squares* (as in (**??**)). Thus, the rectangle $R_{n,m}$ is no longer an (infinite) set of points, but now becomes a (finite) set of grid squares that lie in this rectangle. We label these grid squares by pairs of integers (namely, we label each grid square by the pair (i,j) of Cartesian coordinates of its northeastern corner[4]; thus, the southwesternmost square of $R_{n,m}$ is labeled $(1,1)$, and the eastern neighbor of a square (i,j) is $(i+1,j)$, whereas the northern neighbor of a square (i,j) is $(i,j+1)$). In other words, the square in column i (counted from the left) and row j (counted from the bottom) is labelled by the pair (i,j).

Having thus redefined the rectangle $R_{n,m}$ as a finite set of grid squares, we then do the same for dominos and domino tilings. A domino, too, is not an infinite set any more, but just a set of two adjacent grid squares. It is a vertical domino if these grid squares differ in their y-coordinate (i.e., have the forms (i,j) and $(i,j+1)$ for some $i,j \in \mathbb{Z}$), and it is a horizontal domino if these grid squares differ in their x-coordinate (i.e., have the forms (i,j) and $(i+1,j)$ for some $i,j \in \mathbb{Z}$). From this point of view, two dominos are non-overlapping if they are literally disjoint (because they are sets of grid squares now, and

[4]Here is how $R_{3,4}$ looks like with each square labeled:

$(1,4)$	$(2,4)$	$(3,4)$
$(1,3)$	$(2,3)$	$(3,3)$
$(1,2)$	$(2,2)$	$(3,2)$
$(1,1)$	$(2,1)$	$(3,1)$

thus disjointness means that they have no grid squares in common; it does not matter if they share an edge).

So we have obtained a new model for domino tilings, with simpler definitions and with all sets involved being finite. This kind of model is called a **discrete model**. It is much more manageable than the geometric one, and in particular, almost everything that is visually obvious is actually straightforward to prove in this model (unlike in the geometric one). For example, it is easy to rigorously reproduce our result from Example **??** saying that $R_{3,2}$ has 3 domino tilings. In the discrete model, these domino tilings are[5]

$$\left\{ \underbrace{\{(1,1),(1,2)\}}_{\substack{\text{vertical domino}\\ \text{covering the}\\ \text{leftmost column}}}, \underbrace{\{(2,1),(2,2)\}}_{\substack{\text{vertical domino}\\ \text{covering the}\\ \text{middle column}}}, \underbrace{\{(3,1),(3,2)\}}_{\substack{\text{vertical domino}\\ \text{covering the}\\ \text{righmost column}}} \right\},$$

$$\left\{ \underbrace{\{(1,1),(2,1)\}}_{\substack{\text{horizontal domino}\\ \text{in the bottom row}}}, \underbrace{\{(1,2),(2,2)\}}_{\substack{\text{horizontal domino}\\ \text{in the top row}}}, \underbrace{\{(3,1),(3,2)\}}_{\substack{\text{vertical domino}\\ \text{covering the}\\ \text{righmost column}}} \right\},$$

$$\left\{ \underbrace{\{(1,1),(1,2)\}}_{\substack{\text{vertical domino}\\ \text{covering the}\\ \text{leftmost column}}}, \underbrace{\{(2,1),(3,1)\}}_{\substack{\text{horizontal domino}\\ \text{in the bottom row}}}, \underbrace{\{(2,2),(3,2)\}}_{\substack{\text{horizontal domino}\\ \text{in the top row}}} \right\}.$$

You do need a bit of work to verify that no other domino tilings of $R_{3,2}$ exist; but it is very much doable. The surefire (but boring) way is to simply check all possibilities by **brute force**: There are only 7 dominos that lie inside[6] $R_{3,2}$, and clearly any domino tiling must consist of some of these 7 dominos; now, you can check all the 2^7 possible subsets. (Of course, you can be a lot less stupid.)

From now on, we shall always be using the discrete model when we study domino tilings – i.e., we define $R_{n,m}$, dominos and domino tilings via the combinatorial way.

If $n, m \in \mathbb{N}$, then let us define an integer $d_{n,m}$ by

$$d_{n,m} = (\#\text{ of domino tilings of } R_{n,m}).$$

[5] listed here in the same order in which they appeared in Example **??**

[6] To "lie inside" $R_{3,2}$ means to be a subset of $R_{3,2}$ here.

Here and in the following, the symbol "#" always means "number" (or "the number", depending on context).

Our problem thus asks us to compute $d_{n,m}$. In Example **??**, we have seen that $d_{3,2} = 3$. For any fixed n and m, we can technically compute $d_{n,m}$ by brute force (i.e., trying out all possible subsets of the set of dominoes lying inside $R_{n,m}$, and counting the domino tilings among them). But this becomes forbiddingly slow when n and m get even a little bit large (say, $n = 8$ and $m = 8$). We are looking for something better: for an explicit formula for $d_{n,m}$ if possible, and otherwise at least for faster algorithms that compute $d_{n,m}$.

1.1.2. The odd-by-odd case and the sum rule

We begin with a particularly simple case:

Proposition 1.1.2. Assume that n and m are odd. Then, $d_{n,m} = 0$.

Proof sketch. We have assumed that n and m are odd. Thus, the product nm is odd as well. In other words, the size $|R_{n,m}|$ is odd (since $|R_{n,m}| = nm$).

But each domino has even size (in fact, it has size 2).

If the set $R_{n,m}$ had a domino tiling, then the size $|R_{n,m}|$ of $R_{n,m}$ would equal the sum of the sizes of all the dominos in the tiling (because $R_{n,m}$ is the union of the dominos, and the dominos are disjoint). But the size $|R_{n,m}|$ is odd, whereas the sum of the sizes of all the dominos in the tiling is even (since each domino has even size); thus the former cannot equal the latter. This shows that the set $R_{n,m}$ has no domino tilings. In other words, the # of domino tilings of $R_{n,m}$ is 0. In other words, $d_{n,m} = 0$. □

It is worth being a little bit more detailed once and look under the hood of this proof. We have used the following basic fact:

Theorem 1.1.3 (The sum rule)**.** If a finite set S is the union of k disjoint sets $S_1, S_2, \ldots, S_k$, then

$$|S| = |S_1| + |S_2| + \cdots + |S_k| .$$

Theorem **??** is known as the *sum rule* or the *addition rule*, and is so fundamental for all of mathematics that you have probably not even noticed us tacitly using it in the proof of Proposition **??** above. We shall not prove Theorem **??**, since this is a job for "axiomatic foundations of mathematics" courses and depends on the "implementation details" of your mathematical "standard library" (such as: how do you define the size of a finite set?).[7]

We have also used the visually obvious fact that $|R_{n,m}| = nm$ (that is, $R_{n,m}$ has nm squares). Formally speaking, this is a consequence of another basic fact:

[7] For example, if you define your sets and numbers in the old-fashioned Bourbakist way, then you can find Theorem **??** with proof in [**?**, Chapter III, §3.3, Corollary]. If you are using constructivist foundations, then Theorem **??** can be proven by induction on k (see [**?**, proof of 1.2] for the details of this induction proof), relying on the fact that any two disjoint finite sets A and B of

Theorem 1.1.4 (The product rule for two sets)**.** Let X and Y be two finite sets. Then, $X \times Y$ is a finite set with size $|X \times Y| = |X| \cdot |Y|$.

Again, we will not prove Theorem **??**, as it is sufficiently elementary.[8]

Let us now restate our above proof of Proposition **??** in a way that makes the uses of Theorem **??** and of Theorem **??** in it explicit:

Proof of Proposition **??** *(detailed version).* For each $k \in \mathbb{N}$, we have $[k] = \{1, 2, \ldots, k\}$ and thus $|[k]| = |\{1, 2, \ldots, k\}| = k$. Hence, $|[n]| = n$ and $|[m]| = m$. But $R_{n,m} = [n] \times [m]$ and thus

$$|R_{n,m}| = |[n] \times [m]| = \underbrace{|[n]|}_{=n} \cdot \underbrace{|[m]|}_{=m} \qquad \text{(by Theorem ??, applied to } X = [n] \text{ and } Y = [m]\text{)}$$
$$= nm.$$

We have assumed that n and m are odd. Thus, the product nm is odd as well. In other words, the size $|R_{n,m}|$ is odd (since $|R_{n,m}| = nm$).

Our next goal is to show that the set $R_{n,m}$ has no domino tilings.

Indeed, let T be a domino tiling of $R_{n,m}$. We will derive a contradiction.

Write T in the form $T = \{S_1, S_2, \ldots, S_k\}$, where $S_1, S_2, \ldots, S_k$ are distinct dominos[9]. Then, the sets $S_1, S_2, \ldots, S_k$ are dominos; thus, their sizes $|S_1|, |S_2|, \ldots, |S_k|$ are even (since the size of each domino is even[10]). Hence, the sum $|S_1| + |S_2| + \cdots + |S_k|$ is even (being a sum of even integers).

But the finite set $R_{n,m}$ is a union of the k disjoint sets $S_1, S_2, \ldots, S_k$ (since $\{S_1, S_2, \ldots, S_k\} = T$ is a domino tiling of $R_{n,m}$). Hence, Theorem **??** (applied to $S = R_{n,m}$) yields $|R_{n,m}| = |S_1| + |S_2| + \cdots + |S_k|$. Hence, $|R_{n,m}|$ is even (since $|S_1| + |S_2| + \cdots + |S_k|$ is even). This contradicts the fact that $|R_{n,m}|$ is odd.

Now, forget that we fixed T. We thus have found a contradiction for each domino tiling T of $R_{n,m}$. This shows that there exists no domino tiling of $R_{n,m}$. In other words, the # of domino tilings of $R_{n,m}$ is 0. In other words, $d_{n,m} = 0$. This proves Proposition **??**. □

In the future, we will be using the sum rule (Theorem **??**) many times, usually without even mentioning it.

$|A \cup B| = |A| + |B|$. The latter fact (which is, of course, essentially equivalent to the particular case of Theorem **??** for $k = 2$) can be proven directly by explicitly constructing a bijection $A \cup B \to [n + m]$ out of two bijections $A \to [n]$ and $B \to [m]$ (see [**?**, proof of 1.32] for the details of this construction). But we will not dwell on fundamental issues like this here.

[8] See [**?**, 1.5] for a proof (even of a more general statement).

[9] Here, we are tacitly using the fact that T is finite. Why is T finite? Intuitively it is obvious. More rigorously, you can argue this as follows: There are only finitely many dominos that are subsets of $R_{n,m}$. The set T, being a domino tiling of $R_{n,m}$, must consist entirely of such dominos; thus, it must be a subset of the (finite) set of these dominos. Hence, T is itself finite (since a subset of a finite set is always finite).

We trust you to make such arguments whenever necessary; we will not dwell on them in the future.

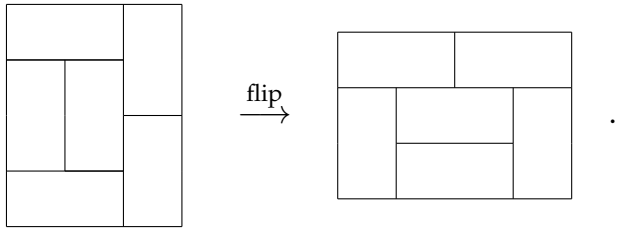
[10] Indeed, the size of each domino is 2.

1.1.3. The symmetry and the bijection rule

Thus we have handled at least one case of our counting problem: the case when n and m are odd. It remains to handle the case when at least one of n and m is even. More precisely, it suffices to handle the case when n is even, because of the following symmetry in the problem:

Proposition 1.1.5. Let $n, m \in \mathbb{N}$. Then, $d_{n,m} = d_{m,n}$.

Proof sketch. The idea is very simple: The rectangle $R_{m,n}$ can be obtained by "flipping" the rectangle $R_{n,m}$ across the line with equation $x = y$ (in Cartesian coordinates). This "flip" operation turns domino tilings of $R_{m,n}$ into domino tilings of $R_{n,m}$ and vice versa; here is an example:



Thus, the domino tilings of $R_{m,n}$ are in 1-to-1 correspondence with the domino tilings of $R_{n,m}$. This entails that the # of the former equals the # of the latter. Since the # of the former is $d_{m,n}$ (by the definition of $d_{m,n}$), whereas the # of the latter is $d_{n,m}$ (by the definition of $d_{n,m}$), we can rewrite this as follows: $d_{m,n} = d_{n,m}$. This proves Proposition **??**. □

It is worth expanding this proof just to see what exactly we have done; again, the underlying principle is very basic but worth stating at least once. The domino tilings of $R_{m,n}$ are not literally the same as the domino tilings of $R_{n,m}$ (unless $n = m$ or one of n and m is 0); yet, we have argued that the former are in 1-to-1 correspondence with the latter, and therefore equinumerous to the latter. Formally, a 1-to-1 correspondence between two sets of objects is given by a map from one set to the other, but it cannot be just any map: It has to be a bijection (i.e., a bijective map)[11]. Thus, what we have used is the fact that if there is a bijection between two sets, then these two sets have the same size. Let us state this a little bit more explicitly:

[11] We refer to places like [**?**, §4.4–4.5], [**?**, §12.2] or [**?**, §3.F] for basic properties of bijections and bijectivity. Here come some brief reminders: A map $f : X \to Y$ between two sets X and Y is said to be

- *injective* if it has the property that $(f(x_1) = f(x_2)) \Longrightarrow (x_1 = x_2)$ for any two elements $x_1, x_2 \in X$ (or, equivalently, if it maps any two distinct elements of X to two distinct elements of Y);
- *surjective* if it has the property that for each $y \in Y$, there exists at least one $x \in X$ satisfying $f(x) = y$ (in other words, every element of Y is a value of f);
- *bijective* if it is both injective and surjective.

It is easy to see that a map is bijective if and only if it is invertible (i.e., has an inverse).

Theorem 1.1.6. If X and Y are two sets, and if $f : X \to Y$ is a bijection (i.e., a bijective map), then

$$|X| = |Y| . \tag{2}$$

Theorem **??** is known as the *bijection principle* or the *bijection rule*, and is sufficiently basic that some authors consider it part of the definition of the cardinality of a set; we are not going to reference it explicitly every time we use it. But just this one time, let us do so, and while at that, also formalize the definition of the "flip" operation that was used in our proof of Proposition **??**:

Proof of Proposition **??** *(detailed version).* Define the map

$$\begin{aligned} F : R_{n,m} &\to R_{m,n}, \\ (i,j) &\mapsto (j,i) . \end{aligned}$$

(This notation is saying "the map F from $R_{n,m}$ to $R_{m,n}$ that sends each element (i,j) of $R_{n,m}$ to the element (j,i) of $R_{m,n}$". To see why this map is well-defined, just recall that $R_{n,m} = [n] \times [m]$ and $R_{m,n} = [m] \times [n]$, which is why $(i,j) \in R_{n,m}$ will always lead to $(j,i) \in R_{m,n}$.)

Visually speaking, this map F simply flips each square of the rectangle $R_{n,m}$ across the line with equation $x = y$. This yields a square of the rectangle $R_{m,n}$, of course. Thus, it is clear that the map F is a bijection. This can be proved rigorously as follows: The map

$$\begin{aligned} G : R_{m,n} &\to R_{n,m}, \\ (i,j) &\mapsto (j,i) \end{aligned}$$

is well-defined[12] and inverse to F [13]. Thus, the map F is invertible, i.e., is a bijection. Note that its inverse map G was defined in the same way as F, but simply with the roles of n and m interchanged.

The map F is a bijection, but it is not the bijection that we are going to apply Theorem **??** to. (If we applied Theorem **??** to $f = F$, then we would conclude that $|R_{n,m}| = |R_{m,n}|$, that is, $nm = mn$, which is reassuring but not what we are trying to prove.)

The map F merely flips the squares of $R_{n,m}$ across the $x = y$ line; we want a map that flips domino tilings. Of course, to flip a domino tiling, we have to flip each domino in it; and to flip a domino, we have to flip each square in it. Thus, we define the following two maps:

- Define the map

$$\begin{aligned} F_{\text{dom}} : \{\text{dominos inside } R_{n,m}\} &\to \{\text{dominos inside } R_{m,n}\}, \\ D &\mapsto \{F(d) \mid d \in D\} . \end{aligned}$$

[12] This is proved in the same way as we showed that F is well-defined.

[13] since $F \circ G = \text{id}$ (because each $(i,j) \in R_{m,n}$ satisfies $(F \circ G)(i,j) = F\left(\underbrace{G(i,j)}_{=(j,i)}\right) = F(j,i) = (i,j) = \text{id}(i,j)$) and $G \circ F = \text{id}$ (for similar reasons)

(To spell this out: The map F_{dom} sends each domino D to the domino $\{F(d) \mid d \in D\}$, which is obtained from D by flipping each square $d \in D$. In other words, it flips each domino by applying the flip map F to each square of the domino.)

It is easy to see that this map F_{dom} is a bijection. (Indeed, it has an inverse G_{dom}, which is defined in the same way but with the roles of n and m interchanged.)

- Define the map

$$F_{\text{til}} : \{\text{domino tilings of } R_{n,m}\} \to \{\text{domino tilings of } R_{m,n}\}, \\ T \mapsto \{F_{\text{dom}}(D) \mid D \in T\}.$$

(To spell this out: The map F_{til} sends each domino tiling T to the domino tiling $\{F_{\text{dom}}(D) \mid D \in T\}$, which is obtained from T by flipping each domino $D \in T$. In other words, it flips each domino tiling by applying the flip map F_{dom} to each domino of the tiling.)

It is easy to see that this map F_{til} is a bijection. (Indeed, it has an inverse G_{til}, which is defined in the same way but with the roles of n and m interchanged.)

To be fully rigorous, we would have to check that these two maps F_{dom} and F_{til} are well-defined (i.e., that flipping a domino inside $R_{n,m}$ really results in a domino inside $R_{m,n}$, and that flipping a domino tiling of $R_{n,m}$ really results in a domino tiling of $R_{m,n}$), but this is intuitively clear, and the formal proof can easily be constructed by just "following your nose"[14], which is why we omit it.

Anyway, we now have constructed a map
$F_{\text{til}} : \{\text{domino tilings of } R_{n,m}\} \to \{\text{domino tilings of } R_{m,n}\}$ and showed that it is a bijection. Hence, (**??**) (applied to $X = \{\text{domino tilings of } R_{n,m}\}$ and $Y = \{\text{domino tilings of } R_{m,n}\}$ and $f = F_{\text{til}}$) shows that

$$|\{\text{domino tilings of } R_{n,m}\}| = |\{\text{domino tilings of } R_{m,n}\}|. \tag{3}$$

But the definition of $d_{n,m}$ yields

$$d_{n,m} = (\#\text{ of domino tilings of } R_{n,m}) = |\{\text{domino tilings of } R_{n,m}\}|.$$

The same reasoning shows that

$$d_{m,n} = |\{\text{domino tilings of } R_{m,n}\}|.$$

In view of the latter two equalities, we can rewrite (**??**) as $d_{n,m} = d_{m,n}$. Thus Proposition **??** is proven. □

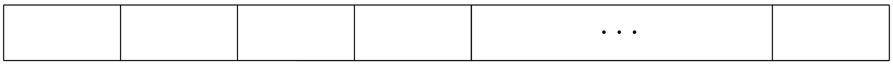
1.1.4. The $m = 1$ case

Let us bite another piece off the problem:

[14] For example: If D is a horizontal domino $\{(i,j),(i+1,j)\}$, then flipping it yields the vertical domino $\{(j,i),(j,i+1)\}$. Likewise, flipping a vertical domino yields a horizontal domino. Thus, flipping a domino yields a domino. (Note that this would be false if our definition of dominos didn't have the symmetry built in!)

Proposition 1.1.7. Assume that $m = 1$ and that n is even. Then, $d_{n,m} = 1$.

Proof sketch. We must show that there is exactly one domino tiling of $R_{n,m}$. But this is visually obvious: The rectangle $R_{n,m}$ has only one row, and this row has an even number of squares, so we can cover it with horizontal dominos in only one way. Here is how this domino tiling looks like:



.

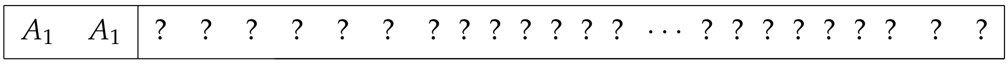
This is not yet a rigorous proof, but it is fairly easy to turn it into one. Rigorously speaking, the domino tiling we just described is[15]

$$\begin{aligned} &\{\{(1,1),(2,1)\},\ \{(3,1),(4,1)\},\ \{(5,1),(6,1)\},\ \ldots,\ \{(n-1,1),(n,1)\}\} \\ &= \{\{(2k-1,1),(2k,1)\} \mid k \in \{1,2,\ldots,n/2\}\}. \end{aligned} \tag{4}$$

It is instantly clear that this is a domino tiling of $R_{n,m}$. In order to show that this is the only domino tiling of $R_{n,m}$, we can let T be any domino tiling of $R_{n,m}$, and then argue as follows:

- The square $(1,1)$ must be contained in some domino $A_1 \in T$ (since T is a domino tiling). This domino A_1 must be either $\{(1,1),(2,1)\}$ or $\{(0,1),(1,1)\}$ or $\{(1,1),(1,2)\}$ or $\{(1,0),(1,1)\}$ (since these are the only dominos that contain $(1,1)$). But out of these four dominos, only $\{(1,1),(2,1)\}$ is a subset of $R_{n,m}$. Hence, A_1 must be the domino $\{(1,1),(2,1)\}$. Thus, $(2,1)$ is also contained in A_1.

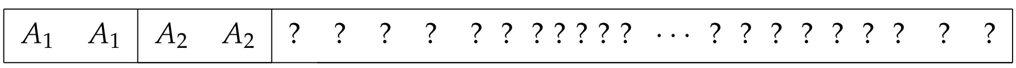
  Now we know that our tiling T looks like this:

  (where we have labeled the leftmost two squares with "A_1" to signify that they are contained in the domino A_1).

- The square $(3,1)$ must be contained in some domino A_2 from T. This domino A_2 must be either $\{(3,1),(4,1)\}$ or $\{(2,1),(3,1)\}$ or $\{(3,1),(3,2)\}$ or $\{(3,0),(3,1)\}$ (since these are the only dominos that contain $(3,1)$). But out of these four dominos, only $\{(3,1),(4,1)\}$ and $\{(2,1),(3,1)\}$ are subsets of $R_{n,m}$. Hence, A_2 must be either $\{(3,1),(4,1)\}$ or $\{(2,1),(3,1)\}$. But the dominos in a domino tiling must be disjoint (by definition); hence, A_2 cannot be $\{(2,1),(3,1)\}$ (because if A_2 was $\{(2,1),(3,1)\}$, then it would fail to be disjoint from $A_1 = \{(1,1),(2,1)\}$). Thus, A_2 must be $\{(3,1),(4,1)\}$. Hence, $(4,1)$ is also contained in A_2.

  Now we know that our tiling T looks like this:

- The square $(5,1)$ must be contained in some domino A_2 from T. This domino A_2 must be either $\{(5,1),(6,1)\}$ or $\{(4,1),(5,1)\}$ or $\{(5,1),(5,2)\}$ or $\{(5,0),(5,1)\}$

[15] Pay attention to where the set braces are! This is a set of sets of pairs of numbers.

(since these are the only dominos that contain $(5,1)$). But out of these four dominos, only $\{(5,1),(6,1)\}$ and $\{(4,1),(5,1)\}$ are subsets of $R_{n,m}$. Hence, A_3 must be either $\{(5,1),(6,1)\}$ or $\{(4,1),(5,1)\}$. But the dominos in a domino tiling must be disjoint (by definition); hence, A_3 cannot be $\{(4,1),(5,1)\}$ (because if A_3 was $\{(4,1),(5,1)\}$, then it would fail to be disjoint from $A_2 = \{(3,1),(4,1)\}$). Thus, A_3 must be $\{(5,1),(6,1)\}$. Hence, $(6,1)$ is also contained in A_3.

Now we know that our tiling T looks like this:

A_1 A_1	A_2 A_2	A_3 A_3	? ? ??????? $\cdots$??? ? ? ? ? ? ?

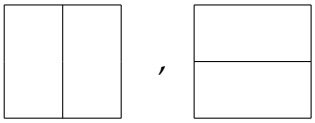
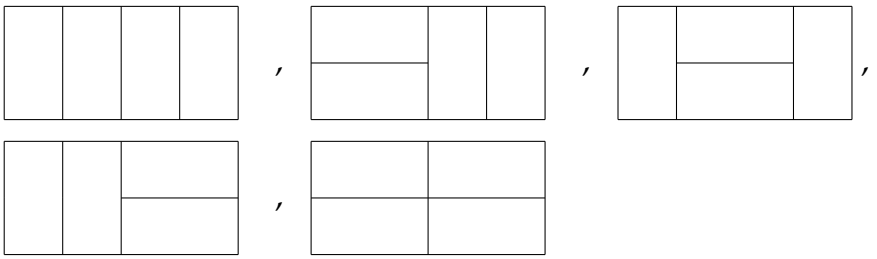
... and so on, proceeding further and further right until you hit the "eastern wall" of $R_{n,m}$ (that is, the square $(n,1)$). Thus, the dominos appearing in T are uniquely determined: They must be $A_1 = \{(1,1),(2,1)\}$, $A_2 = \{(3,1),(4,1)\}$, $A_3 = \{(5,1),(6,1)\}$ and so on. This is precisely the one tiling that we presented in (**??**). Thus, that one tiling is the only domino tiling of $R_{n,m}$.

To be fully rigorous, this argument should be formalized as an induction proof (feel free to do so!), but even if I wake you up at night, you will know how to construct this argument if necessary, because the idea behind it is glaringly obvious (just walk the rectangle $R_{n,m}$ from its western wall to its eastern wall, and observe that at each step, there is only one possible domino that fits in the rectangle without overlapping with the previous domino). □

1.1.5. The $m = 2$ case and Fibonacci numbers

Between Proposition **??** and Proposition **??**, we have fully covered the case $m = 1$ of our problem. Let us now move on to the case $m = 2$. We compute $d_{n,m} = d_{n,2}$

for some small values of n simply by listing all domino tilings of $R_{n,m}$:

n	$d_{n,m}$	domino tilings
0	$d_{0,2} = 1$	
1	$d_{1,2} = 1$	
2	$d_{2,2} = 2$	
3	$d_{3,2} = 3$	
4	$d_{4,2} = 5$	

.

If the $n = 0$ case confuses you, keep in mind that the rectangle $R_{0,2}$ is the empty set (since $R_{0,2} = \underbrace{[0]}_{=\varnothing} \times [2] = \varnothing \times [2] = \varnothing$) and thus has exactly one domino tiling – namely, the tiling that contains no dominos (i.e., the empty set).

Can you find $d_{5,2}$?

Here is a quick way to the answer, at least as far as counting is concerned:

Proposition 1.1.8. For each integer $n \geq 2$, we have $d_{n,2} = d_{n-1,2} + d_{n-2,2}$.

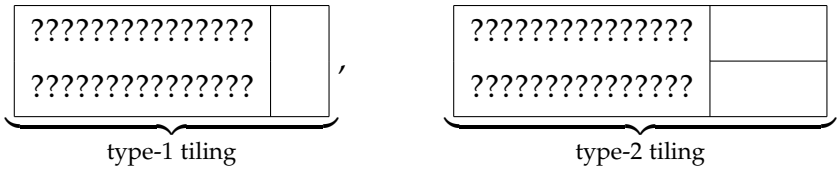
Class of 2019-09-25

Proof sketch. Let $n \geq 2$ be an integer. Consider the last[16] column of $R_{n,2}$ (that is, the set $\{(n,1),(n,2)\}$).

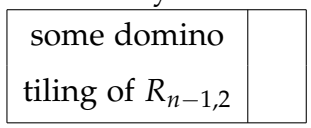
In any domino tiling T of $R_{n,2}$, this last column is **either** covered by 1 vertical domino, **or** covered by (parts of) 2 horizontal dominos.

[16] i.e., easternmost

In the former case, we shall call T a *type-1 tiling*; in the latter case, we shall call T a *type-2 tiling*. Visually, these look as follows:



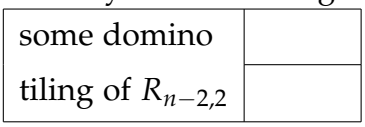
(where the question marks mean an unknown arrangement of dominos).

Let us now analyze type-1 tilings. A type-1 tiling consists of the single vertical domino $\{(n,1),(n,2)\}$ that covers its last column, and a bunch of dominos that cover all the remaining $n-1$ columns. This latter bunch must thus be a domino tiling of $R_{n-1,2}$. Thus, a type-1 tiling consists of the single vertical domino $\{(n,1),(n,2)\}$ and an arbitrary domino tiling of $R_{n-1,2}$. (Visually, this means that it looks as follows: 

.) Hence,[17]

$$\begin{aligned}(\#\text{ of type-1 tilings}) &= (\#\text{ of domino tilings of } R_{n-1,2}) && (5)\\ &= d_{n-1,2} && (6)\end{aligned}$$

(since $d_{n-1,2}$ was defined as the # of domino tilings of $R_{n-1,2}$).

Let us next analyze type-2 tilings. In a type-2 tiling, the last column is covered by (parts of) 2 horizontal dominos. These 2 dominos must extend to the left (because there is no space for them to extend to the right), and thus also cover the second-to-last column. Explicitly speaking, these 2 dominos must be $\{(n-1,1),(n,1)\}$ and $\{(n-1,2),(n,2)\}$. All the other dominos in the tiling must then cover the remaining $n-2$ columns, i.e., must form a domino tiling of $R_{n-2,2}$. Thus, a type-2 tiling consists of the two horizontal dominos $\{(n-1,1),(n,1)\}$ and $\{(n-1,2),(n,2)\}$ and an arbitrary domino tiling of $R_{n-2,2}$. (Visually, this means that it looks as follows: 

.) Hence,

$$\begin{aligned}(\#\text{ of type-2 tilings}) &= (\#\text{ of domino tilings of } R_{n-2,2}) && (7)\\ &= d_{n-2,2} && (8)\end{aligned}$$

(since $d_{n-2,2}$ was defined as the # of domino tilings of $R_{n-2,2}$).

Now, recall that each domino tiling of $R_{n,2}$ is either a type-1 tiling or a type-2 tiling (but cannot be both at the same time). Hence,

$$\begin{aligned}&(\#\text{ of domino tilings of } R_{n,2})\\ &= (\#\text{ of type-1 tilings}) + (\#\text{ of type-2 tilings}) && (9)\\ &= d_{n-1,2} + d_{n-2,2} && (10)\end{aligned}$$

[17] When we say "type-1 tiling", we mean "type-1 tiling of $R_{n,2}$", of course. (The same will apply to "type-2 tiling" later on.)

(by adding the equalities (**??**) and (**??**) together). Now, the definition of $d_{n,2}$ yields

$$d_{n,2} = (\#\text{ of domino tilings of } R_{n,2}) = d_{n-1,2} + d_{n-2,2}$$

(by (**??**)). This proves Proposition **??**. □

Again, let us analyze what we have actually done in this proof:

1. The equality (**??**) follows from the bijection principle. Indeed, our argument for it boils down to the (easily established) fact that there is a bijection

   $$f : \{\text{domino tilings of } R_{n-1,2}\} \to \{\text{type-1 tilings}\}$$

   (which takes any domino tiling of $R_{n-1,2}$, and adds the vertical domino $\{(n,1),(n,2)\}$ to it). Once you have convinced yourself of this fact, you can apply Theorem **??** to $X = \{\text{domino tilings of } R_{n-1,2}\}$ and $Y = \{\text{type-1 tilings of } R_{n,2}\}$, and conclude that

   $$|\{\text{domino tilings of } R_{n-1,2}\}| = |\{\text{type-1 tilings}\}|.$$

   In other words, (# of domino tilings of $R_{n-1,2}$) = (# of type-1 tilings). Thus, the equality (**??**) is proven. The equality (**??**) is obtained similarly.

2. The equality (**??**) follows from the sum rule. Indeed, the sets {type-1 tilings} and {type-2 tilings} are disjoint, and their union is the set {domino tilings of $R_{n,2}$}. Hence, Theorem **??** (applied to $S = \{\text{domino tilings of } R_{n,2}\}$, $k = 2$, $S_1 = \{\text{type-1 tilings}\}$ and $S_2 = \{\text{type-2 tilings}\}$) yields

   $$|\{\text{domino tilings of } R_{n,2}\}| = |\{\text{type-1 tilings}\}| + |\{\text{type-2 tilings}\}|.$$

   In other words, (# of domino tilings of $R_{n,2}$) = (# of type-1 tilings) + (# of type-2 tilings). This proves (**??**).

Proposition **??** lets us compute the numbers $d_{n,2}$ rather easily, if we compute them in the appropriate order (i.e., start with $d_{0,2}$ and $d_{1,2}$, then compute $d_{2,2}$, then compute $d_{3,2}$, then compute $d_{4,2}$, and so on). For example, we get

$$\begin{aligned}
d_{5,2} &= \underbrace{d_{4,2}}_{=5} + \underbrace{d_{3,2}}_{=3} = 5+3 = 8; \\
d_{6,2} &= \underbrace{d_{5,2}}_{=8} + \underbrace{d_{4,2}}_{=5} = 8+5 = 13; \\
d_{7,2} &= \underbrace{d_{6,2}}_{=13} + \underbrace{d_{5,2}}_{=8} = 13+8 = 21; \\
&\cdots.
\end{aligned}$$

But what if we want to compute (say) $d_{900,2}$ without having to first compute all the previous numbers $d_{0,2}, d_{1,2}, \ldots, d_{899,2}$? Is there an explicit formula?

Before we answer this question, let us forget for a moment about domino tilings, and define the sequence of integers:

Definition 1.1.9. The *Fibonacci sequence* is the sequence $(f_0, f_1, f_2, \ldots)$ of nonnegative integers defined recursively by

$$f_0 = 0, \qquad f_1 = 1, \qquad \text{and} \qquad f_n = f_{n-1} + f_{n-2} \text{ for all } n \geq 2.$$

This is a *recursive definition* – i.e., it tells us how to compute f_n assuming that the previous entries $f_0, f_1, \ldots, f_{n-1}$ of the sequence are already known. Thus, if we want to compute f_5 using this definition, we have to compute f_0, f_1, f_2, f_3, f_4 first. (Let us do this: The definition yields $f_0 = 0$ and $f_1 = 1$ immediately. Furthermore, setting $n = 2$ in the equality $f_n = f_{n-1} + f_{n-2}$, we obtain $f_2 = \underbrace{f_1}_{=1} + \underbrace{f_0}_{=0} = 1 + 0 = 1$. Next, setting $n = 3$ in the equality $f_n = f_{n-1} + f_{n-2}$, we obtain $f_3 = \underbrace{f_2}_{=1} + \underbrace{f_1}_{=1} = 1 + 1 = 2$. Likewise, $f_4 = \underbrace{f_3}_{=2} + \underbrace{f_2}_{=1} = 2 + 1 = 3$. Likewise, $f_5 = \underbrace{f_4}_{=3} + \underbrace{f_3}_{=2} = 5$.)

Here is a little table of Fibonacci numbers (this is how the entries f_n of the Fibonacci sequence are called):

n	0	1	2	3	4	5	6	7	8	9	$\cdots$
f_n	0	1	1	2	3	5	8	13	21	34	$\cdots$

.

The Fibonacci sequence is famous – just look at its Wikipedia page! It also has several books dedicated to it, such as Vorobiev's [**?**] (although, to be fully honest, Vorobiev often uses it as a plug to pivot to other mathematics); there is also a journal called *The Fibonacci Quarterly* (again, however, its actual scope is broader). Now, we can reduce our problem of computing $d_{n,2}$ to the problem of computing Fibonacci numbers:

Proposition 1.1.10. We have $d_{n,2} = f_{n+1}$ for each $n \in \mathbb{N}$.

Proof. Here is the main idea of the proof: We must show that the two sequences $(d_{0,2}, d_{1,2}, d_{2,2}, d_{3,2}, \ldots)$ and $(f_1, f_2, f_3, f_4, \ldots)$ are identical. Both of them have the property that their first entries are 1's (that is, $d_{0,2} = 1$ and $f_1 = 1$), their second entries are 1's (that is, $d_{1,2} = 1$ and $f_2 = 1$), and each of their further entries equals the sum of the preceding two entries (because Proposition **??** shows that the $d_{n,2}$ satisfy $d_{n,2} = d_{n-1,2} + d_{n-2,2}$, whereas the definition of Fibonacci numbers shows that $f_{n+1} = f_n + f_{n-1}$). Thus, to put it in practical terms: Both sequences start with the same two entries, and then are built out of these two entries according to the same rule (namely, each further entry is the sum of the preceding two entries). Hence, the two sequences must be the same. This proves Proposition **??**.

If you find this insufficiently rigorous, here is a formal version of this argument:

We shall prove Proposition **??** by strong induction on n.

A strong induction needs no induction base[18]. Thus, we only do the induction step:

Induction step: Let $m \in \mathbb{N}$. Assume (as the induction hypothesis) that Proposition **??** holds for each $n < m$. We must prove that Proposition **??** holds for $n = m$.

Our induction hypothesis says that Proposition **??** holds for each $n < m$. In other words, we have

$$d_{n,2} = f_{n+1} \qquad \text{for each } n \in \mathbb{N} \text{ satisfying } n < m. \tag{11}$$

Now, we must prove that Proposition **??** holds for $n = m$. In other words, we must prove that $d_{m,2} = f_{m+1}$. If $m = 0$, then this is true (since $d_{0,2} = 1 = f_1 = f_{0+1}$). If $m = 1$, then this is also true (since $d_{1,2} = 1 = f_2 = f_{1+1}$). Hence, it remains to prove this in the case $m \geq 2$. So let us WLOG[19] assume that $m \geq 2$. Then, $m - 2 \in \mathbb{N}$. Hence, we can apply (**??**) to $n = m - 2$ (since $m - 2 < m$), and obtain $d_{m-2,2} = f_{(m-2)+1} = f_{m-1}$. Furthermore, $m - 1 \in \mathbb{N}$ (since $m \geq 2 \geq 1$). Thus, we can apply (**??**) to $n = m - 1$ (since $m - 1 < m$), and obtain $d_{m-1,2} = f_{(m-1)+1} = f_m$. However, Proposition **??** (applied to $n = m$) shows that

$$d_{m,2} = \underbrace{d_{m-1,2}}_{=f_{m-1}} + \underbrace{d_{m-2,2}}_{=f_{m-2}} = f_{m-1} + f_{m-2}.$$

Comparing this with

$$f_m = f_{m-1} + f_{m-2} \qquad \text{(by the definition of the Fibonacci sequence)},$$

we obtain $d_{m,2} = f_m$. In other words, Proposition **??** holds for $n = m$. This completes the induction step. Thus, Proposition **??** is proven. □

So we have identified our numbers $d_{n,2}$ as the famous Fibonacci numbers f_{n+1}. Does this help us compute them directly? Yes, because there is a famous formula for the Fibonacci numbers:

Theorem 1.1.11 (Binet's formula)**.** For each $n \in \mathbb{N}$, we have

$$f_n = \frac{1}{\sqrt{5}} \left(\varphi^n - \psi^n \right),$$

where

$$\varphi = \frac{1 + \sqrt{5}}{2} \approx 1.618\ldots \qquad \text{and} \qquad \psi = \frac{1 - \sqrt{5}}{2} \approx -0.618\ldots.$$

A few words about this strange formula are in order. On its left hand side is a nonnegative integer, f_n. On its right side is an expression involving irrational numbers like $\sqrt{5}$ as well as minus signs. How could an explicit formula for a sequence of nonnegative integers require irrational numbers and subtraction?

Well, this is the price of asking for explicit formulas!

[18] See [**?**, §2.8] for how strong induction works.

[19] "WLOG" means "without loss of generality". We can assume that $m \geq 2$ without loss of generality, since we have already proven our claim (that $d_{m,2} = f_{m+2}$) in all other cases.

The numbers φ and ψ in Theorem **??** are known as the *golden ratios* (although usually only φ is considered "the golden ratio"). They are the two roots of the quadratic polynomial $x^2 - x - 1$, so they satisfy $\varphi^2 = \varphi + 1$ and $\psi^2 = \psi + 1$. If this looks like the Fibonacci recursion $f_n = f_{n-1} + f_{n-2}$, don't be surprised! This is the reason why they show up in the explicit formula for the Fibonacci numbers.

How do you compute f_{900} using Theorem **??**? You may be tempted to just plug $n = 900$ into the formula using your favorite computer algebra system; but there is a subtlety involved: Since $\sqrt{5}$ is irrational, the computer may try to work with approximate values, and then when you take n-th powers, the rounding errors will blow up. You will probably not get an integer as a result, and even if you try to round it to the nearest integer, you may well get the wrong value! Such is the price of blindly trusting floating-point arithmetic. Fortunately, $\sqrt{5}$ is an *algebraic number* (i.e., a root of a polynomial with rational coefficients), and this means that it **is** possible to make **exact** computations with it. You just need to restrict yourself to the "$\sqrt{5}$-rationals" (i.e., the numbers of the form $a + b\sqrt{5}$ with $a, b \in \mathbb{Q}$), and instead of approximating them with decimals, you just keep them in the $a + b\sqrt{5}$ form. It is easy to find rules for adding, subtracting, multiplying and dividing $\sqrt{5}$-rationals by one another[20]; thus you don't need approximate values. Using the "exponentiation by squaring" trick, it is now easy to compute very high powers of φ and ψ, and then Theorem **??** yields f_n. For example, a computer will readily tell you all the 188 digits of f_{900}, the last six of which are 938800; you would probably have gotten these wrong if you relied on approximate computation.

Theorem **??** also gives a very easy answer to the question (which we arguably haven't asked) how fast the Fibonacci numbers f_n grow when n gets large. Indeed, $|\psi| < 1$, so that $\psi^n \to 0$ when $n \to \infty$. Hence, for high enough n, we have $f_n \approx \dfrac{1}{\sqrt{5}}\varphi^n$. Of course, the sequence of $\dfrac{1}{\sqrt{5}}\varphi^n$ for $n \in \mathbb{N}$ is a geometric sequence with ratio $\varphi \approx 1.618\ldots$, and grows exponentially. Thus, we see that the Fibonacci numbers f_n grow exponentially – slower than the powers of 2 (since $1.618\ldots < 2$), but still faster than any polynomial. The Fibonacci number f_n will have $\approx \left(\log_{10} \varphi\right) \cdot n \approx 0.209 \cdot n$ many digits.

Theorem **??** is easy to prove:

Proof of Theorem **??** *(sketched).* This can be proven by the same argument that we used for Proposition **??**: We have to show that the sequences $(f_0, f_1, f_2, \ldots)$ and

[20] To wit:

$$\left(a + b\sqrt{5}\right) + \left(c + d\sqrt{5}\right) = (a + c) + (b + d)\sqrt{5};$$
$$\left(a + b\sqrt{5}\right) - \left(c + d\sqrt{5}\right) = (a - c) + (b - d)\sqrt{5};$$
$$\left(a + b\sqrt{5}\right)\left(c + d\sqrt{5}\right) = (ac + 5bd) + (ad + bc)\sqrt{5};$$
$$\frac{a + b\sqrt{5}}{c + d\sqrt{5}} = \frac{\left(a + b\sqrt{5}\right)\left(c - d\sqrt{5}\right)}{\left(c + d\sqrt{5}\right)\left(c - d\sqrt{5}\right)} = \frac{(ac - 5bd) + (bc - ad)\sqrt{5}}{c^2 - 5d^2}.$$

Note in particular the last equation: This is why they taught you to rationalize denominators in high school!

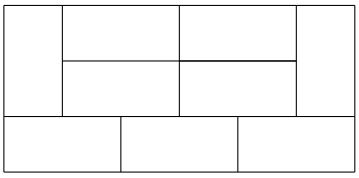
$\left(\frac{1}{\sqrt{5}} \left(\varphi^0 - \psi^0 \right), \frac{1}{\sqrt{5}} \left(\varphi^1 - \psi^1 \right), \frac{1}{\sqrt{5}} \left(\varphi^2 - \psi^2 \right), \ldots \right)$ are identical. Both of them have the property that their first entries are 0's (this is easy to check), their second entries are 1's (this is easy to check), and each of their further entries equals the sum of the preceding two entries[21]. Thus, both sequences start with the same two entries, and then are built out of these two entries according to the same rule. Hence, the two sequences must be the same. This proves Theorem **??**. □

This proof should convince you that Theorem **??** holds, but does not explain how you could have come up with Theorem **??**. If time allows, we will later explain this when we explore the concept of *generating functions*.

1.1.6. Kasteleyn's formula (teaser)

Now we have computed $d_{n,2}$ for each $n \in \mathbb{N}$. What about $d_{n,3}$?

Proposition **??** shows that $d_{n,3} = 0$ when n is odd. But computing $d_{n,3}$ when n is even is a lot harder. You might try to find a recursive formula such as Proposition **??**, but this isn't so easy any more. You can still try to separate the domino tilings of $R_{n,3}$ into types according to how the last column looks like; however, there will be three of these types now, and two of them will not fall into a 1-to-1 correspondence with domino tilings of a smaller rectangle. For example, consider the following domino tiling of $R_{6,3}$:



.

There is no tiling of $R_{5,3}$ anywhere in it, nor of $R_{4,3}$, nor of $R_{3,3}$, nor of $R_{2,3}$, nor of $R_{1,3}$. This thwarts our recursive approach.

Nevertheless, there **is** a nice recursion for $d_{n,3}$:

Proposition 1.1.12. We have $d_{n,3} = 4d_{n-2,3} - d_{n-4,3}$ for each $n \geq 4$.

As I said, there is no proof as easy as the one we gave for Proposition **??**. We will later learn the technique of *generating functions*, which can be used to give

[21] Indeed, for the Fibonacci sequence $(f_0, f_1, f_2, \ldots)$, this is clear. For the second sequence, this boils down to proving the identity

$$\frac{1}{\sqrt{5}} \left(\varphi^n - \psi^n \right) = \frac{1}{\sqrt{5}} \left(\varphi^{n-1} - \psi^{n-1} \right) + \frac{1}{\sqrt{5}} \left(\varphi^{n-2} - \psi^{n-2} \right),$$

which however follows by subtracting the two easily verified identities

$$\varphi^n = \varphi^{n-1} + \varphi^{n-2} \qquad \text{and} \qquad \psi^n = \psi^{n-1} + \psi^{n-2}$$

and dividing the result by $\sqrt{5}$.

a reasonably simple proof (see [**?**]), and probably a more complicated induction could also do the trick.

What about $d_{n,4}$? There is a recursion, too, according to [**?**, §2]:

$$d_{n,4} = d_{n-1,4} + 5d_{n-2,4} + d_{n-3,4} - d_{n-4,4}.$$

As you see, these are getting more complicated. In theory, recurrence relations like these have explicit formulas like Binet's formula for f_n (Theorem **??**). However, these formulas become more and more complicated as well, and in particular they no longer involve "nice" irrational numbers like $\sqrt{5}$, but rather roots of higher-degree polynomials, which at some point can no longer be expressed using rational numbers, sums, differences, products, quotients and radicals (i.e., $\sqrt[a]{b}$ terms)[22]. This looks like a dead end.

In the 20th Century, however, theoretical physicists studying thermodynamics got interested in computing $d_{n,m}$, as they considered a domino tiling to be a (rather idealized) model for a liquid consisting of "dimers" (polymers that take up two adjacent sites in a rectangular lattice; these are exactly our dominos). Even though the 2-dimensionality of a rectangle makes it a somewhat unrealistic approximation for real-world liquids, it found its use as a model for the adsorption of molecules on a surface (see [**?**]). In 1961, Kasteleyn found the following surprising formula for $d_{n,m}$:

Theorem 1.1.13 (Kasteleyn's formula)**.** Assume that m is even and $n \geq 1$. Then,

$$d_{n,m} = 2^{mn/2} \prod_{j=1}^{m/2} \prod_{k=1}^{n} \sqrt{\left(\cos\frac{j\pi}{m+1}\right)^2 + \left(\cos\frac{k\pi}{n+1}\right)^2}.$$

(Here, we are using the product sign $\prod$: That is, if $a_1, a_2, \ldots, a_p$ are any numbers, then $\prod_{i=1}^{p} a_i$ means the product $a_1 a_2 \cdots a_p$. The presence of two product signs directly following one another means that we are taking a product of products.)

Actually, "surprising" is an understatement; why cosines of angles would appear in a formula for the integer $d_{n,m}$ is even less transparent than why $\sqrt{5}$ should appear in a formula for Fibonacci numbers!

We won't even get close to proving Theorem **??**. A proof outline appears in [**?**, Theorem 12.85], serving as a culmination of a graduate-level combinatorics textbook. A more self-contained exposition of the proof has been given by Stucky

[22] The roots of a quadratic polynomial $x^2 + ax + b$ can be expressed in this way: They are $\frac{-a \pm \sqrt{a^2 - 4b}}{2}$. The roots of a degree-3 or degree-4 polynomial can also be expressed in this way, using complicated formulas due to Tartaglia, Ferrari, Cardano and Descartes. But the *Abel–Ruffini theorem* says that for $k \geq 5$, there is no formula that expresses the roots of a (general) degree-k polynomial using only $+$, $-$, $\cdot$, $/$ and $\sqrt{}$.

in [**?**], but even that is only self-contained up to some advanced linear algebra (Pfaffians and Kronecker products of matrices, as well as a good understanding of eigenvectors are required).

For all its seeming extravagance, Theorem **??** actually provides a good way of computing $d_{n,m}$. Indeed, cosines like $\cos\dfrac{j\pi}{m+1}$ and $\cos\dfrac{k\pi}{n+1}$ are algebraic numbers that lend themselves to exact computation (using cyclotomic polynomials – a piece of abstract algebra we are also not coming close to), and the scary-looking products don't scare a good computer algebra system. For example, Kasteleyn's formula can be used to show that $d_{8,8} = 12\ 988\ 816$. This wouldn't be so easy to check by a brute force search for domino tilings!

Class of 2019-09-27

1.1.7. Axisymmetric domino tilings

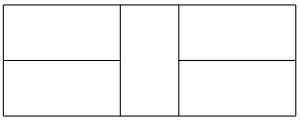
Let us solve a few more counting exercises around domino tilings.

Exercise 1.1.1. Let $n \in \mathbb{N}$. Say that a domino tiling T of $R_{n,2}$ is *axisymmetric* if reflecting it across the vertical axis of symmetry of $R_{n,2}$ leaves it unchanged (i.e., for each domino $\{(i,j),(i',j')\} \in T$, the "mirror domino" $\{(n+1-i,j),(n+1-i',j')\}$ also belongs to T).

For example, the tilings

 and 

are not axisymmetric (indeed, reflecting them across the vertical line transforms them into one another, and they are not the same), but the tilings

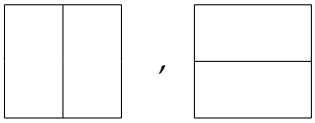
 and 

are axisymmetric.

How many axisymmetric domino tilings does $R_{n,2}$ have?

Example 1.1.14. Let us list the axisymmetric domino tilings for $R_{n,2}$ when n is

small:

n	axisymmetric domino tilings	their number
0		1
1		2
2		1
3		3
4		2

Solution sketch to Exercise **??**. Let us only show the main steps; a (more) detailed solution to Exercise **??** can be found in [**?**, Exercise 5].

There are two cases to consider: the case when n is even, and the case when n is odd.

Let us first consider the case when n is even. In this case, I claim that any axisymmetric domino tiling of $R_{n,2}$ has one of the following two forms:

- **Form 1:** a domino tiling J of $R_{n/2,2}$ covering the left half of $R_{n,2}$, and its mirror image across the vertical axis covering the right half:

domino tiling J of $R_{n/2,2}$	mirror image of J

;

(Note that the vertical axis cuts through the middle of this picture.)

- **Form 2:** a domino tiling J of $R_{n/2-1,2}$ covering the leftmost $n/2-1$ columns of $R_{n,2}$, and two horizontal dominos covering the $(n/2)$-th and $(n/2+1)$-th columns[23], and the mirror image of J across the vertical axis covering the

[23] We count columns from the left.

rightmost $n/2-1$ columns of $R_{n,2}$:

.

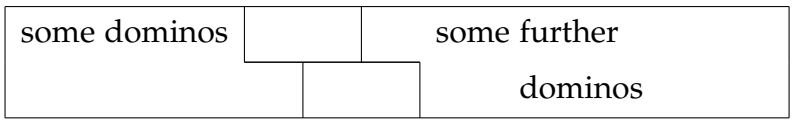
(Again, the vertical axis cuts through the middle of this picture.)

If you agree with me that these are the only possible forms of an axisymmetric domino tiling of $R_{n,2}$, then it follows (by the same logic as in the proof of Proposition **??**) that the # of axisymmetric domino tilings of $R_{n,2}$ is

$$\underbrace{(\#\text{ of domino tilings of } R_{n/2,2})}_{\substack{=d_{n/2,2}=f_{n/2+1} \\ \text{(by Proposition ??, applied to } n/2 \\ \text{instead of } n)}} + \underbrace{(\#\text{ of domino tilings of } R_{n/2-1,2})}_{\substack{=d_{n/2-1,2}=f_{n/2} \\ \text{(by Proposition ??, applied to } n/2-1 \\ \text{instead of } n)}}$$
$$= f_{n/2+1} + f_{n/2} = f_{n/2+2}$$

(by the recursive definition of the Fibonacci numbers). But why are the above-mentioned two forms the only possible forms of an axisymmetric domino tiling of $R_{n,2}$?

To prove this, we fix an axisymmetric domino tiling T of $R_{n,2}$. We must show that T has one of the above two forms. We can consider what dominos cover the $(n/2)$-th column in T. If this column is covered by one vertical domino, then no domino of T straddles the vertical axis, and thus the tiling T has Form 1 (indeed, the part of T to the right of the vertical axis is a mirror image of the part to the left, since T is axisymmetric). It remains to deal with the case when the $(n/2)$-th column is covered by two horizontal dominos in T. The question is now where these horizontal dominos fall. If both of them fall into columns $n/2-1$ and $n/2$, then our tiling T has Form 1. If both of them fall into columns $n/2$ and $n/2+1$, then our tiling T has Form 2. What about the remaining "rogue" possibility, that one of them falls into columns $n/2-1$ and $n/2$, while the other falls into columns $n/2$ and $n/2+1$? Here is a picture (which assumes that the former domino is in the top row and the latter domino is in the bottom row; but the argument will be the same in the opposite case):



.

Note, however, that the part marked "some dominos" in this picture has an odd number of squares (namely, $2(n/2-1)+1 = n-1$ many squares), so that it cannot actually be covered by dominos![24] Thus, we obtain a contradiction. Hence, this "rogue" possibility is actually impossible. This shows that every axisymmetric domino tiling T of $R_{n,2}$ either has Form 1 or has Form 2.

[24]This is the same logic that we used to prove Proposition **??**.

Thus, we have proved that the # of axisymmetric domino tilings of $R_{n,2}$ is $f_{n/2+2}$ in the case when n is even.

What about the case when n is odd? In this case, I claim that any axisymmetric domino tiling of $R_{n,2}$ has the following form:

- **Form 1:** a domino tiling J of $R_{(n-1)/2,2}$ covering the leftmost $(n-1)/2$ columns of $R_{n,2}$, and one vertical domino covering the $(n+1)/2$-th column, and the mirror image of J across the vertical axis covering the rightmost $(n-1)/2$ columns of $R_{n,2}$:

domino tiling J of $R_{(n-1)/2,2}$		mirror image of J

.

(Again, the vertical axis cuts through the middle of this picture, which in this case means cutting through the middle of the vertical domino.)

This time, there is no Form 2. Again, we need to prove that this is the only possible form. This is even easier than in the previous case; the main idea is that if an axisymmetric tiling T of $R_{n,2}$ contained any horizontal domino intersect with its $(n+1)/2$-th column, then it would also contain the mirror image of this tile across the vertical axis (since T is axisymmetric), but then this tile and its mirror image would be distinct but overlapping[25], which would contradict the definition of a domino tiling.

Thus, when n is odd, the # of axisymmetric domino tilings of $R_{n,2}$ equals the # of domino tilings of $R_{(n-1)/2,2}$, which (by Proposition **??**) is $f_{(n-1)/2+1} = f_{(n+1)/2}$.

Thus, the general formula for the # of axisymmetric domino tilings of $R_{n,2}$ is:

$$(\#\text{ of axisymmetric domino tilings of } R_{n,2}) = \begin{cases} f_{n/2+2}, & \text{if } n \text{ is even;} \\ f_{(n+1)/2}, & \text{if } n \text{ is odd} \end{cases}.$$

□

1.1.8. Tiling rectangles with k-bricks

Now, let us look at another related problem.

For the rest of Subsection **??**, fix a positive integer k.

A *k-brick* shall mean a $1 \times k$-rectangle or a $k \times 1$-rectangle. More specifically: A *vertical k-brick* shall mean a $1 \times k$-rectangle (i.e., a set of the form

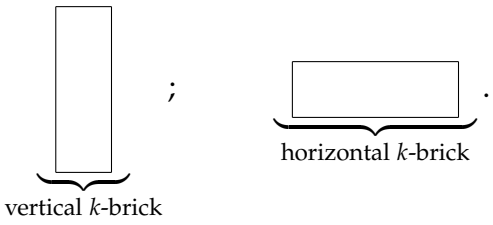
$$\{(i,j), (i,j+1), (i,j+2), \ldots, (i,j+k-1)\}$$

for some $i, j \in \mathbb{Z}$); a *horizontal k-brick* shall mean a $k \times 1$-rectangle (i.e., a set of the form

$$\{(i,j), (i+1,j), (i+2,j), \ldots, (i+k-1,j)\}$$

[25] They would overlap in the $(n+1)/2$-th column.

for some $i, j \in \mathbb{Z}$). Here is how they look like:



If S is a set of squares, then a *k-brick tiling* of S means a way to cover the set S with non-overlapping k-bricks (i.e., formally speaking: a set of disjoint k-bricks whose union is S).

Thus, k-brick tilings are a generalization of domino tilings. More specifically: If $k = 2$, then k-bricks are the same as dominos, and thus k-brick tilings are the same as domino tilings.

When does a rectangle $R_{n,m}$ have a k-brick tiling? The answer is surprisingly simple:

Proposition 1.1.15. Let $n, m \in \mathbb{N}$, and let k be a positive integer. Then, the rectangle $R_{n,m}$ has a k-brick tiling if and only if we have $k \mid m$ or $k \mid n$.

In the case when $k = 2$, Proposition **??** says that the rectangle $R_{n,m}$ has a domino tiling if and only if we have $2 \mid m$ or $2 \mid n$ (that is, at least one of the numbers m and n is even). This should not be surprising: One direction of this equivalence (namely, the "$\Longrightarrow$" direction: i.e., the direction saying that if $R_{n,m}$ has a domino tiling, then at least one of m and n is even) follows from Proposition **??**, whereas the other direction is easy. We could use a similar argument to prove Proposition **??** whenever the number k is prime. Indeed, our proof of Proposition **??** generalizes to show that if the rectangle $R_{n,m}$ has a k-brick tiling, then $k \mid mn$. When k is prime, the divisibility $k \mid mn$ implies that $k \mid m$ or $k \mid n$; thus, the "$\Longrightarrow$" direction of Proposition **??** would follow immediately in this case. But when k is prime, we need a better argument.

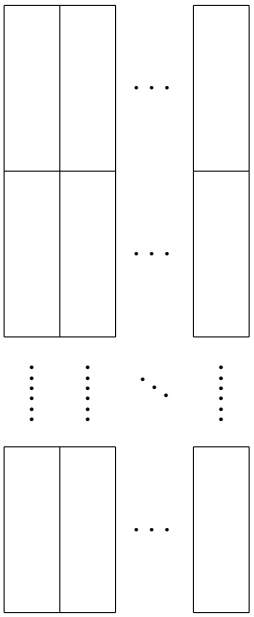
Proof of Proposition **??**. The claim we want to prove is an "if and only if" claim, so it has two directions: the "$\Longleftarrow$" direction (also known as the "if" direction), and the "$\Longrightarrow$" direction (also known as the "only if" direction). The former direction claims that if we have $k \mid m$ or $k \mid n$, then the rectangle $R_{n,m}$ has a k-brick tiling. The latter claims the converse of this statement.

We shall prove these two directions separately:

$\Longleftarrow$: Assume that $k \mid m$ or $k \mid n$.

If $k \mid m$, then $R_{n,m}$ has a k-brick tiling consisting entirely of vertical k-bricks. It

looks as follows:



(with each column being covered by m/k many vertical k-bricks). Similarly, if $k \mid n$, then $R_{n,m}$ has a k-brick tiling consisting entirely of horizontal k-bricks. Thus, in either case, $R_{n,m}$ has a k-brick tiling. This proves the "$\Longleftarrow$" direction of Proposition **??**.

$\Longrightarrow$: Assume that the rectangle $R_{n,m}$ has a k-brick tiling. We must prove that we have $k \mid m$ or $k \mid n$.

Assume the contrary (for the sake of contradiction). Thus, $k \nmid m$ and $k \nmid n$.

We shall use the following notations from elementary number theory: If a is an integer and b is a positive integer, then

- we let $a \,//\, b$ denote the quotient obtained when dividing a by b (in the sense of division with remainder);
- we let $a\%b$ denote the remainder obtained when dividing a by b.

Both of these numbers $a \,//\, b$ and $a\%b$ are integers, and they satisfy $0 \leq a\%b < b$ and $a = (a \,//\, b) \cdot b + (a\%b)$.

Let $r = n\%k$ and $s = m\%k$. We have $r = n\%k \neq 0$ (since $k \nmid n$) and thus $0 < r < k$ (since r is a remainder upon division by k). Likewise, $0 < s < k$.

Now, we are going to color the squares of $R_{n,m}$ with k colors. The k colors we are going to use will be numbered $0, 1, \ldots, k-1$. Each square $(i, j) \in R_{n,m}$ will be colored with the color $(i + j - 2)\,\%k$. Here is how this coloring looks like (in the

example where $k = 4$, $n = 10$ and $m = 7$):

2	3	0	1	2	3	0	1	2	3
1	2	3	0	1	2	3	0	1	2
0	1	2	3	0	1	2	3	0	1
3	0	1	2	3	0	1	2	3	0
2	3	0	1	2	3	0	1	2	3
1	2	3	0	1	2	3	0	1	2
0	1	2	3	0	1	2	3	0	1

(where we have written the color of each square as a number into this square). Note that if $k = 2$, then there are two colors only, numbered 0 and 1; in this case, our coloring is precisely the usual chessboard coloring (if we regard color 0 as black and color 1 as white).

Let us make a few observation about how our coloring behaves:

- The southwesternmost square $R_{n,m}$ is $(1, 1)$, and thus has color $(1 + 1 - 2) \,\%2 = 0\%2 = 0$.
- As we move eastwards, the colors of the squares increase by 1 at each step, until they reach $k - 1$, at which point they "fall back down" to 0 at the next step.
- The same happens as we move northwards.
- Along each "northwest-to-southeast" diagonal (i.e., each line with slope -1), the color stays constant.

Let us say that a finite set S of squares is *balanced* if it has equally many squares of each color. In other words, a finite set S of squares is *balanced* if and only if for any color $h \in \{0, 1, \ldots, k-1\}$, the # of all squares in S that have color h does not depend on h.

Each horizontal k-brick has exactly 1 square of each color. Indeed, the colors of the k squares in a horizontal k-brick look as follows:

u	$u+1$	$\cdots$	$k-1$	0	1	$\cdots$	$u-1$

(where u is the color of the leftmost square of the k-brik). Thus, each horizontal k-brick is balanced. Likewise, each vertical k-brick is balanced. Thus, we have shown that each k-brick is balanced.

But we have assumed that the rectangle $R_{n,m}$ has a k-brick tiling. Hence, $R_{n,m}$ must, too, be balanced (by the sum rule)[26].

[26]Here is the argument in details:

Let us now subdivide the rectangle $R_{n,m}$ into several (disjoint) *zones* $Z_{u,v}$ by cutting it with several lines[27]. Namely, we cut it with horizontal lines every k squares (counted from the bottom), and with vertical lines every k squares (counted from the left). Thus, we obtain the following subdivision of $R_{n,m}$:

$Z_{0,2}$	$Z_{1,2}$	$Z_{2,2}$	$Z_{3,2}$
$Z_{0,1}$	$Z_{1,1}$	$Z_{2,1}$	$Z_{3,1}$
$Z_{0,0}$	$Z_{1,0}$	$Z_{2,0}$	$Z_{3,0}$

.

Formally speaking, our zones $Z_{u,v}$ are defined as follows:

$$Z_{u,v} = \{(i,j) \in R_{n,m} \mid (i-1) \,//\, k = u \text{ and } (j-1) \,//\, k = v\}$$
$$\text{for all } u \in \{0,1,\ldots,(n-1) \,//\, k\} \text{ and } v \in \{0,1,\ldots,(m-1) \,//\, k\}.$$

All these zones $Z_{u,v}$ are rectangles. More precisely: Let us call a zone $Z_{u,v}$ (with $u \in \{0,1,\ldots,(n-1) \,//\, k\}$ and $v \in \{0,1,\ldots,(m-1) \,//\, k\}$)

We have assumed that the rectangle $R_{n,m}$ has a k-brick tiling. Let T be this k-brick tiling. Write T in the form $T = \{S_1, S_2, \ldots, S_j\}$, where $S_1, S_2, \ldots, S_j$ are distinct k-bricks. Hence, each square in $R_{n,m}$ belongs to exactly one of $S_1, S_2, \ldots, S_j$. Now, for each color $h \in \{0,1,\ldots,k-1\}$, we have

$$\begin{aligned}
&(\#\text{ of all squares in } R_{n,m} \text{ that have color } h) \\
&= (\#\text{ of all squares in } R_{n,m} \text{ that belong to } S_1 \text{ and have color } h) \\
&\quad + (\#\text{ of all squares in } R_{n,m} \text{ that belong to } S_2 \text{ and have color } h) \\
&\quad + \cdots \\
&\quad + (\#\text{ of all squares in } R_{n,m} \text{ that belong to } S_j \text{ and have color } h) \\
&\qquad \left(\begin{array}{c} \text{by the sum rule, since each square in } R_{n,m} \\ \text{belongs to exactly one of } S_1, S_2, \ldots, S_j \end{array} \right) \\
&= \sum_{i=1}^{j} \underbrace{(\#\text{ of all squares in } R_{n,m} \text{ that belong to } S_i \text{ and have color } h)}_{\substack{=(\#\text{ of all squares in } S_i \text{ that have color } h) \\ (\text{since } S_i \subseteq R_{n,m})}} \\
&= \sum_{i=1}^{j} \underbrace{(\#\text{ of all squares in } S_i \text{ that have color } h)}_{\substack{\text{independent of } h \\ (\text{since } S_i \text{ is balanced})}}.
\end{aligned}$$

Hence, for each $h \in \{0,1,\ldots,k-1\}$, the # of all squares in $R_{n,m}$ that have color h is independent of h (since it is a sum of j numbers that are each independent of h). In other words, $R_{n,m}$ is balanced.

[27] We will give a rigorous definition further below.

- *generic* if $u < (n-1) \,//\, k$ and $v < (m-1) \,//\, k$;
- *northern* if $u < (n-1) \,//\, k$ and $v = (m-1) \,//\, k$;
- *eastern* if $u = (n-1) \,//\, k$ and $v < (m-1) \,//\, k$;
- *northeastern* if $u = (n-1) \,//\, k$ and $v = (m-1) \,//\, k$.

Then,[28]

- each generic zone is a $k \times k$-rectangle;
- each northern zone is a $k \times s$-rectangle;
- each eastern zone is an $r \times k$-rectangle;
- the northeastern zone is an $r \times s$-rectangle.

Note that all these zones are nonempty, since $r > 0$ and $s > 0$ and $k > 0$.

Each northern zone is a $k \times s$-rectangle, and thus can be tiled with horizontal k-bricks; therefore it is balanced (since each k-brick is balanced). Likewise, each eastern zone is balanced, and each generic zone is balanced. Thus, summarizing, we have shown that each zone Z except for the northeastern zone is balanced. In other words, if Z is a zone that is not the northeastern zone, then, for any color $h \in \{0, 1, \ldots, k-1\}$,

$$\text{the \# of all squares in } Z \text{ that have color } h \text{ is independent of } h. \tag{12}$$

But the zones are disjoint and their union is the whole rectangle $R_{n,m}$. Hence, it follows that the northeastern zone is balanced as well[29].

[28] Recall that $r = n\%k$ and $s = m\%k$.

[29] *Proof.* For any color $h \in \{0, 1, \ldots, k-1\}$, we have

$$\begin{aligned}
&(\text{\# of all squares in } R_{n,m} \text{ that have color } h) \\
&= \sum_{Z \text{ is a zone}} (\text{\# of all squares in } Z \text{ that have color } h) \\
&= (\text{\# of all squares in the northeastern zone that have color } h) \\
&\qquad + \sum_{\substack{Z \text{ is a zone;} \\ Z \text{ is not the northeastern zone}}} (\text{\# of all squares in } Z \text{ that have color } h),
\end{aligned}$$

so that

$$\begin{aligned}
&(\text{\# of all squares in the northeastern zone that have color } h) \\
&= \underbrace{(\text{\# of all squares in } R_{n,m} \text{ that have color } h)}_{\substack{\text{independent of } h \\ (\text{since } R_{n,m} \text{ is balanced})}} \\
&\qquad - \sum_{\substack{Z \text{ is a zone;} \\ Z \text{ is not the northeastern zone}}} \underbrace{(\text{\# of all squares in } Z \text{ that have color } h)}_{\substack{\text{independent of } h \\ (\text{by } (??))}},
\end{aligned}$$

and thus the # of all squares in the northeastern zone that have color h is independent of h. In other words, the northeastern zone is balanced.

But let us take a closer look at the northeastern zone. We denote this zone by $\mathfrak{Z}$. This zone $\mathfrak{Z}$ is an $r \times s$-rectangle, and (just as for any zone) its southwestern corner has color 0. Thus, the colors of its squares look as follows:

$$\begin{array}{|cccccc|}
\hline
s-1 & s & * & \cdots & * & * \\
s-2 & s-1 & s & \cdots & * & * \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
1 & 2 & 3 & \cdots & r-1 & r \\
0 & 1 & 2 & \cdots & r-2 & r-1 \\
\hline
\end{array} \tag{13}$$

(where each asterisk $*$ stands for some entry we don't need to know about).

We shall now show that this northeastern zone is **not** balanced. This will give us a contradiction.

We WLOG assume that $s \leq r$ (since otherwise, the argument is similar). Recall that $0 < r < k$ and $0 < s < k$. From $0 < s$, we obtain $s \geq 1$, so that $s - 1 \leq 0$. Thus, $0 \leq \underbrace{s}_{\leq r} - 1 \leq r - 1$. Therefore, each row of $\mathfrak{Z}$ contains at least one square with color $s - 1$. Since $\mathfrak{Z}$ has s rows, we thus conclude that

$$\mathfrak{Z} \text{ contains at least } s \text{ squares with color } s-1. \tag{14}$$

(We can easily see that $\mathfrak{Z}$ contains exactly s squares with color $s - 1$, but we won't need this.)

On the other hand, the bottommost row of $\mathfrak{Z}$ only contains squares with colors $0, 1, \ldots, r - 1$. Hence, it does not contain a square with color r (since $s - 1 < s \leq r$). But each row of $\mathfrak{Z}$ has width $r < k$, and thus contains **at most** one square with color r (because if it contained two distinct squares with color r, then these two squares would be a distance of $\geq k$ apart from one another, but the row only has width $r < k$). So we know that the zone $\mathfrak{Z}$ has s rows, one of which (the bottommost one) contains no square with color r, while each of the other $s - 1$ rows contains **at most** one square with color r. Hence, altogether,

$$\mathfrak{Z} \text{ contains at most } s-1 \text{ squares with color } r. \tag{15}$$

Comparing this with (**??**), we conclude that the # of squares with color r contained in $\mathfrak{Z}$ is different from the # of squares with color $s - 1$ contained in $\mathfrak{Z}$ (since the former number is $\leq s - 1$, while the latter number is $\geq s$). Hence, $\mathfrak{Z}$ is not balanced. In other words, the northeastern zone is not balanced (since $\mathfrak{Z}$ is the northeastern zone). This contradicts the fact that the northeastern zone is balanced. This contradiction shows that our assumption was false. Hence, the "$\Longrightarrow$" direction of Proposition **??** is proven. $\square$

This was just a little taste of the theory of tilings of discrete (plane) shapes. See the survey [**?**] by Ardila and Stanley for an introduction to the varied questions and ideas of this theory. See also [**?**, Exercises 4 and 5] and [**?**, Exercises 1, 2, 3] for further exercises on counting tilings, and [**?**, §5.1.5] for a neat exercise in proving existence of tilings. Furthermore, the book [**?**] by Benjamin and Quinn provides a lot of applications of tilings to enumerative combinatorics, such as proofs of identities between Fibonacci numbers using their domino-tiling interpretation (Proposition **??**). (See [**?**] for a "best-of" of sorts.)

Class of 2019-09-30

1.2. Sums of powers

1.2.1. The sum $1+2+...+n$

We now switch the subject and recall a famous result, known colloquially as the *"Little Gauss" formula* due to the anecdote of Gauss inventing it in primary school:[30]

Theorem 1.2.1 ("Little Gauss" formula)**.** Let $n \in \mathbb{N}$. Then,

$$1+2+\cdots+n = \frac{n\left(n+1\right)}{2}.$$

Keep in mind that $0 \in \mathbb{N}$ according to our conventions; thus, Theorem **??** applies to $n = 0$. And indeed, Theorem **??** holds for $n = 0$, since an *empty sum* (i.e., a sum that consists of no addends) is defined to be 0, and thus we have

$$1+2+\cdots+n = 1+2+\cdots+0 = (\text{empty sum}) = 0 = \frac{0\cdot(0+1)}{2}.$$

First proof of Theorem **??**. Induction on n. The details are completely straightforward and LTTR.

(The abbreviation "LTTR" stands for "left to the reader". I will usually leave arguments to the reader when they are straightforward or easy variations of arguments shown before.) □

Second proof of Theorem **??**. We observe the following fact: If $a_1, a_2, \ldots, a_n$ are n numbers (say, real numbers), and $b_1, b_2, \ldots, b_n$ are n further numbers, then

$$\begin{aligned}&\left(a_1+a_2+\cdots+a_n\right)+\left(b_1+b_2+\cdots+b_n\right)\\&=\left(a_1+b_1\right)+\left(a_2+b_2\right)+\cdots+\left(a_n+b_n\right).\end{aligned} \tag{16}$$

(Indeed, both sides of this equality are just two ways to add all the $2n$ numbers $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$. This should convince you that they are the same, although formally speaking, this is not a proof. See Subsection **??** below for some references to a formal proof.)

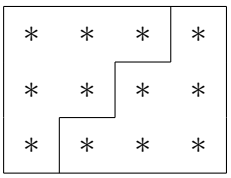
[30] In truth, this formula was known to the Ancient Greeks.

Now,

$$\begin{aligned}
&2 \cdot (1+2+\cdots+n) \\
&= (1+2+\cdots+n) + \underbrace{(1+2+\cdots+n)}_{\substack{=n+(n-1)+\cdots+1 \\ \text{(here, we have just reversed} \\ \text{the order of summation)}}} \\
&= (1+2+\cdots+n) + (n+(n-1)+\cdots+1) \\
&= (1+n) + (2+(n-1)) + \cdots + (n+1) \\
&\qquad \text{(by (??), applied to } a_i = i \text{ and } b_i = n+1-i) \\
&= \underbrace{(n+1)+(n+1)+\cdots+(n+1)}_{n \text{ many addends}} \\
&\qquad \text{(since each of the numbers } 1+n, 2+(n-1), \ldots, n+1 \text{ equals } n+1) \\
&= n(n+1).
\end{aligned}$$

Dividing this equality by 2, we obtain $1+2+\cdots+n = \dfrac{n(n+1)}{2}$. This proves Theorem **??** again. □

Third proof of Theorem **??** *(sketched).* Here is a picture proof (drawn for the case $n = 3$):



.

This is a 4×3-rectangle[31] (i.e., a rectangle of width 4 and height 3), subdivided into two parts by a broken line which starts in the southwestern corner and winds its way to the northeastern corner, making steps of length 1 eastwards and northwards (by turns). The two parts (the one below and the one above the broken line) have the same area, since they are symmetric to each other with respect to the center of the square.[32] Hence, the area of either part equals half the area of the whole rectangle. Since the area of the whole is $3 \cdot 4$, we thus conclude that the area of either part equals $\dfrac{3 \cdot 4}{2}$.

On the other hand, here is a different way to compute this area: Let us look at the part below the broken line. This part has 0 squares in the 1-st column[33], 1 square in the 2-nd column, 2 squares in the 3-rd column, and 3 squares in the 4-th

[31] We have put an asterisk into each little square of the rectangle in order to make the squares easier to discern.

[32] Note that "area" is normally a geometric concept, but since both parts consist of full squares, we can redefine it combinatorially as the number of squares in the respective part.

[33] We count columns from the left.

column. Hence, in total, it has $0+1+2+3$ many squares. In other words, its area is $0+1+2+3$.

Now we know that the area of the part of the rectangle below the broken line is $0+1+2+3$, but we also know (from before) that it equals $\dfrac{3\cdot 4}{2}$. Hence, $0+1+2+3=\dfrac{3\cdot 4}{2}$. In view of $0+1+2+3=1+2+3$, this rewrites as $=\dfrac{3\cdot 4}{2}$. This is precisely the statement of Theorem **??** for $n=3$.

The same argument (but using an $(n+1)\times n$-rectangle instead of a 4×3-rectangle) proves Theorem **??** for arbitrary n.

How can we formalize this argument? Once again (as with domino tilings), it makes sense to think of the $(n+1)\times n$-rectangle as the finite set $[n+1]\times[n]$ (where, as before, we set $[k]=\{1,2,\ldots,k\}$ for each $k\in\mathbb{N}$) rather than a geometric shape in the real plane.[34] The two parts into which this rectangle is divided by the broken line become the two sets

$$A := \{(i,j)\in[n+1]\times[n] \mid i\leq j\} \qquad \text{and}$$
$$B := \{(i,j)\in[n+1]\times[n] \mid i>j\}.$$

(The letters A and B here stand for "above" and "below".)

So let us redo this proof from scratch, using the combinatorial model (i.e., finite sets) throughout it. We consider the finite set $[n+1]\times[n]$, which has $(n+1)\cdot n = n(n+1)$ many elements. Define two subsets

$$A := \{(i,j)\in[n+1]\times[n] \mid i\leq j\} \qquad \text{and}$$
$$B := \{(i,j)\in[n+1]\times[n] \mid i>j\}$$

of $[n+1]\times[n]$. These two subsets A and B are disjoint (since no $(i,j)\in[n+1]\times[n]$ can satisfy $i\leq j$ and $i>j$ at the same time), and their union is $[n+1]\times[n]$ (since each $(i,j)\in[n+1]\times[n]$ satisfies either $i\leq j$ or $i>j$). Hence, the sum rule shows that $|[n+1]\times[n]| = |A|+|B|$, so that

$$\begin{aligned}|A|+|B| &= |[n+1]\times[n]| = \underbrace{|[n+1]|}_{=n+1}\cdot\underbrace{|[n]|}_{=n} \qquad \text{(by Theorem **??**)}\\ &= (n+1)\cdot n = n(n+1). \end{aligned} \tag{17}$$

On the other hand, the map

$$\begin{aligned} A &\to B,\\ (i,j) &\mapsto (n+2-i, n+1-j)\end{aligned}$$

(which, visually, is just the reflection around the center of the $(n+1)\times n$-rectangle) is a bijection[35]. Hence, the bijection rule yields $|A|=|B|$. Thus, $\underbrace{|A|}_{=|B|}+|B| = |B|+|B| = 2\cdot|B|$. Comparing this with (**??**), we obtain $2\cdot|B| = n(n+1)$, so that

$$|B| = \frac{n(n+1)}{2}. \tag{18}$$

[34] Again, we let (i,j) be the square in column i (counted from the left) and row j (counted from the bottom).

[35] This needs to be proven, but the proof is really straightforward. First, you need to check that this

On the other hand, let us count the squares in B "by column". Recall that, in our combinatorial model, the x-coordinate i of a square (i,j) tells us which column it belongs to. Thus, the squares $c \in B$ in the k-th column (for any given $k \in \{1,2,\ldots,n+1\}$) are precisely the squares $c \in B$ that have x-coordinate k. Hence, we should count the squares in B according to their x-coordinate. Formally speaking, this means applying the sum rule to the set B and its $n+1$ disjoint subsets $\{c \in B \mid c \text{ has x-coordinate } k\}$ for $k \in \{1,2,\ldots,n+1\}$. These $n+1$ subsets are disjoint (since any square $c \in B$ has only one x-coordinate) and their union is B (since each square $c \in B$ has x-coordinate 1 or 2 or $\cdots$ or $n+1$). Hence, the sum rule yields

$$\begin{aligned} |B| &= |\{c \in B \mid c \text{ has x-coordinate } 1\}| + |\{c \in B \mid c \text{ has x-coordinate } 2\}| \\ &\qquad + \cdots + |\{c \in B \mid c \text{ has x-coordinate } n+1\}| \\ &= \sum_{i=1}^{n+1} |\{c \in B \mid c \text{ has x-coordinate } i\}| . \end{aligned} \tag{19}$$

But the addends on the right hand side of this equality are easily computed:[36] For each $i \in \{1,2,\ldots,n+1\}$, we have

$$\begin{aligned} &\{c \in B \mid c \text{ has x-coordinate } i\} \\ &= \{c \in B \mid c = (i,j) \text{ for some } j \in [n]\} \\ &\qquad \left(\begin{array}{c} \text{since a square of } [n+1] \times [n] \text{ has x-coordinate } i \\ \text{if and only if it has the form } (i,j) \text{ for some } j \in [n] \end{array} \right) \\ &= \{(i,j) \mid j \in [n] \text{ such that } (i,j) \in B\} \\ &= \{(i,j) \mid j \in [n] \text{ such that } i > j\} \\ &\qquad \left(\begin{array}{c} \text{since a square } (i,j) \in [n+1] \times [n] \text{ satisfies } (i,j) \in B \\ \text{if and only if } i > j \text{ (by the definition of } B) \end{array} \right) \\ &= \{(i,1), (i,2), \ldots, (i,i-1)\} \end{aligned}$$

and thus

$$\begin{aligned} |\{c \in B \mid c \text{ has x-coordinate } i\}| &= |\{(i,1), (i,2), \ldots, (i,i-1)\}| \\ &= i-1. \end{aligned} \tag{20}$$

---

map is well-defined. This means showing that $(n+2-i, n+1-j) \in B$ for each $(i,j) \in A$. Once this is shown, you can show the bijectivity of this map by explicitly constructing an inverse; namely, the inverse is the map

$$\begin{aligned} B &\to A, \\ (i,j) &\mapsto (n+2-i, n+1-j). \end{aligned}$$

(Don't be surprised that it is given by the same formula: In order to undo a reflection around a point, you have to reflect again around the same point.)

[36] This is just the formalization of our (visually obvious) observation that the part below the broken line (which we are now calling B) has 0 squares in the 1-st column, 1 square in the 2-nd column, 2 squares in the 3-rd column, etc..

Hence, (**??**) becomes

$$|B| = \sum_{i=1}^{n+1} \underbrace{|\{c \in B \mid c \text{ has x-coordinate } i\}|}_{\substack{=i-1 \\ \text{(by (??))}}}$$
$$= \sum_{i=1}^{n+1} (i-1) = 0+1+2+\cdots+n = 1+2+\cdots+n.$$

Comparing this with (**??**), we obtain $1+2+\cdots+n = \dfrac{n(n+1)}{2}$. This finally completes our formalized picture proof of Theorem **??**.

The moral of the story: A picture is worth a thousand words! □

1.2.2. What is a sum, actually?

Throughout Subsection **??**, we have been freely working with expressions like $1+2+\cdots+n$ and (more generally) $a_1+a_2+\cdots+a_n$, where $a_1, a_2, \ldots, a_n$ are n numbers. Looking back, you might wonder: Why are these expressions well-defined? How is "the sum of n numbers" defined to begin with?

Let me explain what I mean by this. Let "number" mean "rational number", just to be specific here (although the same question and the same answer apply to integers, real numbers and complex numbers). I assume you know what the sum of **two** numbers is. Even when two numbers a and b are given without specifying their order, their sum is well-defined, because the *commutativity of addition* (i.e., the rule saying that $a+b = b+a$ for any two numbers a and b) guarantees that the two possible ways of adding them together yield the same result.

Now, suppose you are given **three** numbers a, b and c. To add them all together means to add two of them and then add the third one to the result. How many ways are there to do this? There are 12, as you can easily check, namely

$$\begin{array}{llll} (a+b)+c, & (a+c)+b, & (b+a)+c, & (b+c)+a, \\ (c+a)+b, & (c+b)+a, & a+(b+c), & a+(c+b), \\ b+(a+c), & b+(c+a), & c+(a+b), & c+(b+a). \end{array}$$

If we want to make sense of a "sum of three numbers" without having to specify the precise procedure of summation, we better hope that these 12 ways all lead to the same result! And indeed, they do. This is not hard to derive from the above-mentioned commutativity of addition, combined with the *associativity of addition* (i.e., the rule saying that $(a+b)+c = a+(b+c)$ for any three numbers a, b and

c). For example,

$$\begin{aligned} a+(b+c) &= \underbrace{(a+b)}_{\substack{=b+a \\ \text{(by commutativity)}}} + c \qquad \text{(by associativity)} \\ &= (b+a)+c = b+\underbrace{(a+c)}_{\substack{=c+a \\ \text{(by commutativity)}}} \qquad \text{(by associativity)} \\ &= b+(c+a) = (b+c)+a \qquad \text{(by associativity)} \end{aligned}$$

and so on (in the sense that similar reasoning shows that all 12 ways give equal results). Once we know that these 12 ways all lead to the same result, we can call the result "the sum of the three numbers a, b and c" with a clear conscience, and denote it by $a+b+c$ without specifying the order in which the sum is taken through well-placed parentheses.

Next, suppose you are given **four** numbers a, b, c and d. There are now 120 ways of adding them together, including such ways as

$$((a+b)+c)+d, \qquad (a+b)+(c+d), \qquad (b+d)+(c+a), \qquad (b+(d+a))+c$$

and many others. How can we tell that they all lead to the same result? They do, and this can be proven as for three numbers (we don't need any new rules; commutativity and associativity suffice), but this is of course more laborious than the case of three numbers.

Now, suppose you are given n numbers $a_1, a_2, \ldots, a_n$, and you want to prove that all ways of adding them together give the same result. For example, two of these ways are

$$(\cdots(((a_1+a_2)+a_3)+a_4)+\cdots)+a_n \qquad \text{("left-associative summation")}$$

and

$$a_1+(\cdots+(a_{n-3}+(a_{n-2}+(a_{n-1}+a_n)))\cdots) \qquad \text{("right-associative summation");}$$

how do we know they are equivalent?

Let us close a ring around this problem. As long as we don't know that all ways to add n numbers, we cannot define "**the** sum" of n numbers. But we can define the **set of all possible sums** of n numbers, i.e., the set of all possible results that can be obtained by adding them together in all possible orders. For example, for three numbers a, b, c, this set will be

$$\{(a+b)+c, (a+c)+b, \ldots, c+(b+a)\}$$

(containing all the 12 ways to add a, b, c together, listed above). This definition can easily be done by recursion:

- There is only one way to add 0 numbers together. Namely, adding 0 numbers always yields 0.
- There is only one way to add 1 number together. Namely, adding 1 number a always yields a itself.
- If $n > 1$, then any way of adding n numbers together is given by splitting them into two (disjoint) groups[37], then adding the numbers in the first group together (in one of the many possible ways), then adding the numbers in the second group together, and finally adding the two results together.

Thus we can define the **set of all possible sums** of n numbers. Our goal is to prove that there is only one such sum, i.e., that this set is a 1-element set. This is now a rigorously stated claim[38], which we can try to prove! Thus, at the very least, we have formalized our claim that we can add n numbers together in a well-defined way. Proving it is another story, but it can be done with rather elementary tools. See [**?**, §2.14] for a very detailed proof.[39]

The upshot is that we have a well-defined notion of the sum of any finite family of numbers, even if it is given without specifying an order. For example, "the sum of all prime numbers smaller than 10" is well-defined (and equals $2+3+5+7=17$). Likewise, if you are given a polygon, then "the sum of the sidelengths of the polygon" is well-defined.

It is helpful to have a notation for these kinds of sums, so let us introduce it (although we have already used it to some extent):

Definition 1.2.2. Let S be a finite set. For each $s \in S$, let a_s be a number (e.g., an integer or a rational number or a real number or a complex number).

Then, $\sum_{s \in S} a_s$ shall denote the sum of the numbers a_s for all $s \in S$. This notation is read as "the sum of a_s over all $s \in S$" or "the sum of a_s for s ranging over S" or "the sum of a_s where s runs through S". The "$\sum$" sign in this notation is called the *summation sign*.

Example 1.2.3. **(a)** We have $\sum_{s \in \{1,2,3,4\}} s = 1+2+3+4 = 10$.

(b) We have $\sum_{s \in \{1,2,3,4\}} s^2 = 1^2+2^2+3^2+4^2 = 30$.

(c) We have $\sum_{s \in \{1,2,3,4\}} \frac{1}{s} = \frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4} = \frac{25}{12}$.

(d) We have $\sum_{s \in \{3,4,9\}} s = 3+4+9 = 16$.

[37] "Groups" in the sense of "batches", not in the sense of group theory.

[38] known as the *general commutativity theorem* ("general" because it applies to any finite number of addends rather than two)

[39] More precisely, the proof is in [**?**, §2.14.1–§2.14.6]. The remaining subsections of [**?**, §2.14] prove properties of sums.

We need to note a few things about the summation sign:

- The letter "s" in the notation "$\sum_{s \in S} a_s$" is an instance of what is called a *bound variable* (or *running index*, or *dummy variable*): Its only purpose is to mark the "moving part" of the sum. So it plays the same role as the letter "s" in the set-comprehension notation "$\{s \in \mathbb{Z} \mid s > 2\}$", or the letter "$s$" in "the map $\mathbb{Z} \to \mathbb{Z}$, $s \mapsto s+3$", or the letter "s" in the sentence "There exists no $s \in \mathbb{Z}$ such that $s^2 = 2$". Thus, it is perfectly legitimate to replace it by any other symbol (that is not used otherwise). For example, we can rewrite the sum $\sum_{s \in \{1,2,3,4\}} \frac{1}{s}$ as $\sum_{i \in \{1,2,3,4\}} \frac{1}{i}$ or as $\sum_{k \in \{1,2,3,4\}} \frac{1}{k}$ or as $\sum_{\mathfrak{S} \in \{1,2,3,4\}} \frac{1}{\mathfrak{S}}$ or as $\sum_{\spadesuit \in \{1,2,3,4\}} \frac{1}{\spadesuit}$; it will still be the same sum.

- In the notation "$\sum_{s \in S} a_s$", the letter "s" is called the *summation index*; the set S is called the *indexing set*; and the numbers a_s are called the *addends* of the sum.

- Our definition of sums forces $\sum_{s \in \varnothing} a_s$ to always be 0. This is called an *empty sum*. Thus, empty sums are 0. If this sounds like an arbitrary convention to you, you should check that it is the only convention that makes (**??**) hold for one-element sets S.

- Sums can have equal addends. For example,

$$\sum_{s \in \{-2,-1,0,1,2\}} s^2 = (-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 = 4 + 1 + 0 + 1 + 4 = 10.$$

  However,

$$\sum_{s \in \left\{(-2)^2, (-1)^2, 0^2, 1^2, 2^2\right\}} s = 0^2 + 1^2 + 2^2 = 0 + 1 + 4 = 5,$$

  because the **set** $\left\{(-2)^2, (-1)^2, 0^2, 1^2, 2^2\right\}$ is only a 3-element set (with elements $0^2, 1^2, 2^2$). Thus, if you want to write down a sum with some of its entries equal, you still need to ensure that each of the addends gets a distinct index in the indexing set. In general, the sum of 5 numbers is not the sum of the elements of the **set** of these 5 numbers, because if some of these 5 numbers are equal, then the set "forgets" that they appear more than once.

- The summation index does not always have to be a single letter. For instance, if S is a set of pairs, then we can write $\sum_{(x,y) \in S} a_{(x,y)}$ (meaning the same as $\sum_{s \in S} a_s$). Here is an example of this notation:

$$\sum_{(x,y) \in \{1,2,3\}^2} \frac{x}{y} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{2}{1} + \frac{2}{2} + \frac{2}{3} + \frac{3}{1} + \frac{3}{2} + \frac{3}{3}.$$

Here are some more examples, coming from more combinatorial questions:

Example 1.2.4. (a) Let $\mathcal{P}(A)$ denote the powerset of a set A (that is, the set of all subsets of A). Then, we can write the sum of the sizes of all subsets of $\{1,2,3\}$ as follows:

$$\begin{aligned}\sum_{B\in\mathcal{P}(\{1,2,3\})} |B| &= |\varnothing| + |\{1\}| + |\{2\}| + |\{3\}| \\ &\quad + |\{1,2\}| + |\{1,3\}| + |\{2,3\}| + |\{1,2,3\}| \\ &= 0+1+1+1+2+2+2+3 = 12.\end{aligned}$$

(b) If A and B are two sets, then B^A denotes the set of all maps from A to B. Then, we can write the sum of the sizes of the images of all maps from $\{1,2\}$ to $\{1,2\}$ as follows:

$$\begin{aligned}\sum_{f\in B^A} |f(A)| = &\underbrace{|\{1\}|}_{\substack{\text{corresponding}\\ \text{to the map } \left(\begin{smallmatrix}1&2\\1&1\end{smallmatrix}\right)}} + \underbrace{|\{1,2\}|}_{\substack{\text{corresponding}\\ \text{to the map } \left(\begin{smallmatrix}1&2\\1&2\end{smallmatrix}\right)}} \\ &+ \underbrace{|\{1,2\}|}_{\substack{\text{corresponding}\\ \text{to the map } \left(\begin{smallmatrix}1&2\\2&1\end{smallmatrix}\right)}} + \underbrace{|\{2\}|}_{\substack{\text{corresponding}\\ \text{to the map } \left(\begin{smallmatrix}1&2\\2&2\end{smallmatrix}\right)}} \\ &= 1+2+2+1 = 6.\end{aligned}$$

(Here, we have written out each map in two-line notation: That is, the map sending 1 and 2 to a_1 and a_2 has been written as $\begin{pmatrix}1&2\\a_1&a_2\end{pmatrix}$.)

There are several variants of the summation sign:

- The expression $\sum\limits_{\substack{s\in\{1,2,3,4,5,6\};\\ s \text{ is odd}}} s^2$ stands for the sum of the squares of all **odd** elements of $\{1,2,3,4,5,6\}$. In general, if S is a set, and if $\mathcal{A}(s)$ is a logical statement defined for each $s \in S$, then the notation $\sum\limits_{\substack{s\in S;\\ \mathcal{A}(s)}} a_s$ stands for the sum $\sum\limits_{s\in\{t\in S \mid \mathcal{A}(t)\}} a_s$ (that is, the sum of a_s not over all $s \in S$, but only over the ones which satisfy $\mathcal{A}(s)$). For this sum to be well-defined, we do not need S to be finite; we only need the subset $\{t \in S \mid \mathcal{A}(t)\}$ to be finite (i.e., we need there to only be finitely many $s \in S$ satisfying $\mathcal{A}(s)$).
- The expression $\sum\limits_{i=5}^{8} i^2$ stands for the sum $\sum\limits_{i\in\{5,6,7,8\}} i^2$. In general, if p and q are

two integers, then the notation $\sum_{i=p}^{q} a_i$ stands for the sum $\sum_{i \in \{p,p+1,\ldots,q\}} a_i$. Here it should be kept in mind that the set $\{p, p+1, \ldots, q\}$ is understood to be empty if $p > q$. (Be warned that some authors have different conventions for the latter case.)

- The summation sign $\sum_{B \subseteq A}$ (where A is a given set) is shorthand for $\sum_{B \in \mathcal{P}(A)}$. Thus, the result of Example **??** **(a)** could be rewritten as $\sum_{B \subseteq \{1,2,3\}} |B| = 12$.
- The summation sign $\sum_{f: A \to B}$ (where A and B are two sets) is shorthand for $\sum_{f \in B^A}$. Thus, the result of Example **??** **(b)** could be rewritten as $\sum_{f: A \to B} |f(A)| = 6$.

These notations can be mixed and matched. For example, you can write $\sum_{\substack{f: A \to B; \\ f \text{ is injective}}} |f(A)|$ to obtain a sum of the sizes of the images of all **injective** maps from A to B.

1.2.3. The sums $1^k + 2^k + ... + n^k$

Let us now go back to Theorem **??**. That theorem gave an explicit formula for the sum of the first n positive integers. A similar formula exists for the sum of the squares of the first n positive integers:

Proposition 1.2.5. Let $n \in \mathbb{N}$. Then,

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

Proof of Proposition **??** *(sketched).* Again, this can be shown by a straightforward induction (LTTR). □

Is there a picture proof for Proposition **??** as well, similar to the one we gave for Theorem **??**? Yes, but the picture will be 3-dimensional and thus much harder to reason about.

Is there a proof of Proposition **??** that is analogous to our second proof of Theorem **??** above? Not to my knowledge. The numbers $1^2 + n^2, 2^2 + (n-1)^2, \ldots, n^2 + 1^2$ are not equal, and I don't know of a way to reorder the addends in the (expanded) sum $6 \cdot \left(1^2 + 2^2 + \cdots + n^2\right)$ to obtain $n(n+1)(2n+1)$.

Next, let us sum the cubes of the first n positive integers:

Proposition 1.2.6. Let $n \in \mathbb{N}$. Then,

$$1^3 + 2^3 + \cdots + n^3 = \frac{n^2 (n+1)^2}{4}.$$

Proof of Proposition **??** *(sketched).* Again, a straightforward induction (LTTR). □

Class of 2019-09-30 (to-do)

TODO: properties of $\sum$ sign.

TODO: $\prod$ and product signs.

Question: Given $k \in \mathbb{N}$, what is $1^k + 2^k + \cdots + n^k$? Is it a $(k+1)$-degree polynomial in n ?

Definition 1.2.7. If $a_p, a_{p+1}, \ldots, a_q$ (for some integers p and q) are numbers, then $\sum_{i=p}^{q} a_i$ means $a_p + a_{p+1} + \cdots + a_q$ (or, equivalently, $\sum_{i \in \{p, p+1, \ldots, q\}} a_i$).

Thus, $1^k + 2^k + \cdots + n^k$ rewrites as

$$\sum_{i=1}^{n} i^k.$$

Theorem 1.2.8. (18th Century?) Let $n, k \in \mathbb{N}$ with $k > 0$. Then,

$$\sum_{i=1}^{n} i^k = \sum_{i=0}^{k} \operatorname{sur}(k,i) \cdot \binom{n+1}{i+1},$$

where:

- as before, we set $[k] = \{1, 2, \ldots, k\}$ for each $k \in \mathbb{N}$;
- $\operatorname{sur}(k,i)$ means the number of surjective maps from $[k] \to [i]$ (remember: $[k] = \{1, 2, \ldots, k\}$ and $[i] = \{1, 2, \ldots, i\}$, and $[0] = \varnothing$).
- $\dbinom{x}{j} = \dfrac{x(x-1)(x-2)\cdots(x-j+1)}{j(j-1)(j-2)\cdots 1}$ for all $x \in \mathbb{R}$ and $j \in \mathbb{N}$.

Example 1.2.9. Let $k = 2$. Then, this theorem yields

$$\begin{aligned}
1^2 + 2^2 + \cdots + n^2 &= \sum_{i=0}^{2} \operatorname{sur}(2,i) \cdot \binom{n+1}{i+1} \\
&= \underbrace{\operatorname{sur}(2,0)}_{=0} \cdot \underbrace{\binom{n+1}{0+1}}_{=\binom{n+1}{1}=\frac{n+1}{1}} + \underbrace{\operatorname{sur}(2,1)}_{=1} \cdot \underbrace{\binom{n+1}{1+1}}_{=\binom{n+1}{2}=\frac{(n+1)n}{2\cdot 1}} \\
&\quad + \underbrace{\operatorname{sur}(2,2)}_{=2} \cdot \underbrace{\binom{n+1}{2+1}}_{=\binom{n+1}{3}=\frac{(n+1)n(n-1)}{3\cdot 2\cdot 1}} \\
&= 0 \cdot \frac{n+1}{1} + 1 \cdot \frac{(n+1)n}{2\cdot 1} + 2 \cdot \frac{(n+1)n(n-1)}{3\cdot 2\cdot 1} \\
&= \frac{n(n+1)(2n+1)}{6}.
\end{aligned}$$

Thus, our formula for $1^2 + 2^2 + \cdots + n^2$ follows from the Theorem. Similarly we can get formulas for sums of all other powers.

We will prove the Theorem later.

1.3. Factorials and binomial coefficients

Definition 1.3.1. For any $n \in \mathbb{N}$, set $n! = 1 \cdot 2 \cdot \cdots \cdot n$.

This is understood to be 1 when $n = 0$, since generally, empty products are defined to be 1. (Same logic as for why $0^0 = 1$, or more generally $k^0 = 1$ for all k.)

We shall refer to $n!$ as "n factorial".

Remark 1.3.2. If n is a positive integer, then

$$\begin{aligned}
n! &= 1 \cdot 2 \cdot \cdots \cdot n = \underbrace{(1 \cdot 2 \cdot \cdots \cdot (n-1))}_{=(n-1)!} \cdot n \\
&= (n-1)! \cdot n.
\end{aligned}$$

Example 1.3.3.

$$\begin{aligned}
0! &= 1, \\
1! &= 1, \\
2! &= 1 \cdot 2 = 2, \\
3! &= 1 \cdot 2 \cdot 3 = 6, \\
4! &= 1 \cdot 2 \cdot 3 \cdot 4 = 24, \\
5! &= 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120, \\
6! &= 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720, \\
7! &= 5\,040, \\
8! &= 40\,320, \\
&\vdots
\end{aligned}$$

Definition 1.3.4. Let n be any number ($n \in \mathbb{N}$ or $n \in \mathbb{Z}$ or $n \in \mathbb{R}$ or $n \in \mathbb{C}$).

(a) For any $k \in \mathbb{N}$, we set

$$\binom{n}{k} = \frac{n\,(n-1)\,(n-2)\cdots(n-k+1)}{k!}.$$

(b) If $k \notin \mathbb{N}$, we set

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

(This is the convention of [Graham, Knuth and Patashnik].)

We call $\binom{n}{k}$ a **binomial coefficient**, and we refer to it as "n **choose** k".

Remark 1.3.5. For any number n and any $k \in \mathbb{N}$, we have

$$\binom{n}{k} = \frac{n^{\underline{k}}}{k!} \qquad \text{using the notation of HW0 exercise 2.}$$

Example 1.3.6.

$$\binom{n}{0} = \frac{(\text{empty product})}{0!} = \frac{1}{1} = 1;$$
$$\binom{n}{1} = \frac{n}{1!} = \frac{n}{1} = n;$$
$$\binom{n}{2} = \frac{n(n-1)}{2!} = \frac{n(n-1)}{2};$$
$$\binom{n}{3} = \frac{n(n-1)(n-2)}{3!} = \frac{n(n-1)(n-2)}{6};$$
$$\binom{-1}{5} = \frac{(-1)(-2)(-3)(-4)(-5)}{5!} = \frac{(-1)(-2)(-3)(-4)(-5)}{1\cdot 2\cdot 3\cdot 4\cdot 5} = -1;$$
$$\binom{-1}{k} = \frac{(-1)(-2)\cdots(-k)}{k!} = (-1)^k = \begin{cases} 1, & \text{if } k \text{ is even;} \\ -1, & \text{if } k \text{ is odd} \end{cases}.$$

——————— 2019-10-02

Example 1.3.7.

$$\binom{\sqrt{2}}{2} = \frac{\sqrt{2}\left(\sqrt{2}-1\right)}{2};$$
$$\binom{2}{\sqrt{2}} = 0 \qquad \text{since } \sqrt{2} \notin \mathbb{N}.$$

One of the major topics in combinatorics is properties of binomial coefficients.

Proposition 1.3.8. If $n \in \mathbb{N}$ and $k > n$, then $\binom{n}{k} = 0$.

Proof. WLOG $k \in \mathbb{N}$, since otherwise $\binom{n}{k} = 0$ by definition.

So

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} = 0,$$

because one of the factors $n, n-1, n-2, \ldots, n-k+1$ is $n-n=0$. □

(Cf. the old math joke: Simplify $(x-a)(x-b)\cdots(x-z)$.)

Proposition 1.3.9. ("Upper negation") Let $n \in \mathbb{R}$ and $k \in \mathbb{Z}$. Then,

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

Proof. We can have two cases:
Case 1: We have $k \notin \mathbb{N}$. Then, both sides are 0, qed.
Case 2: We have $k \in \mathbb{N}$. Then,

$$\begin{aligned}
\binom{-n}{k} &= \frac{(-n)(-n-1)(-n-2)\cdots(-n-k+1)}{k!} \\
&= \frac{(-1)^k \cdot n(n+1)(n+2)\cdots(n+k-1)}{k!} \\
&= \frac{(-1)^k \cdot (n+k-1)(n+k-2)(n+k-3)\cdots n}{k!} \\
&= (-1)^k \cdot \underbrace{\frac{(n+k-1)(n+k-2)(n+k-3)\cdots n}{k!}}_{=\binom{n+k-1}{k}} \\
&= (-1)^k \cdot \binom{n+k-1}{k}.
\end{aligned}$$

□

Here are some more results, to be proven later.

Theorem 1.3.10. (Recurrence of the BC). Let $n \in \mathbb{R}$ and $k \in \mathbb{Z}$. Then,

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

This theorem gives a quick way to compute a given $\binom{n}{k}$ when $n, k \in \mathbb{N}$. Usually, these numbers $\binom{n}{k}$ with $n, k \in \mathbb{N}$ are tabulated in a triangular table called *Pascal's triangle*, where the rows correspond to the values of n and the diagonals (of $\diagup$ form) correspond to the values of k. In this triangle, each number is the sum of two numbers on the previous row, namely the one left-above and the one right-above. (See Wikipedia.)

Theorem 1.3.11. (Symmetry of the BC) Let $n \in \mathbb{N}$ and $k \in \mathbb{Z}$. Then,

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

Theorem 1.3.12. (Integrality of the BC) Let $n \in \mathbb{Z}$ and $k \in \mathbb{Z}$. Then, $\binom{n}{k} \in \mathbb{Z}$.

Theorem 1.3.13. (Combinatorial interpretation of the BC)
Let $n \in \mathbb{N}$ and $k \in \mathbb{Z}$. Let S be an n-element set. Then,

$$\binom{n}{k} = (\#\text{ of } k\text{-element subsets of } S).$$

Example 1.3.14. Let $n = 4$ and $k = 2$ and $S = \{1,2,3,4\}$. Then, the 2-element subsets of S are $\{1,2\}$, $\{1,3\}$, $\{1,4\}$, $\{2,3\}$, $\{2,4\}$ and $\{3,4\}$. Their number is $6 = \binom{4}{2}$, as the theorem above predicts.

We will prove these later ([detnotes, Ch. 3], or [Graham, Knuth, Patashnik], or essentially any other text on combinatorics).

The combinatorial interpretation of the BC is the reason why $\binom{n}{k}$ is called "n choose k": It is the number of ways to **choose** k distinct elements (without an order) from n given elements. The k-element subsets of S are called the **k-combinations** of S.

Warning 1.3.15. The combinatorial interpretation of the BC says nothing about $\binom{n}{k}$ when $n \notin \mathbb{N}$. Neither does the symmetry property.

Theorem 1.3.16. ("Hockey-stick identity"). Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$. Then,

$$\binom{0}{k} + \binom{1}{k} + \binom{2}{k} + \cdots + \binom{n}{k} = \binom{n+1}{k+1}.$$

Proof. HW0 Exercise 2 says that

$$0^{\underline{k}} + 1^{\underline{k}} + \cdots + n^{\underline{k}} = \frac{1}{k+1}(n+1)^{\underline{k+1}}.$$

Dividing it by $k!$ yields

$$\begin{aligned}\binom{0}{k} + \binom{1}{k} + \binom{2}{k} + \cdots + \binom{n}{k} &= \frac{1}{(k+1)\cdot k!}(n+1)^{\underline{k+1}} \\ &= \frac{1}{(k+1)!}\cdot(n+1)^{\underline{k+1}} = \binom{n+1}{k+1}.\end{aligned}$$

□

Alternatively, this can be proven by induction or combinatorially; we will see those proofs later.

Theorem 1.3.17. (The binomial formula) Let $x, y \in \mathbb{R}$. Let $n \in \mathbb{N}$. Then,

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

This is why the $\binom{n}{k}$ are called *binomial coefficients* – they appear as coefficients in the binomial formula.

Corollary 1.3.18. Let $n \in \mathbb{N}$. Then, $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

Proof. Apply the binomial formula to $x = 1$ and $y = 1$. □

Proposition 1.3.19. Let $n \in \mathbb{N}$. Let $a, b \in \{0, 1, \ldots, 2^n - 1\}$. Then:

(a) $\binom{2^n + a}{b} \equiv \binom{a}{b} \bmod 2$.

(b) $\binom{2^n + a}{2^n + b} \equiv \binom{a}{b} \bmod 2$.

Theorem 1.3.20. Let p be a prime. Let $k \in \{1, 2, \ldots, p-1\}$. Then, $p \mid \binom{p}{k}$.

Proposition 1.3.21. Let $n \in \mathbb{N}$. Then, the Fibonacci number f_{n+1} is

$$f_{n+1} = \sum_{k=0}^{n} \binom{n-k}{k} = \binom{n-0}{0} + \binom{n-1}{1} + \binom{n-2}{2} + \cdots + \binom{n-n}{n}.$$

Theorem 1.3.22. (Cauchy's binomial formula) Let $x, y, z \in \mathbb{R}$. Then,

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

Theorem 1.3.23. (Abel's binomial formula) Let $x, y, z \in \mathbb{R}$. Then,

$$\sum_{k=0}^{n} \binom{n}{k} \underbrace{x (x+kz)^{k-1}}_{\substack{\text{This should be read} \\ \text{as 1 if } k=0}} (y-kz)^{n-k} = (x+y)^n.$$

Proof. (Proof of the Recurrence of the BC).
We are in one of the following three cases:
Case 1: We have $k \in \{1,2,3,\ldots\}$.
Case 2: We have $k = 0$.
Case 3: We have $k \notin \mathbb{N}$.
Case 3 is easy: If $k \notin \mathbb{N}$, then $k-1 \notin \mathbb{N}$, so $\binom{n-1}{k-1} = 0$, but also $\binom{n}{k} = 0$ and $\binom{n-1}{k} = 0$ (since $k \notin \mathbb{N}$). Thus, we have to prove that $0 = 0+0$.
Case 2: Assume $k = 0$. Then, $\binom{n}{k} = \binom{n}{0} = 1$, and similarly $\binom{n-1}{k} = 1$. Also, $\binom{n-1}{k-1} = \binom{n-1}{-1} = 0$. So we have to prove that $1 = 0+1$. □

Class of 2019-09-04

Proof. Case 1: Assume $k \in \{1,2,3,\ldots\}$. Thus, both k and $k-1$ belong to $\mathbb{N}$.
Now, by the definition of BCs,

$$\begin{aligned}
&\binom{n-1}{k-1} + \binom{n-1}{k} \\
&= \frac{(n-1)(n-2)(n-3)\cdots(n-k+1)}{(k-1)!} + \frac{(n-1)(n-2)(n-3)\cdots(n-k)}{k!} \\
&= \frac{(n-1)(n-2)(n-3)\cdots(n-k+1)}{(k-1)!} \underbrace{\left(1 + \frac{n-k}{k}\right)}_{=\frac{n}{k}} \qquad (\text{since } k! = (k-1)!\cdot k) \\
&= \frac{(n-1)(n-2)(n-3)\cdots(n-k+1)}{(k-1)!} \cdot \frac{n}{k} \\
&= \frac{n(n-1)(n-2)(n-3)\cdots(n-k+1)}{(k-1)!\cdot k} = \frac{n(n-1)(n-2)(n-3)\cdots(n-k+1)}{k!} = \binom{n}{k}.
\end{aligned}$$

Thus, the theorem is proven in Case 1 as well. □

Theorem 1.3.24. (Factorial formula for the BCs)
Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$ be such that $k \leq n$. Then,

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}.$$

Proof.

$$\begin{aligned} k! \cdot (n-k)! \cdot \binom{n}{k} &= k! \cdot (n-k)! \cdot \frac{n\,(n-1)\,(n-2)\cdots(n-k+1)}{k!} \\ &= (n-k)! \cdot (n\,(n-1)\,(n-2)\cdots(n-k+1)) \\ &= n\,(n-1)\cdots 1 = n!. \end{aligned}$$

□

Theorem 1.3.25. (Symmetry of the BC) Let $n \in \mathbb{N}$ and $k \in \mathbb{Z}$. Then,

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

Proof. Case 1: $k \in \{0, 1, \ldots, n\}$.
Case 2: $k \notin \mathbb{Z}$.
Case 3: Neither.
In Case 1, the previous theorem yields

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} \qquad \text{and}$$
$$\binom{n}{n-k} = \frac{n!}{(n-k)! \cdot (n-(n-k))!} = \frac{n!}{(n-k)! \cdot k!} = \frac{n!}{k! \cdot (n-k)!}.$$

Comparing these yields $\binom{n}{k} = \binom{n}{n-k}$. Thus, we are done in Case 1.

In Case 2, we just need to prove $0 = 0$.

Now let us look at Case 3. Here, $k \in \mathbb{Z}$ but $k \notin \{0, 1, \ldots, n\}$. Thus, either $k < 0$ or $k > n$. Hence, again, we just need to prove $0 = 0$, because we know that $\binom{n}{\text{something negative}} = 0$ and $\binom{n}{\text{an integer } > n} = 0$. □

Proposition 1.3.26. Let $k \in \mathbb{R}$. Then, $\binom{0}{k} = [k = 0]$.

Recall: We are using the *Iverson bracket notation*: i.e., for any statement $\mathcal{A}$, we let $[\mathcal{A}]$ denote the *truth value* of $\mathcal{A}$, defined to be $\begin{cases} 1, & \text{if } \mathcal{A} \text{ is true;} \\ 0, & \text{if } \mathcal{A} \text{ is false} \end{cases}$.

Proof. (Proof of the Proposition.) Again, there are three cases to consider: $k = 0$, $k \notin \mathbb{N}$ or $k \in \{1, 2, 3, \ldots\}$. □

Theorem 1.3.27. (Combinatorial interpretation of the BC)
Let $n \in \mathbb{N}$ and $k \in \mathbb{Z}$. Let S be an n-element set. Then,

$$\binom{n}{k} = (\#\text{ of } k\text{-element subsets of } S)\,.$$

Proof. Forget that we fixed n and k. Induction on n.
Induction base: If S is a 0-element set, then $S = \varnothing$ and thus

$$\begin{aligned}
&(\#\text{ of } k\text{-element subsets of } S)\\
&= (\#\text{ of } k\text{-element subsets of } \varnothing)\\
&= \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k \neq 0 \end{cases} = [k = 0] = \binom{0}{k} \qquad \text{(by the previous proposition)}\,.
\end{aligned}$$

Thus, the theorem is proven for $n = 0$.
Induction step: Let $m \in \mathbb{N}$. Assume (as IH) that the Theorem holds for $n = m$.
Let S be an $(m+1)$-element set. We must prove that

$$\binom{m+1}{k} = (\#\text{ of } k\text{-element subsets of } S)\,.$$

The set S is nonempty, since its size is $|S| = m + 1 \geq 1 > 0$. Thus, there exists a $t \in S$. Fix such a t.
Now, there are two types of subsets of S:

- **red subsets**: those that contain t;
- **green subsets:** those that don't contain t.

So, by the sum rule,

$$\begin{aligned}
&(\#\text{ of } k\text{-element subsets of } S)\\
&= (\#\text{ of } k\text{-element red subsets of } S) + (\#\text{ of } k\text{-element green subsets of } S)\,.
\end{aligned}$$

But the green subsets of S are exactly the subsets of $S \setminus \{t\}$. Hence,

$$\begin{aligned}
&(\#\text{ of } k\text{-element green subsets of } S)\\
&= (\#\text{ of } k\text{-element subsets of } S \setminus \{t\})\\
&= \binom{m}{k} \qquad \text{(by the IH, since } S \setminus \{t\} \text{ is an } m\text{-element set)}\,.
\end{aligned}$$

What about the red subsets?
Informally: The k-element red subsets of S "correspond to" the $(k-1)$-element subsets of $S \setminus \{t\}$.

Formally: The map

$$\{k\text{-element red subsets of } S\} \to \{(k-1)\text{-element subsets of } S \setminus \{t\}\}, \\ Q \mapsto Q \setminus \{t\}$$

is a bijection. (Its inverse is

$$\{(k-1)\text{-element subsets of } S \setminus \{t\}\} \to \{k\text{-element red subsets of } S\}, \\ P \mapsto P \cup \{t\}.$$

) Therefore, the bijection principle yields

$$|\{k\text{-element red subsets of } S\}| \\ = |\{(k-1)\text{-element subsets of } S \setminus \{t\}\}|.$$

In other words,

$$(\#\text{ of } k\text{-element red subsets of } S) \\ = (\#\text{ of } (k-1)\text{-element subsets of } S \setminus \{t\}) \\ = \binom{m-1}{k-1} \qquad (\text{by the IH, applied to } k-1 \text{ instead of } k).$$

Now,

$$(\#\text{ of } k\text{-element subsets of } S) \\ = \underbrace{(\#\text{ of } k\text{-element red subsets of } S)}_{=\binom{m-1}{k-1}} + \underbrace{(\#\text{ of } k\text{-element green subsets of } S)}_{=\binom{m-1}{k}} \\ = \binom{m-1}{k-1} + \binom{m-1}{k} = \binom{m}{k} \qquad (\text{by the recurrence of the BCs}).$$

And this completes the induction step. □

Theorem 1.3.28. (Integrality of the BC) Let $n \in \mathbb{Z}$ and $k \in \mathbb{Z}$. Then, $\binom{n}{k} \in \mathbb{Z}$.

Proof. Three cases are possible:
Case 1: $k \notin \mathbb{N}$.
Case 2: $k \in \mathbb{N}$ and $n \in \mathbb{N}$.
Case 3: $k \in \mathbb{N}$ and $n \notin \mathbb{N}$.
In Case 1, we just need to prove that $0 \in \mathbb{Z}$.
In Case 2, the claim follows from the Combinatorial interpretation of the BCs.

In Case 3, we must have $n < 0$. Now, the Upper negation formula (applied to $-n$ instead of n) yields

$$\binom{n}{k} = (-1)^k \underbrace{\binom{-n+k-1}{k}}_{\substack{\in \mathbb{Z} \\ \text{(according to the Combinatorial interpretation of the BCs, because } -n+k-1 \in \mathbb{N})}} \in \mathbb{Z}.$$

□

Theorem 1.3.29. (The binomial formula) Let $x, y \in \mathbb{R}$. Let $n \in \mathbb{N}$. Then,

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

Proof. I will rewrite $\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$ as $\sum_{k \in \mathbb{Z}} \binom{n}{k} x^k y^{n-k}$ (that is, a sum over **all** integers k).

Why is this latter **infinite** sum well-defined?

Let us look at the case $n = 3$. In this case, this sum has the form

$$\cdots + 0 + 0 + 0 + y^3 + 3xy^2 + 3x^2y + x^3 + 0 + 0 + 0 + \cdots.$$

So it is infinite, but only finitely many of its addends are nonzero. The same is true for any $n \in \mathbb{N}$.

An infinite sum that has only finitely many nonzero addends is always well-defined: Its value is obtained by discarding the 0's (and summing the nonzero addends).

(To be fully precise: We have to agree to interpret a product of the form ab as 0 whenever $a = 0$, even if b is undefined. This matters, because of $x = 0$, then for negative k, the power x^k will be undefined.)

We have now made sense of $\sum_{k \in \mathbb{Z}} \binom{n}{k} x^k y^{n-k}$. Since $\binom{n}{k} = 0$ whenever $k \notin \{0, 1, \ldots, n\}$, this sum just equals $\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

Thus, it remains to prove that

$$(x+y)^n = \sum_{k \in \mathbb{Z}} \binom{n}{k} x^k y^{n-k}. \tag{21}$$

We will prove this by induction:

Base: $1 = 1$.

Step: Fix $m \in \mathbb{N}$. Assume (2) holds for $n = m$. We must prove that (2) holds for $n = m + 1$.

We have

$$\begin{aligned}
(x+y)^{m+1} &= \underbrace{(x+y)^m}_{\substack{=\sum_{k\in\mathbb{Z}}\binom{m}{k}x^k y^{m-k} \\ \text{(by our IH)}}} (x+y) \\
&= \left(\sum_{k\in\mathbb{Z}}\binom{m}{k}x^k y^{m-k}\right)(x+y) \\
&= \left(\sum_{k\in\mathbb{Z}}\binom{m}{k}x^k y^{m-k}\right)x + \left(\sum_{k\in\mathbb{Z}}\binom{m}{k}x^k y^{m-k}\right)y \\
&= \sum_{k\in\mathbb{Z}} x\binom{m}{k}x^k y^{m-k} + \sum_{k\in\mathbb{Z}} y\binom{m}{k}x^k y^{m-k} \\
&= \sum_{k\in\mathbb{Z}}\binom{m}{k}x^{k+1}y^{m-k} + \sum_{k\in\mathbb{Z}}\binom{m}{k}x^k y^{m-k+1} \\
&= \sum_{k\in\mathbb{Z}}\binom{m}{k-1}x^{(k-1)+1}y^{m-(k-1)} + \sum_{k\in\mathbb{Z}}\binom{m}{k}x^k y^{m-k+1} \\
&\qquad \left(\text{here, we substituted } k-1 \text{ for } k \text{ in the first sum}\right) \\
&= \sum_{k\in\mathbb{Z}}\binom{m}{k-1}x^k y^{m-k+1} + \sum_{k\in\mathbb{Z}}\binom{m}{k}x^k y^{m-k+1} \\
&= \sum_{k\in\mathbb{Z}} \underbrace{\left(\binom{m}{k-1}+\binom{m}{k}\right)}_{\substack{=\binom{m+1}{k} \\ \text{(by the recurrence of the BCs)}}} x^k \underbrace{y^{m-k+1}}_{=y^{m+1-k}} \\
&= \sum_{k\in\mathbb{Z}}\binom{m+1}{k}x^k y^{m+1-k}.
\end{aligned}$$

Thus, (2) holds for $n = m+1$. This completes the induction step. □

So why did we take the trouble to turn our finite sum into an infinite sum? Because this saved us the headache of having to combine two sums with different ranges.

Knuth advises to always make sums infinite when possible. It is not always possible, because sometimes when you extend a sum to a larger range, the new addends you get will not all be 0.

1.4. Counting

1.4.1. Subsets

We have already seen the following enumerations:

- An n-element set has 2^n subsets. (This was HW0 exercise 1 **(a)**.)
- An n-element set has $\binom{n}{k}$ many k-element subsets.

Now, we can ask subtler questions.

Definition 1.4.1. A set S of integers is called **lacunar** if it contains no two consecutive integers (i.e., there is no integer i such that both $i \in S$ and $i+1 \in S$).

For example, $\{1,5,7\}$ is lacunar, but $\{1,5,6\}$ is not.
Some people say "sparse" instead of "lacunar".
Questions:
(a) How many lacunar subsets does $[n]$ have?
(b) How many k-element lacunar subsets does $[n]$ have?
(c) What is the largest size of a lacunar subset of $[n]$?
Let us start with **(c)**:

Definition 1.4.2. Let $x \in \mathbb{R}$. Then:

- $\lfloor x \rfloor$ denotes the largest integer $\leq x$ (called the **floor of** x, or "**rounding down** x").
- $\lceil x \rceil$ denotes the smallest integer $\geq x$ (called the **ceiling of** x, or "**rounding up** x").

Example 1.4.3. $\lfloor 3 \rfloor = 3$, $\left\lfloor \sqrt{2} \right\rfloor = 1$, $\lfloor \pi \rfloor = 3$, $\lfloor -\pi \rfloor = -4$.
$\lceil 3 \rceil = 3$, $\left\lceil \sqrt{2} \right\rceil = 2$, $\lceil \pi \rceil = 4$, $\lceil -\pi \rceil = -3$.

Proposition 1.4.4. Let $n \in \mathbb{N}$. Then, the largest size of a lacunar subset of $[n]$ is $\lceil n/2 \rceil$.

Proof. The lacunar subset

$$\{1,3,5,\ldots,(n \text{ or } n-1)\} = \{1 < 3 < 5 < \cdots < (n \text{ or } n-1)\}$$

(where the last element is n if n is odd, and $n-1$ if n is even) has size $\lceil n/2 \rceil$.
Thus, we only need to prove that no higher size is possible. □

Class of 2019-09-07

Proof. (Continuing the proof:) Let L be a lacunar subset of $[n]$. We want to prove that $|L| \leq \lceil n/2 \rceil$.
Indeed, let L^+ be the set $\{s+1 \mid s \in L\}$.

Then, the sets L and L^+ are disjoint, since L is lacunar. Hence, the sum principle yields $|L \cup L^+| = |L| + \underbrace{\left|L^+\right|}_{=|L|} = 2 \cdot |L|$.

But $L \cup L^+$ is a subset of $[n+1]$. Hence,

$$\left|L \cup L^+\right| \leq |[n+1]| = n+1.$$

Since $|L \cup L^+| = 2 \cdot |L|$, this becomes $2 \cdot |L| \leq n+1$. Thus,

$$|L| \leq \frac{n+1}{2} < \frac{n}{2} + 1 \leq \left\lceil \frac{n}{2} \right\rceil + 1.$$

Since $|L|$ and $\left\lceil \frac{n}{2} \right\rceil + 1$ are integers, this yields

$$|L| \leq \left(\left\lceil \frac{n}{2} \right\rceil + 1 \right) - 1 = \left\lceil \frac{n}{2} \right\rceil.$$

This is what we wanted to prove. □

1.4.2. Intermezzo: SageMath

Now to Question **(b)**: How many lacunar subsets does $[n]$ have?

Let us try it for small values of n:

1.4.3. Counting lacunar subsets

Definition 1.4.5. If $k \in \mathbb{Z}$, then $[k]$ means the set $\{1, 2, \ldots, k\}$. If $k \leq 0$, then this is understood to be $\varnothing$.

Proposition 1.4.6. Let $n \in \{-1, 0, 1, \ldots\}$. Then,

$$(\#\text{ of lacunar subsets of } [n]) = f_{n+2}.$$

One way to prove this is by induction, similarly to how we counted k-element subsets of n. Indeed, we can call a subset of $[n]$

- **red** if it contains n, and
- **green** if it does not contain n.

Then, for each $n \geq 1$, we have

$$\begin{aligned}
&(\#\text{ of lacunar subsets of } [n]) \\
&= (\#\text{ of lacunar red subsets of } [n]) + (\#\text{ of lacunar green subsets of } [n]) \\
&= (\#\text{ of lacunar subsets of } [n-2]) + (\#\text{ of lacunar subsets of } [n-1]).
\end{aligned}$$

So the sequence of $(\#\text{ of lacunar subsets of } [n])$ follows the same recursion as the Fibonacci sequence. Now check the starting values and you're done.

Sketch of a different proof: There is a bijection between

$$\{\text{lacunar subsets of } [n]\} \qquad \text{and} \qquad \{\text{domino tilings of } R_{n+1,2}\}.$$

1.4.4. Counting k-element lacunar subsets

Let us answer part **(c)** of the question now:

Proposition 1.4.7. Let $n \in \mathbb{Z}$ and $k \in \mathbb{N}$ be such that $k \leq n+1$. Then,

$$(\#\text{ of } k\text{-element lacunar subsets of } [n]) = \binom{n+1-k}{k}.$$

This is plausible but not quite obvious.

One way to prove this proposition is by strong induction on n, just as we proved the previous proposition.

Here is a different proof – a *bijective proof*:

Proof. Define a map

$$A : \{k\text{-element lacunar subsets of } [n]\} \to \{k\text{-element subsets of } [n+1-k]\},$$
$$\{s_1 < s_2 < \cdots < s_k\} \mapsto \{s_1 - 0 < s_2 - 1 < s_3 - 2 < \cdots < s_k - (k-1)\}.$$

This is a bijection, with inverse

$$B : \{k\text{-element subsets of } [n+1-k]\} \to \{k\text{-element lacunar subsets of } [n]\},$$
$$\{t_1 < t_2 < \cdots < t_k\} \mapsto \{t_1 + 0 < t_2 + 1 < t_3 + 2 < \cdots < t_k + (k-1)\}.$$

Thus, the bijection principle yields

$$|\{k\text{-element lacunar subsets of } [n]\}|$$
$$= |\{k\text{-element subsets of } [n+1-k]\}| = \binom{n+1-k}{k}.$$

□

Now, we can also prove:

Proposition 1.4.8. Let $n \in \mathbb{N}$. Then, the Fibonacci number f_{n+1} is

$$f_{n+1} = \sum_{k=0}^{n} \binom{n-k}{k} = \binom{n-0}{0} + \binom{n-1}{1} + \binom{n-2}{2} + \cdots + \binom{n-n}{n}.$$

Proof. Consider the lacunar subsets of $[n-1]$. Their sizes range from 0 to n (actually, n can only be a size when $n=0$, but that's OK). So the sum principle yields

$$\begin{aligned}
&(\#\text{ of lacunar subsets of } [n-1]) \\
&= \sum_{k=0}^{n} \underbrace{(\#\text{ of } k\text{-element lacunar subsets of } [n-1])}_{= \binom{(n-1)+1-k}{k}} \\
&\qquad \text{(by previous proposition, applied to } n-1 \text{ instead of } n) \\
&= \sum_{k=0}^{n} \binom{(n-1)+1-k}{k} = \sum_{k=0}^{n} \binom{n-k}{k}.
\end{aligned}$$

But the pre-previous proposition yields

$$(\#\text{ of lacunar subsets of } [n-1]) = f_{(n-1)+2} = f_{n+1}.$$

Comparing these, we get

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

□

We have now answered all our three questions on lacunar subsets.

1.4.5. A little exercise

Exercise 1.4.1. Let n, a, b be nonnegative integers such that n is even. Find

$$(\#\text{ of subsets of } [n] \text{ that have } a \text{ even and } b \text{ odd elements}).$$

Solution idea: In order to choose such a subset, we have to choose two independent things:

- an a-element subset of $\{2,4,6,\ldots,n\}$;
- a b-element subset of $\{1,3,5,\ldots,n-1\}$.

The former can be chosen in $\binom{n/2}{a}$ many ways, while the latter can be chosen in $\binom{n/2}{b}$ many ways. So the total # of subsets is

$$\binom{n/2}{a} \cdot \binom{n/2}{b}.$$

Why exactly? This is the product rule in action. Formally speaking, there is a bijection

$$\{\text{subsets of } [n] \text{ that have } a \text{ even and } b \text{ odd elements}\}$$
$$\to \{a\text{-element subsets of } \{2,4,6,\ldots,n\}\} \times \{b\text{-element subsets of } \{1,3,5,\ldots,n-1\}\},$$
$$S \mapsto (S \cap \{2,4,6,\ldots,n\}, S \cap \{1,3,5,\ldots,n-1\}).$$

Class of 2019-09-09

1.4.6. The Fibonacci addition formula

Theorem 1.4.9. Let $m, n \in \mathbb{N}$. Then, $f_{m+n+1} = f_m f_n + f_{m+1} f_{n+1}$.

Proof. It is not hard to prove this by induction (see references in notes).
Proof by "double counting":
WLOG assume that $m, n > 0$.
How many lacunar subsets of $[m+n-1]$ are there?
1st method: One of our propositions says that the # is $f_{(m+n-1)+2} = f_{m+n+1}$.
2nd method: Call a lacunar subset S of $[m+n-1]$

- **red** if it contains m;
- **green** if it does not contain m.

A green lacunar subset of $[m+n-1]$ is just a union of

- a lacunar subset of $[m-1]$, and
- a lacunar subset of $\{m+1, m+2, \ldots, m+n-1\}$.

Conversely, any such union is a green lacunar subset of $[m+n-1]$.
Thus, we get a bijection

$$\{\text{green lacunar subsets of } [m+n-1]\}$$
$$\to \{\text{lacunar subsets of } [m-1]\} \times \{\text{lacunar subsets of } \{m+1, m+2, \ldots, m+n-1\}\}.$$

Thus, by the bijection principle and the product rule,

$$\begin{aligned}
&(\#\text{ of green lacunar subsets of } [m+n-1]) \\
&= \underbrace{(\#\text{ of lacunar subsets of } [m-1])}_{=f_{(m-1)+2}=f_{m+1}} \\
&\qquad \cdot \underbrace{(\#\text{ of lacunar subsets of } \{m+1, m+2, \ldots, m+n-1\})}_{\substack{=f_{(n-1)+1}=f_{n+1} \\ \text{(here, we have shifted our subsets} \\ \text{to the left by } m\text{)}}} \\
&= f_{m+1} f_{n+1}.
\end{aligned}$$

A red lacunar subset of $[m+n-1]$ is just a union of

- a lacunar subset of $[m-2]$,
- a lacunar subset of $\{m+2, m+3, \ldots, m+n-1\}$, and
- the one-element set $\{m\}$.

Conversely, any such union is a red lacunar subset of $[m+n-1]$.
Hence, we have a bijection

$$\{\text{red lacunar subsets of } [m+n-1]\}$$
$$\to \{\text{lacunar subsets of } [m-2]\} \times \{\text{lacunar subsets of } \{m+2, m+3, \ldots, m+n-1\}\}.$$

Thus, by the bijection principle and the product rule,

$$\begin{aligned}
&(\#\text{ of red lacunar subsets of } [m+n-1]) \\
&= \underbrace{(\#\text{ of lacunar subsets of } [m-2])}_{=f_{(m-2)+2}=f_m} \\
&\qquad \cdot \underbrace{(\#\text{ of lacunar subsets of } \{m+2, m+3, \ldots, m+n-1\})}_{\substack{=f_{n+1}=f_n \\ \text{(here, we have shifted our subsets} \\ \text{to the left by } m+1)}} \\
&= f_m f_n.
\end{aligned}$$

Hence,

$$\begin{aligned}
&(\#\text{ of lacunar subsets of } [m+n-1]) \\
&= \underbrace{(\#\text{ of red lacunar subsets of } [m+n-1])}_{=f_m f_n} \\
&\qquad + \underbrace{(\#\text{ of green lacunar subsets of } [m+n-1])}_{=f_{m+1} f_{n+1}} \\
&= f_m f_n + f_{m+1} f_{n+1}.
\end{aligned}$$

So we have computed $(\#\text{ of lacunar subsets of } [m+n-1])$ in two ways, and we have obtained f_{m+n+1} and $f_m f_n + f_{m+1} f_{n+1}$, respectively. Comparing these, we obtain $f_{m+n+1} = f_m f_n + f_{m+1} f_{n+1}$. □

See [Benjamin & Quinn] for more examples of "double counting".

1.4.7. Some counting exercises

Exercise 1.4.2. A set S of integers is said to be **self-counting** if $|S| \in S$.

(For example, $\{1,3,5\}$ is self-counting, since $|\{1,3,5\}| = 3 \in \{1,3,5\}$. But $\{4\}$ and $\{1,3\}$ and $\{1,2,4\}$ are not self-counting.)

Let n be a positive integer.

(a) For a given $k \in [n]$, how many self-counting k-element subsets does $[n]$ have?

(b) How many self-counting subsets does $[n]$ have?

Proof. **(a)** Fix $k \in [n]$. A k-element subset of $[n]$ is self-counting if and only if it contains k. So

$$\begin{aligned}
&(\#\text{ of self-counting } k\text{-element subsets of } [n]) \\
&= (\#\text{ of } k\text{-element subsets of } [n] \text{ that contain } k) \\
&= (\#\text{ of } (k-1)\text{-element subsets of } [n] \setminus \{k\}) \\
&\qquad \text{(by the bijection principle)} \\
&= \binom{n-1}{k-1} \qquad \text{(by the Combinatorial interpretation of BCs)}.
\end{aligned}$$

(b) Each self-counting subset of $[n]$ has size 1 or 2 or $\cdots$ or n. Thus, the sum rule shows that

$$\begin{aligned}
&(\#\text{ of self-counting subsets of } [n]) \\
&= \sum_{k=1}^{n} (\#\text{ of self-counting subsets of } [n] \text{ of size } k) \\
&= \sum_{k=1}^{n} \underbrace{(\#\text{ of self-counting } k\text{-element subsets of } [n])}_{= \binom{n-1}{k-1}} \\
&= \sum_{k=1}^{n} \binom{n-1}{k-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} \qquad \text{(here, we substituted } k \text{ for } k-1) \\
&= 2^{n-1} \qquad \left(\text{by the } \sum_{k=0}^{n} \binom{n}{k} = 2^n \text{ formula, applied to } n-1 \text{ instead of } n \right).
\end{aligned}$$

□

Exercise 1.4.3. A set S of integers is said to be **self-starting** if $|S|$ is the smallest element of S.

(For example, $\{1\}$ and $\{2,5\}$ and $\{3,5,9\}$ are self-starting. But $\{4\}$ and $\{1,3\}$ and $\{1,2,4\}$ and $\{1,3,5\}$ are not self-starting.)

Let n be a positive integer.

(a) For a given $k \in [n]$, how many self-starting k-element subsets does $[n]$ have?

(b) How many self-starting subsets does $[n]$ have?

Proof. **(a)** Fix $k \in [n]$. Then, the self-starting k-element subsets of $[n]$ are just the k-element subsets of $[n]$ that contain k and whose all other elements are in

$\{k+1, k+2, \ldots, n\}$. Thus, we have

$$\begin{aligned}
&(\#\text{ of self-starting } k\text{-element subsets of } [n]) \\
&= \left(\begin{array}{c} \#\text{ of } k\text{-element subsets of } [n] \text{ that contain } k \\ \text{and whose all other elements are in } \{k+1, k+2, \ldots, n\} \end{array} \right) \\
&= (\#\text{ of } (k-1)\text{-element subsets of } \{k+1, k+2, \ldots, n\}) \\
&\qquad (\text{by the bijection principle}) \\
&= \binom{n-k}{k-1} \qquad (\text{by the Combinatorial interpretation of BCs}) .
\end{aligned}$$

(b) The sum rule again yields

$$\begin{aligned}
&(\#\text{ of self-starting subsets of } [n]) \\
&= \sum_{k=1}^{n} \underbrace{(\#\text{ of self-starting } k\text{-element subsets of } [n])}_{= \binom{n-k}{k-1}} \\
&= \sum_{k=1}^{n} \binom{n-k}{k-1} = \sum_{k=0}^{n-1} \binom{n-k-1}{k} \qquad (\text{here, we substituted } k \text{ for } k-1) \\
&= \sum_{k=0}^{n-1} \binom{n-1-k}{k} = f_n,
\end{aligned}$$

since one of our previous propositions said that

$$f_{n+1} = \sum_{k=0}^{n} \binom{n-k}{k}$$

(we have applied this to $n-1$ instead of n). □

Exercise 1.4.4. A set S of integers is said to be **self-ending** if $|S|$ is the largest element of S.

(For example, $\{1\}$ and $\{1,2,3\}$ are self-ending. But $\{4\}$ and $\{1,3\}$ and $\{1,2,4\}$ and $\{1,3,5\}$ are not self-starting.)

Let n be a positive integer.

(a) For a given $k \in [n]$, how many self-ending k-element subsets does $[n]$ have?

(b) How many self-ending subsets does $[n]$ have?

Proof. **(a)** Fix $k \in [n]$. The only self-ending k-element subset of $[n]$ is $[k]$. So the # of such subsets is 1.

(b) By the sum rule, the # is $\sum_{k=1}^{n} 1 = n$. □

1.4.8. Counting maps and tuples

Recall: If $A_1, A_2, \ldots, A_n$ are n sets, then $A_1 \times A_2 \times \cdots \times A_n$ means the set of all n-tuples

$$(a_1, a_2, \ldots, a_n) \qquad \text{with } a_1 \in A_1 \text{ and } a_2 \in A_2 \text{ and } \cdots \text{ and } a_n \in A_n.$$

This is called the **Cartesian product** of $A_1, A_2, \ldots, A_n$. If $n = 0$, then this is an empty Cartesian product, and is a 1-element set, consisting only of the 0-tuple $()$. If $n = 1$, then this is "more or less" the set A_1. Formally, this means that the map

$$\begin{aligned} &(\text{the set } A_1) \to (\text{the Cartesian product of the single set } A_1)\,, \\ &\qquad a_1 \mapsto (a_1) \end{aligned}$$

is a bijection.

The Cartesian is not literally associative: If A, B and C are three sets, then

$$A \times B \times C \qquad \text{and} \qquad (A \times B) \times C \qquad \text{and} \qquad A \times (B \times C)$$

are not the same set. Their elements have the forms

$$(a, b, c) \qquad \text{and} \qquad ((a, b), c) \qquad \text{and} \qquad (a, (b, c))\,.$$

But there are bijections between these sets, so at least they all have the same size.

More generally, if $A_1, A_2, \ldots, A_n$ are any n sets, and $k \in \{0, 1, \ldots, n\}$, then the sets

$$A_1 \times A_2 \times \cdots \times A_n \qquad \text{and} \qquad (A_1 \times A_2 \times \cdots \times A_k) \times (A_{k+1} \times A_{k+2} \times \cdots \times A_n)$$

have the same size.

Theorem 1.4.10. (The product rule for n sets)
Let $A_1, A_2, \ldots, A_n$ be any n finite sets. Then, $A_1 \times A_2 \times \cdots \times A_n$ is also finite and its size is

$$|A_1 \times A_2 \times \cdots \times A_n| = |A_1| \cdot |A_2| \cdot \cdots \cdot |A_n|\,.$$

Proof. Induction on n. The induction base is $|\{()\}| = 1 =$ (empty product). The induction step uses the product rule for 2 sets. □

Intuitively, this is just saying that when choosing an n-tuple, we can choose each of its entries separately, and then the total # of options is the product of the #s of choices at each step.

Now, this theorem has a corollary about counting maps between sets:

Theorem 1.4.11. (The power rule)
Let A and B be two finite sets. Let $B^A = \{\text{maps } A \to B\}$. Then,

$$\left|B^A\right| = |B|^{|A|}.$$

Proof. Write A as $A = \{a_1, a_2, \ldots, a_k\}$ with k distinct elements $a_1, a_2, \ldots, a_k$. Then, each map $f : A \to B$ can be uniquely written in two-line notation as follows:

$$f = \begin{pmatrix} a_1 & a_2 & \cdots & a_k \\ b_1 & b_2 & \cdots & b_k \end{pmatrix}$$

for some $b_1, b_2, \ldots, b_k \in B$. Thus, there a bijection

$$\begin{aligned} \{\text{maps } A \to B\} &\to \underbrace{B \times B \times \cdots \times B}_{k \text{ times}}, \\ f &\mapsto (f(a_1), f(a_2), \ldots, f(a_k)). \end{aligned}$$

Thus, the bijection principle yields

$$\begin{aligned} &(\#\text{ of maps } A \to B) \\ &= \left|\underbrace{B \times B \times \cdots \times B}_{k \text{ times}}\right| = \underbrace{|B| \cdot |B| \cdot \cdots \cdot |B|}_{k \text{ times}} = |B|^k = |B|^{|A|} \end{aligned}$$

(since $k = |A|$). In other words, $\left|B^A\right| = |B|^{|A|}$. □

For example, the # of maps from $[3]$ to $[5]$ is $5^3 = 125$.

In one of the next chapters, we will count certain kinds of maps (injective, surjective, ...).

Class of 2019-09-11

1.5. Interchanging summation signs

Theorem 1.5.1. (Finite Fubini's principle)
Let X and Y be two finite sets. For each pair $(x, y) \in X \times Y$, let $a_{(x,y)}$ be a number. Then,

$$\sum_{x \in X} \sum_{y \in Y} a_{(x,y)} = \sum_{(x,y) \in X \times Y} a_{(x,y)} = \sum_{y \in Y} \sum_{x \in X} a_{(x,y)}.$$

Example 1.5.2. Let $X = [n]$ and $Y = [m]$ for some $n, m \in \mathbb{N}$. Then, this becomes

$$\sum_{x\in[n]} \sum_{y\in[m]} a_{(x,y)} = \sum_{(x,y)\in[n]\times[m]} a_{(x,y)} = \sum_{y\in[m]} \sum_{x\in[n]} a_{(x,y)},$$

i.e.

$$\sum_{x=1}^{n} \sum_{y=1}^{m} a_{(x,y)} = \sum_{(x,y)\in[n]\times[m]} a_{(x,y)} = \sum_{y=1}^{m} \sum_{x=1}^{n} a_{(x,y)},$$

i.e.

$$\begin{aligned}
&\left(a_{(1,1)} + a_{(1,2)} + \cdots + a_{(1,m)}\right) \\
&\qquad + \left(a_{(2,1)} + a_{(2,2)} + \cdots + a_{(2,m)}\right) \\
&\qquad + \cdots \\
&\qquad + \left(a_{(n,1)} + a_{(n,2)} + \cdots + a_{(n,m)}\right) \\
&= \left(\text{the sum of } \mathbf{all}\ a_{(x,y)}\right) \\
&= \left(a_{(1,1)} + a_{(2,1)} + \cdots + a_{(n,1)}\right) \\
&\qquad + \left(a_{(1,2)} + a_{(2,2)} + \cdots + a_{(n,2)}\right) \\
&\qquad + \cdots \\
&\qquad + \left(a_{(1,m)} + a_{(2,m)} + \cdots + a_{(n,m)}\right).
\end{aligned}$$

In other words, if we have a rectangular table of numbers:

$$\begin{pmatrix}
a_{(1,1)} & a_{(1,2)} & \cdots & a_{(1,m)} \\
a_{(2,1)} & a_{(2,2)} & \cdots & a_{(2,m)} \\
\vdots & \vdots & \ddots & \vdots \\
a_{(n,1)} & a_{(n,2)} & \cdots & a_{(n,m)}
\end{pmatrix},$$

then the following three procedures lead to the same result:

- summing together the m entries of each row, and then summing these "row tallies";
- summing all nm entries of the table;
- summing together the n entries of each column, and then summing these "column tallies".

Fubini's principle allows us to swap any two summation signs that stand near each other and don't "depend" upon each other.

Exercise 1.5.1. Let S be an n-element set. Find the sum of the sizes of all subsets of S, that is,

$$\sum_{T \subseteq S} |T| .$$

We shall prove this using the following simple fact:

Proposition 1.5.3. ("Counting by roll call") Let S be a finite set. Let T be a subset of S. Then,

$$|T| = \sum_{s \in S} [s \in T] .$$

Proof of Proposition.

$$\begin{aligned} \sum_{s \in S} [s \in T] &= \sum_{\substack{s \in S; \\ s \in T}} \underbrace{[s \in T]}_{=1} + \sum_{\substack{s \in S; \\ s \notin T}} \underbrace{[s \in T]}_{=0} \\ &= \sum_{\substack{s \in S; \\ s \in T}} 1 + \underbrace{\sum_{\substack{s \in S; \\ s \notin T}} 0}_{=0} = \sum_{\substack{s \in S; \\ s \in T}} 1 = \sum_{s \in S \cap T} 1 = |S \cap T| = |T| , \end{aligned}$$

since $T \subseteq S$ yields $S \cap T = T$. □

Solution of the exercise. Applying the Proposition, we have

$$\sum_{T \subseteq S} \underbrace{|T|}_{= \sum_{s \in S} [s \in T]} = \sum_{T \subseteq S} \sum_{s \in S} [s \in T] = \sum_{s \in S} \sum_{T \subseteq S} [s \in T]$$

$$\left(\begin{array}{c} \text{we interchanged the summation signs} \\ \text{using Fubini's principle} \end{array} \right) .$$

But for a given $s \in S$, we have

$$\begin{aligned} \sum_{T \subseteq S} [s \in T] &= \sum_{\substack{T \subseteq S; \\ s \in T}} \underbrace{[s \in T]}_{=1} + \sum_{\substack{T \subseteq S; \\ s \notin T}} \underbrace{[s \in T]}_{=0} \\ &= \sum_{\substack{T \subseteq S; \\ s \in T}} 1 = \sum_{\substack{T \text{ is a subset of } S \\ \text{that contains } s}} 1 = (\text{\# of subsets } T \text{ of } S \text{ that contain } s) \\ &= (\text{\# of subsets of } S \setminus \{s\}) \\ &\qquad \left(\begin{array}{c} \text{since there is a bijection} \\ \{\text{subsets } T \text{ of } S \text{ that contain } s\} \to \{\text{subsets of } S \setminus \{s\}\} \end{array} \right) \\ &= 2^{n-1} \qquad (\text{since } S \setminus \{s\} \text{ is an } (n-1)\text{-element set}) . \end{aligned}$$

Hence,

$$\sum_{T\subseteq S}|T| = \sum_{s\in S}\underbrace{\sum_{T\subseteq S}[s\in T]}_{=2^{n-1}} = \sum_{s\in S}2^{n-1} = \underbrace{|S|}_{=n}\cdot 2^{n-1} = n\cdot 2^{n-1}.$$

(2nd solution:) Each subset of S has size in $\{0,1,\ldots,n\}$. Thus, by the "splitting out" principle for sums,

$$\begin{aligned}\sum_{T\subseteq S}|T| &= \sum_{k=0}^{n}\sum_{\substack{T\subseteq S;\\ |T|=k}}\underbrace{|T|}_{=k} = \sum_{k=0}^{n}\underbrace{\sum_{\substack{T\subseteq S;\\ |T|=k}}k}_{=k\cdot(\#\text{ of }k\text{-element subsets of }S)}\\ &= \sum_{k=0}^{n}k\cdot\underbrace{(\#\text{ of }k\text{-element subsets of }S)}_{=\binom{n}{k}}\\ &\qquad\qquad\text{(by the Combinatorial interpretation of BCs)}\\ &= \sum_{k=0}^{n}k\binom{n}{k}.\end{aligned}$$

By comparing the results of the two solutions, we get the following identity for free:

$$\sum_{k=0}^{n}k\binom{n}{k} = n\cdot 2^{n-1}.$$

□

Some variants of Fubini's principle are:

Theorem 1.5.4. (Finite Fubini's principle with a predicate)
Let X and Y be two finite sets. Let $\mathcal{A}(x,y)$ be some statement for each $(x,y)\in X\times Y$. For each pair $(x,y)\in X\times Y$ satisfying $\mathcal{A}(x,y)$, let $a_{(x,y)}$ be a number. Then,

$$\sum_{x\in X}\sum_{\substack{y\in Y;\\ \mathcal{A}(x,y)}}a_{(x,y)} = \sum_{\substack{(x,y)\in X\times Y;\\ \mathcal{A}(x,y)}}a_{(x,y)} = \sum_{y\in Y}\sum_{\substack{x\in X;\\ \mathcal{A}(x,y)}}a_{(x,y)}.$$

For example:

$$\sum_{x=1}^{n}\sum_{\substack{y\in[m];\\ x+y\text{ is even}}}a_{(x,y)} = \sum_{\substack{(x,y)\in[n]\times[m];\\ x+y\text{ is even}}}a_{(x,y)} = \sum_{y=1}^{m}\sum_{\substack{x\in[n];\\ x+y\text{ is even}}}a_{(x,y)}.$$

Corollary 1.5.5. (Triangular Fubini's principle I)
Let $n \in \mathbb{N}$. For each pair $(x,y) \in [n] \times [n]$ with $x+y \leq n$, let $a_{(x,y)}$ be a number. Then,

$$\sum_{x=1}^{n} \sum_{y=1}^{n-x} a_{(x,y)} = \sum_{\substack{(x,y) \in [n] \times [n]; \\ x+y \leq n}} a_{(x,y)} = \sum_{y=1}^{n} \sum_{x=1}^{n-y} a_{(x,y)}.$$

Corollary 1.5.6. (Triangular Fubini's principle II)
Let $n \in \mathbb{N}$. For each pair $(x,y) \in [n] \times [n]$ with $x \leq y$, let $a_{(x,y)}$ be a number. Then,

$$\sum_{x=1}^{n} \sum_{y=1}^{x} a_{(x,y)} = \sum_{\substack{(x,y) \in [n] \times [n]; \\ x \leq y}} a_{(x,y)} = \sum_{y=1}^{n} \sum_{x=y}^{n} a_{(x,y)}.$$

1.6. Counting permutations – an introduction

There are two different things called "permutations" in mathematics. One is a kind of maps ("active permutations"); the other is a kind of lists ("passive permutations"). We are always going to use the word "permutation" for the former, at least in the context of "permutation of a set".

Definition 1.6.1. A **permutation** of a set X is a bijection from X to X.

For example,

$$\begin{pmatrix} 0 & 1 & 5 & 7 & 9 \\ 1 & 7 & 5 & 0 & 9 \end{pmatrix}$$

is a permutation of $\{0,1,5,7,9\}$.

Theorem 1.6.2. Let $n \in \mathbb{N}$. Let X be an n-element set. Then,

$$(\#\text{ of permutations of } X) = n!.$$

We shall prove this later.

Definition 1.6.3. A **derangement** of a set X means a permutation σ of X such that

$$\sigma(x) \neq x \qquad \text{for all } x \in X.$$

For example,

$$\begin{pmatrix} 0 & 1 & 5 & 7 & 9 \\ 1 & 7 & 9 & 0 & 5 \end{pmatrix}$$

is a derangement of $\{0,1,5,7,9\}$.

How many derangements does an n-element set have?

A first simplification: Instead of studying an arbitrary n-element set, it suffices to study $[n]$, because:

Lemma 1.6.4. Let X be any n-element set. Then,

$$(\#\text{ of derangements of } X) = (\#\text{ of derangements of } [n])\,.$$

Proof. The sets X and $[n]$ have the same number of elements.

Thus, there is a bijection $\phi : X \to [n]$. Fix such a ϕ.

Now, any derangement of X can be transformed into a derangement of $[n]$ by "relabeling" the elements of X as $1, 2, \ldots, n$ using ϕ.

For example, if X is a 3-element set $\{x, y, z\}$, and ϕ is the bijection $\begin{pmatrix} x & y & z \\ 1 & 2 & 3 \end{pmatrix} : X \to [3]$, then the derangement

$$\omega = \begin{pmatrix} x & y & z \\ y & z & x \end{pmatrix} \qquad \text{of } X$$

becomes the derangement

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \qquad \text{of } [3]\,.$$

Essentially, what we do is: we take the two-line notation of our derangement of X, and replace each element of X by its label in $[n]$.

More formally: We consider the map

$$\{\text{derangements of } X\} \to \{\text{derangements of } [n]\}\,,$$
$$\omega \mapsto \phi \circ \omega \circ \phi^{-1}.$$

(Keep in mind: The composition sign is defined so that $\alpha \circ \beta$ means "apply β first, then α".) This map is a bijection; its inverse is

$$\{\text{derangements of } [n]\} \to \{\text{derangements of } X\}\,,$$
$$\alpha \mapsto \phi^{-1} \circ \alpha \circ \phi.$$

Thus, by the bijection principle,

$$(\#\text{ of derangements of } X) = (\#\text{ of derangements of } [n])\,.$$

□

So it suffices to study derangements of $[n]$.

Definition 1.6.5. For each $n \in \mathbb{N}$, let

$$D_n = (\#\text{ of derangements of } [n]).$$

Definition 1.6.6. Let $n \in \mathbb{N}$. The **one-line notation** for a permutation σ of $[n]$ is the n-tuple $(\sigma(1), \sigma(2), \ldots, \sigma(n))$.

(**Warning:** This notation looks like the cycle notation in algebra textbooks, but is not the cycle notation from algebra textbooks.)

Example 1.6.7. The permutations of $[3]$ are (in one-line notation)

$$(1,2,3), \quad (1,3,2), \quad (2,1,3), \quad (2,3,1), \quad (3,1,2), \quad (3,2,1).$$

In two-line notation, these are

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

The only derangements among these are

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

Thus, $D_3 = 2$.

Example 1.6.8.

$D_0 = 1$ (since $\mathrm{id} : \varnothing \to \varnothing$ is a derangement);
$D_1 = 0$ (since $\mathrm{id} : [1] \to [1]$ is **not** a derangement);
$D_2 = 1$ $\left(\text{since the only derangement of } [2] \text{ is } \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\right)$;
$D_3 = 2$;
$D_4 = 9$.

https://oeis.org suggested A000166 Subfactorial or rencontres numbers, or derangements: number of permutations of n elements with no fixed points.

Theorem 1.6.9. **(a)** We have $D_n = (n-1)(D_{n-1} + D_{n-2})$ for all $n \geq 2$.
(b) We have $D_n = nD_{n-1} + (-1)^n$ for all $n \geq 1$.

(c) We have

$$n! = \sum_{k=0}^{n} \binom{n}{k} D_{n-k} \qquad \text{for all } n \in \mathbb{N}.$$

(d) We have

$$D_n = \sum_{k=0}^{n} (-1)^k \frac{n!}{k!} \qquad \text{for all } n \in \mathbb{N}.$$

(e) We have

$$D_n = \operatorname{round}\left(\frac{n!}{e}\right) \qquad \text{for all } n \geq 1,$$

where $e = \sum_{k=0}^{\infty} \frac{1}{k!} \approx 2.718\ldots$ and $\operatorname{round}(x) = \left\lfloor x + \frac{1}{2} \right\rfloor$.

Some of these formulas will be proven later. (Part **(b)** follows from **(a)** via a HW1 exercise.)

Class of 2019-09-16

1.6.1. Short-legged permutations

Definition 1.6.10. Let $n \in \mathbb{N}$. A permutation σ of $[n]$ is said to be **short-legged** if each $i \in [n]$ satisfies $|\sigma(i) - i| \leq 1$.

Question: How many short-legged permutations does $[n]$ have?

Example 1.6.11. For $n = 2$, the short-legged permutations of $[2]$ are (in one-line notation)

$$(1,2), \qquad (2,1).$$

For $n = 3$, the short-legged permutations of $[3]$ are (in one-line notation)

$$(1,2,3), \qquad (1,3,2), \qquad (2,1,3).$$

For $n = 4$, the short-legged permutations of $[4]$ are (in one-line notation)

$$(1,2,3,4), \qquad (1,2,4,3), \qquad (1,3,2,4), \qquad (2,1,3,4), \qquad (2,1,4,3).$$

(We have used "backtracking".)

Proposition 1.6.12. Let $n \in \mathbb{N}$. Then,

$$(\#\text{ of short-legged permutations of } [n]) = f_{n+1}.$$

First proof. Strong induction on n; see notes. (Main idea: We can equate each permutation σ of $[n]$ with its one-line notation $(\sigma(1), \sigma(2), \ldots, \sigma(n))$. If σ is short-legged, then the last entry of this one-line notation is $\sigma(n) \in \{n, n-1\}$.)

[Second proof.] Bijection

$$\{\text{short-legged permutations of } [n]\} \to \{\text{domino tilings of } R_{n,2}\}$$

(see blackboard). □

1.6.2. Long-legged permutations

Definition 1.6.13. Let $n \in \mathbb{N}$. A permutation σ of $[n]$ is said to be **long-legged** if each $i \in [n]$ satisfies $|\sigma(i) - i| > 1$.

Question: How many long-legged permutations does $[n]$ have?

This is Sequence A001883 in the OEIS.

2. Binomial coefficients

Let us study binomial coefficients more systematically, and see some techniques for manipulating sums and products of them without going into combinatorics.

2.1. The alternating sum of a row of Pascal's triangle

Proposition 2.1.1. Let $n \in \mathbb{N}$. Then,

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = [n = 0].$$

First proof. Apply the binomial formula to -1 and 1. You get

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

Now compare this with $((-1)+1)^n = 0^n = [n = 0]$. □

For the second proof, we will use the **telescope principle**:

Theorem 2.1.2. (Telescoping sum principle).

Let u and v be two integers with $u \leq v+1$. Let $a_u, a_{u+1}, \ldots, a_v$ be any numbers. Then,

$$\sum_{j=u}^{v} \left(a_{j+1} - a_j\right) = a_{v+1} - a_u.$$

Proof of Theorem (outline). We have

$$\begin{aligned}\sum_{j=u}^{v}\left(a_{j+1}-a_j\right) &= \left(a_{u+1}-a_u\right)+\left(a_{u+2}-a_{u+1}\right)+\left(a_{u+3}-a_{u+2}\right)+\cdots+\left(a_{v+1}-a_v\right)\\ &= a_{v+1}-a_u \qquad \left(\text{since all other addends cancel}\right).\end{aligned}$$

Formally, this can be proven by induction. See also HW0, where this was proved for $u = 0$ (but of course, the general case can be reduced to this by shifting the index).

[Second proof of Proposition.] WLOG assume $n > 0$ (since the $n = 0$ case is just straightforward verification). Now,

$$\begin{aligned}&\sum_{k=0}^{n}(-1)^k \underbrace{\binom{n}{k}}_{=\binom{n-1}{k-1}+\binom{n-1}{k}\ \text{(by the recurrence of the BCs)}}\\ &= \sum_{k=0}^{n}(-1)^k\left(\binom{n-1}{k-1}+\binom{n-1}{k}\right)\\ &= \sum_{k=0}^{n}\left(\underbrace{(-1)^k}_{=-(-1)^{k-1}}\binom{n-1}{k-1}+(-1)^k\binom{n-1}{k}\right)\\ &= \sum_{k=0}^{n}\left(-(-1)^{k-1}\binom{n-1}{k-1}+(-1)^k\binom{n-1}{k}\right)\\ &= \sum_{k=0}^{n}\left((-1)^k\binom{n-1}{k}-(-1)^{k-1}\binom{n-1}{k-1}\right)\\ &= \sum_{j=-1}^{n-1}\left((-1)^{j+1}\binom{n-1}{j+1}-(-1)^j\binom{n-1}{j}\right)\\ &\qquad \left(\text{here, we substituted } j+1 \text{ for } k\right)\\ &= (-1)^n \underbrace{\binom{n-1}{n}}_{\substack{=0\\ (\text{since } n-1\in\mathbb{N} \text{ and } n>n-1)}} -(-1)^{-1}\underbrace{\binom{n-1}{-1}}_{=0}\\ &= 0-0=0.\end{aligned}$$

□

Third proof of Proposition. WLOG assume $n > 0$ (since the $n = 0$ case is just straightforward verification). We thus need to show that

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

Rewrite $\sum_{k=0}^{n} (-1)^k \binom{n}{k}$ as

$$\sum_{k \text{ even}} \binom{n}{k} - \sum_{k \text{ odd}} \binom{n}{k}$$

(where both sums range only over $k \in \{0, 1, \ldots, n\}$). Thus, we only need to prove that

$$\sum_{k \text{ even}} \binom{n}{k} = \sum_{k \text{ odd}} \binom{n}{k}. \tag{22}$$

We can prove this bijectively:

The LHS of (1) is the # of subsets of $[n]$ of even size (by the Combinatorial interpretation of BCs).

The RHS of (1) is the # of subsets of $[n]$ of odd size.

Thus, we need a bijection

$$\{\text{subsets of } [n] \text{ of even size}\} \to \{\text{subsets of } [n] \text{ of odd size}\}.$$

One bijection we can use for this is

$$A \mapsto \begin{cases} A \cup \{1\}, & \text{if } 1 \notin A; \\ A \setminus \{1\}, & \text{if } 1 \in A. \end{cases}$$

This map is a bijection (and its inverse is given by the same formula). This proves (1). Thus, the Proposition is proved. □

Remark 2.1.3. If X and Y are two sets, then the **symmetric difference** $X \triangle Y$ of X and Y is defined to be the set

$$\begin{aligned} X \triangle Y &:= \{\text{all elements that lie in exactly one of } X \text{ and } Y\} \\ &= (X \setminus Y) \cup (Y \setminus X) = (X \cup Y) \setminus (X \cap Y). \end{aligned}$$

This concept has the following properties:

$$\begin{aligned} X \triangle X &= \varnothing; \\ X \triangle \varnothing &= X; \\ X \triangle Y &= Y \triangle X; \\ (X \triangle Y) \triangle Z &= X \triangle (Y \triangle Z); \\ X \cap (Y \triangle Z) &= (X \cap Y) \triangle (X \cap Z). \end{aligned}$$

The bijection we used in the above third proof is thus given by

$$A \mapsto A \triangle \{1\}.$$

2.2. The trinomial revision formula

Proposition 2.2.1. (Trinomial revision)
Let $n, a, b \in \mathbb{R}$. Then,

$$\binom{n}{a}\binom{a}{b} = \binom{n}{b}\binom{n-b}{a-b}.$$

Proof. Four cases are possible:
Case 1: $b \notin \mathbb{N}$.
Case 2: $b \in \mathbb{N}$ and $a \notin \mathbb{N}$.
Case 3: $b \in \mathbb{N}$ and $a \in \mathbb{N}$ but $a < b$.
Case 4: $b \in \mathbb{N}$ and $a \in \mathbb{N}$ and $a \geq b$.

With a bit of busywork, you can check that the formula boils down to $0 = 0$ in the first three cases. Thus, we only need to do Case 4. Here, b, a and $a - b$ all belong to $\mathbb{N}$. Hence, $\dbinom{a}{b} = \dfrac{a!}{b!\,(a-b)!}$. Now,

$$\begin{aligned}\binom{n}{b}\binom{n-b}{a-b} &= \frac{n\,(n-1)\,(n-2)\cdots(n-b+1)}{b!} \cdot \frac{(n-b)\,(n-b-1)\,(n-b-2)\cdots(n-a+1)}{(a-b)!} \\ &= \frac{n\,(n-1)\,(n-2)\cdots(n-a+1)}{b!\,(a-b)!}.\end{aligned}$$

Comparing this with

$$\begin{aligned}\binom{n}{a}\binom{a}{b} &= \frac{n\,(n-1)\,(n-2)\cdots(n-a+1)}{a!} \cdot \frac{a!}{b!\,(a-b)!} \\ &= \frac{n\,(n-1)\,(n-2)\cdots(n-a+1)}{b!\,(a-b)!},\end{aligned}$$

we obtain

$$\binom{n}{a}\binom{a}{b} = \binom{n}{b}\binom{n-b}{a-b}.$$

□

Corollary 2.2.2. (Absorption)
Let $m \in \mathbb{R}$ and $n \in \mathbb{R} \setminus \{0\}$. We have

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

Proof. Apply trinomial revision to m, n and 1 instead of n, a and b. You get

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

But recall $\binom{n}{1} = n$ and $\binom{m}{1} = m$; so this becomes

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

Now divide both sides by n. □

Remark 2.2.3. The trinomial revision formula

$$\binom{n}{a}\binom{a}{b} = \binom{n}{b}\binom{n-b}{a-b}$$

has a nice bijective proof at least in the case when $n \in \mathbb{N}$. Can you find it?

Class of 2019-09-18

Bijective proof of the trinomial revision formula for $n \in \mathbb{N}$. Assume $n \in \mathbb{N}$.

Fix a set N of n people.

Then, $\binom{n}{a}\binom{a}{b}$ is the # of ways to choose a committee of a people from N and then to choose a subcommittee of b people from this committee.

On the other hand, $\binom{n}{b}\binom{n-b}{a-b}$ is the # of ways to choose the same thing, but in a different way: first choose the subcommittee, then choose $a-b$ further people to add to it, forming the committee. □

If we want to formalize this argument, we run into trouble: We would want to apply the product rule, so we want to interpret $\binom{n}{a}\binom{a}{b}$ as $|X \times Y|$ for two sets X and Y. We expect $X = \{a\text{-element subsets of } N\}$ and $Y = \{b\text{-element subsets of an } a\text{-element set}\}$, which does not work ($Y$ would depend on the choice of an a-element set). In other words, our two choices are dependent: the choice of the subcommittee depends on the choice of the committee. Our product rule is not made for this.

The easiest way to correctly formalize this argument is by using the sum rule:

Theorem 2.2.4. Let S and W be finite sets. Let $f : S \to W$ be a map. Then,

$$|S| = \sum_{w \in W} \left(\#\text{ of } s \in S \text{ with } f(s) = w\right).$$

Above proof of trinomial revision for $n \in \mathbb{N}$, rewritten formally. Fix an n-element set N.

Let

$$S = \{(B, A) \mid B \subseteq A \subseteq N \text{ and } |B| = b \text{ and } |A| = a\}.$$

We want to compute $|S|$ in two ways.

First way:

$$\begin{aligned}
|S| &= \sum_{\substack{C\subseteq N;\\ |C|=a}} \underbrace{(\#\text{ of } (B,A)\in S \text{ with } A=C)}_{=(\#\text{ of } b\text{-element subsets of } C)} \\
&= \sum_{\substack{C\subseteq N;\\ |C|=a}} \underbrace{(\#\text{ of } b\text{-element subsets of } C)}_{=\binom{|C|}{b}=\binom{a}{b} \text{ (since } |C|=a)} \\
&= \sum_{\substack{C\subseteq N;\\ |C|=a}} \binom{a}{b} = \underbrace{(\#\text{ of } C\subseteq N \text{ satisfying } |C|=a)}_{=(\#\text{ of } a\text{-element subsets of } N)=\binom{n}{a}} \cdot \binom{a}{b} \\
&= \binom{n}{a}\binom{a}{b}.
\end{aligned}$$

Second way:

$$\begin{aligned}
|S| &= \sum_{\substack{D\subseteq N;\\ |D|=b}} \underbrace{(\#\text{ of } (B,A)\in S \text{ with } A=D)}_{\substack{=(\#\text{ of } a\text{-element sets } A \text{ such that } D\subseteq A\subseteq N)\\ =(\#\text{ of } (a-b)\text{-element subsets of } N\setminus D)\\ \text{(by the bijection principle)}}} \\
&= \sum_{\substack{D\subseteq N;\\ |D|=b}} \underbrace{(\#\text{ of } (a-b)\text{-element subsets of } N\setminus D)}_{=\binom{|N\setminus D|}{a-b}=\binom{n-b}{a-b} \text{ (since } |N\setminus D|=n-b)} \\
&= \sum_{\substack{D\subseteq N;\\ |D|=b}} \binom{n-b}{a-b} = \underbrace{(\#\text{ of } D\subseteq N \text{ such that } |D|=b)}_{=\binom{n}{b}} \cdot \binom{n-b}{a-b} \\
&= \binom{n}{b}\binom{n-b}{a-b}.
\end{aligned}$$

Comparing these results, we get

$$\binom{n}{a}\binom{a}{b} = \binom{n}{b}\binom{n-b}{a-b}.$$

□

Three remarks:

- What we did in our bijective proof is a kind of "product principle" that is not actually the product principle we stated. In the product principle we stated,

we are counting things that can be determined by 2 (or more) **independent** choices. Here, we have 2 choices of which the second depends on the first, but only the **number of options** for the second does not depend on the first. So we had to use the sum principle instead of the product principle.

There **is** a version of the product principle that allows for such dependent choices. Informally speaking, it says that if you are doing a sequence of k choices, such that you have a_1 options in choice 1, then you have a_2 options in choice 2 (no matter what you chose in choice 1), then you have a_3 options in choice 3 (no matter what you chose in choices 1 and 2), and so on, then the total # of options will be $a_1 a_2 a_3 \cdots a_k$. This is called the **dependent choice principle**, and is formalized in [Loehr, §1.8] and [Newstead, Strategy 6.2.21].

I will not formalize this principle, because in practice I find it easier to just apply the sum rule (and induction on k if necessary).

- We have assumed that $n \in \mathbb{N}$, but we have not assumed that $a, b \in \mathbb{N}$. How were to able to avoid those latter assumptions? Didn't we use $\binom{a}{b} =$ (# of b-element subsets of an a-element set)? Didn't this require $a \in \mathbb{N}$?

  No, because we were doing these arguments only in relation to an already fixed a-element subset of N. Of course, given an a-element subset of N, you automatically know that $a \in \mathbb{N}$. If $a \notin \mathbb{N}$, then there is no a-element subset of N, so we are manipulating an empty sum.

- What about the case $n \notin \mathbb{N}$? Can we salvage our above proof of

$$\binom{n}{a}\binom{a}{b} = \binom{n}{b}\binom{n-b}{a-b}$$

  to also work in this case?

  There is a way, but it requires us to change our point of view. See one of the later sections.

2.3. The hockey-stick formula revisited

Recall the hockey-stick formula: It says that

$$\binom{0}{k} + \binom{1}{k} + \cdots + \binom{n}{k} = \binom{n+1}{k+1}$$

for all $n, k \in \mathbb{N}$. We proved this algebraically. Let us see three more proofs:

2nd proof of the hockey-stick formula. The recurrence of the BCs yields

$$\begin{aligned}
\binom{n+1}{k+1} &= \binom{n}{k} + \underbrace{\binom{n}{k+1}}_{=\binom{n-1}{k}+\binom{n-1}{k+1}} \\
&= \binom{n}{k} + \binom{n-1}{k} + \underbrace{\binom{n-1}{k+1}}_{=\binom{n-2}{k}+\binom{n-2}{k+1}} \\
&= \binom{n}{k} + \binom{n-1}{k} + \binom{n-2}{k} + \underbrace{\binom{n-2}{k+1}}_{=\binom{n-3}{k}+\binom{n-3}{k+1}} \\
&= \binom{n}{k} + \binom{n-1}{k} + \binom{n-2}{k} + \binom{n-3}{k} + \binom{n-3}{k+1} \\
&= \cdots\cdots\cdots \qquad \text{(keep doing the same transformation)} \\
&= \binom{n}{k} + \binom{n-1}{k} + \binom{n-2}{k} + \cdots + \binom{0}{k} + \underbrace{\binom{0}{k+1}}_{\substack{=0 \\ \text{(since } k+1>0\text{)}}} \\
&= \binom{n}{k} + \binom{n-1}{k} + \binom{n-2}{k} + \cdots + \binom{0}{k} \\
&= \binom{0}{k} + \binom{1}{k} + \cdots + \binom{n}{k}.
\end{aligned}$$

This can be easily formalized as an induction proof. □

3rd proof of the hockey-stick formula. For each $i \in \mathbb{Z}$, we have

$$\binom{i}{k} = \binom{i+1}{k+1} - \binom{i}{k+1}, \qquad \text{since } \binom{i+1}{k+1} = \binom{i}{k} + \binom{i}{k+1}.$$

Thus,

$$\begin{aligned}
\binom{0}{k} + \binom{1}{k} + \cdots + \binom{n}{k} &= \sum_{i=0}^{n} \underbrace{\binom{i}{k}}_{= \binom{i+1}{k+1} - \binom{i}{k+1}} \\
&= \sum_{i=0}^{n} \left(\binom{i+1}{k+1} - \binom{i}{k+1} \right) \\
&= \binom{n+1}{k+1} - \underbrace{\binom{0}{k+1}}_{=0} \qquad \text{(by the telescoping sum theorem)} \\
&= \binom{n+1}{k+1}.
\end{aligned}$$

□

4th proof of the hockey-stick formula. Recall that

$$\begin{aligned}
&\binom{n+1}{k+1} \\
&= (\#\text{ of } (k+1)\text{-element subsets of } [n+1]) \\
&= \sum_{j=1}^{n+1} (\#\text{ of } (k+1)\text{-element subsets of } [n+1] \text{ whose largest element is } j).
\end{aligned}$$

Now, fix $j \in [n+1]$. How many $(k+1)$-element subsets of $[n+1]$ have the property that their largest element is j ? The answer is $\binom{j-1}{k}$, since the element j has already been chosen, and the remaining k elements must be chosen from $\{1, 2, \ldots, j-1\} = [j-1]$, which has $j-1$ elements. (Formally, this is saying that the map

$$\begin{aligned}
&\{(k+1)\text{-element subsets of } [n+1] \text{ whose largest element is } j\} \\
&\to \{k\text{-element subsets of } [j-1]\}
\end{aligned}$$

that removes j from each set is a bijection.) Thus, the above computation becomes

$$\begin{aligned}
&\binom{n+1}{k+1} \\
&= \sum_{j=1}^{n+1} \underbrace{(\#\text{ of } (k+1)\text{-element subsets of } [n+1] \text{ whose largest element is } j)}_{=\binom{j-1}{k}} \\
&= \sum_{j=1}^{n+1} \binom{j-1}{k} = \sum_{i=0}^{n} \binom{i}{k} \qquad \text{(here, we have substituted } i \text{ for } j-1) \\
&= \binom{0}{k} + \binom{1}{k} + \cdots + \binom{n}{k}.
\end{aligned}$$

□

2.4. Counting maps

Theorem 2.4.1. Let $m, n \in \mathbb{N}$. Let A be an m-element set. Let B be an n-element set. Then,

$$(\#\text{ of maps from } A \text{ to } B) = n^m.$$

Proof.

$$\begin{aligned}
(\#\text{ of maps from } A \text{ to } B) &= |\{\text{maps } A \to B\}| = \left|B^A\right| = |B|^{|A|} \qquad \text{(as we proved above)} \\
&= n^m \qquad (\text{since } |B| = n \text{ and } |A| = m).
\end{aligned}$$

□

Now, what if we want to count not all maps, but only some maps?

2.4.1. Injective maps

Recall the definition of the lower factorial:

Definition 2.4.2. Let $n \in \mathbb{R}$ and $m \in \mathbb{N}$. Then,

$$n^{\underline{m}} = n\,(n-1)\,(n-2)\cdots(n-m+1).$$

Theorem 2.4.3. Let $m, n \in \mathbb{N}$. Let A be an m-element set. Let B be an n-element set. Then,

$$(\#\text{ of injective maps from } A \text{ to } B) = n^{\underline{m}}.$$

Remark 2.4.4. (a) If $m = 0$, then the RHS is $n^{\underline{0}} = (\text{empty product}) = 1$. And indeed, there is exactly one injective map from $A = \varnothing$ to B (namely, the "empty map", which sends nothing anywhere).

(b) If $m > n$, then the RHS is

$$n^{\underline{m}} = n(n-1)(n-2)\cdots\underbrace{(n-n)}_{=0}\cdots(n-m+1) = 0.$$

And thus, the theorem yields that there are no injective maps from A to B in this case. This is intuitively clear; we are later going to state this as "Pigeonhole Principle for Injections".

Informal proof of the Theorem. Let $a_1, a_2, \ldots, a_m$ be the m elements of A (listed without repetitions). Then, in order to construct a map f from A to B, we only have to choose the values $f(a_1), f(a_2), \ldots, f(a_m)$. Moreover, to ensure that this map f is injective, we need to choose these values to be distinct. We can construct such an injective map f as follows:

- choose $f(a_1)$ (there are n options for this, since we want $f(a_1) \in B$);
- choose $f(a_2)$ to be distinct from $f(a_1)$ (there are $n-1$ options for this, since we want $f(a_2) \in B \setminus \{f(a_1)\}$);
- choose $f(a_3)$ to be distinct from $f(a_1)$ and $f(a_2)$ (there are $n-2$ options for this, since we want $f(a_3) \in B \setminus \underbrace{\{f(a_1), f(a_2)\}}_{\substack{\text{a 2-element set} \\ \text{(since } f(a_1) \text{ and } f(a_2) \text{ are distinct)}}}$);
- choose $f(a_4)$ to be distinct from $f(a_1)$, $f(a_2)$ and $f(a_3)$ (there are $n-3$ options for this, since we want $f(a_4) \in B \setminus \underbrace{\{f(a_1), f(a_2), f(a_3)\}}_{\substack{\text{a 3-element set} \\ \text{(since } f(a_1), f(a_2) \text{ and } f(a_3) \text{ are distinct)}}}$);
- $\ldots$;
- last step: choose $f(a_m)$; there are $n - (m-1)$ options.

Thus, by the dependent choice principle, the total # of options is

$$\begin{aligned} &n(n-1)(n-2)\cdots(n-(m-1)) \\ &= n(n-1)(n-2)\cdots(n-m+1) = n^{\underline{m}}. \end{aligned}$$

□

Class of 2019-09-21

How do we make this proof rigorous? By induction on the # of choices (which is m in our case).

Notation: We let $\operatorname{Inj}(A, B)$ denote the set of all injective maps from A to B.

Thus, the theorem we are proving states that $|\operatorname{Inj}(A, B)| = n^{\underline{m}}$, where $n = |B|$ and $m = |A|$.

Proof. (Formal proof.) We proceed by induction on m.

Induction base: $m = 0$. Here, $A = \varnothing$. This is trivial (see Remark **(a)** above).

Induction step: Let $k \in \mathbb{N}$. Assume (as the IH) that the theorem holds for $m = k$. We must now prove it for $m = k + 1$.

Let A be a $(k+1)$-element set, and let B be an n-element set for some $n \in \mathbb{N}$. We must show that $|\operatorname{Inj}(A, B)| = n^{\underline{k+1}}$.

Fix $a \in A$. (This exists because $|A| = k + 1 > k \geq 0$.)

The set $A \setminus \{a\}$ has size $|A \setminus \{a\}| = |A| - 1 = (k+1) - 1 = k$. Thus, the IH yields

$$|\operatorname{Inj}(A \setminus \{a\}, B)| = n^{\underline{k}}.$$

The map

$$\begin{aligned} R : \operatorname{Inj}(A, B) &\to \operatorname{Inj}(A \setminus \{a\}, B), \\ h &\mapsto h|_{A \setminus \{a\}} \end{aligned}$$

is well-defined (since a restriction of an injective map to a subset is still injective).

> *Observation 1:* Let $g \in \operatorname{Inj}(A \setminus \{a\}, B)$. Then, there are precisely $n - k$ many maps $h \in \operatorname{Inj}(A, B)$ such that $R(h) = g$.

[*Proof of Observation 1:* To construct an $h \in \operatorname{Inj}(A, B)$ such that $R(h) = g$, we only need to choose a value $h(a)$. This value $h(a)$ must be distinct from all values of g (in order for h to be injective); there are precisely k values to avoid (since g is injective and thus takes exactly k different values). Thus, in total, there are $n - k$ many options for $h(a)$. This proves Observation 1.

(This argument, too, can be formalized: We have a bijection

$$\begin{aligned} \{h \in \operatorname{Inj}(A, B) \mid R(h) = g\} &\to B \setminus g(A), \\ h &\mapsto h(a). \end{aligned}$$

You can easily check this, as well as that $|B \setminus g(A)| = n - k$.)]

Now,

$$\begin{aligned}
|\operatorname{Inj}(A,B)| &= (\#\text{ of all } h \in \operatorname{Inj}(A,B)) \\
&= \sum_{g \in \operatorname{Inj}(A\setminus\{a\},B)} \underbrace{(\#\text{ of all } h \in \operatorname{Inj}(A,B) \text{ such that } R(h) = g)}_{\substack{=n-k \\ \text{(by Observation 1)}}} \\
&\qquad \text{(by the sum rule)} \\
&= \sum_{g \in \operatorname{Inj}(A\setminus\{a\},B)} (n-k) = \underbrace{|\operatorname{Inj}(A \setminus \{a\}, B)|}_{=n^{\underline{k}}} \cdot (n-k) \\
&= n^{\underline{k}} \cdot (n-k) = n(n-1)(n-2)\cdots(n-k+1)\cdot(n-k) \\
&= n(n-1)(n-2)\cdots(n-k) = n^{\underline{k+1}}.
\end{aligned}$$

This completes the induction step. □

Theorem 2.4.5. (Pigeonhole Principle for Injections)
Let $f : A \to B$ be an injective map between finite sets. Then:
(a) $|A| \leq |B|$.
(b) If $|A| = |B|$, then f is bijective.

Proof. Let $a_1, a_2, \ldots, a_k$ be the elements of A (listed without repetitions). Thus, $|A| = k$. Now, $f(a_1), f(a_2), \ldots, f(a_k)$ are distinct (since f is injective). Hence,

$$|\{f(a_1), f(a_2), \ldots, f(a_k)\}| = k = |A|.$$

But $\{f(a_1), f(a_2), \ldots, f(a_k)\} \subseteq B$, so that

$$|\{f(a_1), f(a_2), \ldots, f(a_k)\}| \leq |B|.$$

Hence, $|A| \leq |B|$. This proves part **(a)**.

For part **(b)**, assume that $|A| = |B|$. Then, the above inequality must be an equality, and this entails $\{f(a_1), f(a_2), \ldots, f(a_k)\} = B$. But this means that f is surjective. Thus, f is bijective. □

Theorem 2.4.6. (Pigeonhole Principle for Surjections)
Let $f : A \to B$ be an surjective map between finite sets. Then:
(a) $|A| \geq |B|$.
(b) If $|A| = |B|$, then f is bijective.

Proof. Let $a_1, a_2, \ldots, a_k$ be the elements of A (listed without repetitions). Thus, $|A| = k$. Now,

$$|\{f(a_1), f(a_2), \ldots, f(a_k)\}| \leq k = |A|.$$

But $\{f(a_1), f(a_2), \ldots, f(a_k)\} = B$ (since f is surjective), so that

$$|\{f(a_1), f(a_2), \ldots, f(a_k)\}| = |B|.$$

Hence, $|A| \geq |B|$. This proves part **(a)**.

(b) Let $|A| = |B|$. Then, the inequalities above must be equalities. Thus, $|\{f(a_1), f(a_2), \ldots, f(a_k)\}| = k$. Thus, $f(a_1), f(a_2), \ldots, f(a_k)$ are distinct (since otherwise, we could omit one of them and would get the same set; but this would entail $|\{f(a_1), f(a_2), \ldots, f(a_k)\}| \leq k-1 < k$). Hence, f is injective. Thus, f is bijective. □

(These pigeonhole principles are combinatorial analogues of parts of the "Inverse Matrix Theorem" from linear algebra – that long theorem that gives many different criteria for a matrix to be invertible.)

Remark 2.4.7. The parts **(b)** of both Pigeonhole Principles are false if the sets A and B are not finite. For example:

- The map $\mathbb{N} \to \mathbb{N}$, $i \mapsto i+1$ is injective, but not bijective.
- The map $\mathbb{N} \to \mathbb{N}$, $i \mapsto \begin{cases} 0, & \text{if } i = 0; \\ i-1, & \text{if } i > 0 \end{cases}$ is surjective, but not bijective.

We can now prove something that we claimed in the previous chapter:

Corollary 2.4.8. Let X be a finite set. Then:

$$(\#\text{ of permutations of } X) = |X|!.$$

Proof. The Pigeonhole Principle for Injections shows that every injective map from X to X is bijective. The converse also holds (by definition). Thus,

$$\begin{aligned} &\{\text{injective maps } X \to X\} \\ &= \{\text{bijective maps } X \to X\} \\ &= \{\text{permutations of } X\}. \end{aligned}$$

Hence,

$$\begin{aligned} &(\#\text{ of injective maps } X \to X) \\ &= (\#\text{ of permutations of } X). \end{aligned}$$

But the LHS is (by the previous Theorem, applied to $m = |X|$, $n = |X|$, $A = X$ and $B = X$) equal to

$$\begin{aligned} |X|^{\underline{|X|}} &= |X| \cdot (|X|-1) \cdot (|X|-2) \cdot \cdots \cdot \underbrace{(|X| - |X| + 1)}_{=1} \\ &= |X| \cdot (|X|-1) \cdot (|X|-2) \cdot \cdots \cdot 1 = |X|!. \end{aligned}$$

Thus, the RHS is $|X|!$ as well. □

After having counted injective maps, let us count surjective maps.

Definition 2.4.9. Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Then, $\operatorname{sur}(m,n)$ means the # of surjective maps from $[m]$ to $[n]$.

Proposition 2.4.10. Let $m, n \in \mathbb{N}$. Let A be an m-element set. Let B be an n-element set. Then,

$$(\#\text{ of surjective maps from } A \text{ to } B) = \operatorname{sur}(m,n).$$

Proof. This is a relabeling argument, just as our proof of that lemma about derangements: Relabel the m elements of A as $1, 2, \ldots, m$, and relabel the n elements of B as $1, 2, \ldots, n$. Details are LTTR. □

Proposition 2.4.11. **(a)** $\operatorname{sur}(m,0) = [m = 0]$ for all $m \in \mathbb{N}$.
(b) $\operatorname{sur}(m,1) = [m \neq 0] = 1 - [m = 0]$ for all $m \in \mathbb{N}$.
(c) $\operatorname{sur}(m,2) = 2^m - 2 + [m = 0]$ for all $m \in \mathbb{N}$.
(d) $\operatorname{sur}(0,k) = [k = 0]$ for all $k \in \mathbb{N}$.
(e) $\operatorname{sur}(1,k) = [k = 1]$ for all $k \in \mathbb{N}$.
(f) $\operatorname{sur}(m,n) = 0$ if $m < n$.

Proof. **(a)** If $m \neq 0$, then there are no surjections from $[m]$ to $[0]$ (since the arrows have nowhere to point to). If $m = 0$, then there is exactly 1.

(b) There is always exactly 1 map from $[m]$ to $[1]$. It is surjective if and only if $m \neq 0$.

(c), **(d)** and **(e)** are LTTR.

(f) Part **(a)** of the Pigeonhole Principle for Surjections says that surjections from $[m]$ to $[n]$ can only exist when $m \geq n$; thus their # is 0 when $m < n$. □

Now, let us look for a recursive formula for $\operatorname{sur}(m,n)$.

1st approach: Fix $m \in \mathbb{N}$ and $n > 0$.

Given a surjective map $f : [m] \to [n]$, we let J_f be the set of all $i \in [m]$ such that

$f(i) = n$. Clearly, $J_f \neq \varnothing$, since f is surjective. Thus, the sum rule yields

$$\begin{aligned}
&(\#\text{ of all surjective maps } f : [m] \to [n]) \\
&= \sum_{\substack{J \subseteq [m]; \\ J \neq \varnothing}} \underbrace{(\#\text{ of all surjective maps } f : [m] \to [n] \text{ with } J_f = J)}_{\substack{=(\#\text{ of all surjective maps } [m]\setminus J \to [n-1]) \\ \text{(since there is a bijection} \\ \text{from } \{\text{surjective maps } f:[m]\to[n] \text{ with } J_f = J\} \\ \text{to } \{\text{surjective maps } [m]\setminus J \to [n-1]\} \\ \text{which sends each } f \text{ to } f|_{[m]\setminus J})}} \\
&= \sum_{\substack{J \subseteq [m]; \\ J \neq \varnothing}} \underbrace{(\#\text{ of all surjective maps } [m] \setminus J \to [n-1])}_{\substack{=\operatorname{sur}(|[m]\setminus J|, |[n-1]|) \\ \text{(by the previous proposition)}}} \\
&= \sum_{\substack{J \subseteq [m]; \\ J \neq \varnothing}} \operatorname{sur}\left(\underbrace{|[m] \setminus J|}_{=m-|J|}, \underbrace{|[n-1]|}_{=n-1} \right) = \sum_{\substack{J \subseteq [m]; \\ J \neq \varnothing}} \operatorname{sur}(m - |J|, n-1) \\
&= \sum_{j=1}^{m} \sum_{\substack{J \subseteq [m]; \\ J \neq \varnothing; \\ |J| = j}} \operatorname{sur}\left(m - \underbrace{|J|}_{=j}, n-1 \right) \\
&= \sum_{j=1}^{m} \underbrace{\sum_{\substack{J \subseteq [m]; \\ J \neq \varnothing; \\ |J| = j}} \operatorname{sur}(m-j, n-1)}_{=(\#\text{ of } J \subseteq [m] \text{ such that } J \neq \varnothing \text{ and } |J| = j) \cdot \operatorname{sur}(m-j, n-1)} \\
&= \sum_{j=1}^{m} \underbrace{(\#\text{ of } J \subseteq [m] \text{ such that } J \neq \varnothing \text{ and } |J| = j)}_{\substack{=(\#\text{ of } J \subseteq [m] \text{ such that } |J| = j) \\ = \binom{m}{j}}} \cdot \operatorname{sur}(m-j, n-1) \\
&= \sum_{j=1}^{m} \binom{m}{j} \cdot \operatorname{sur}(m-j, n-1) = \sum_{j=0}^{m-1} \underbrace{\binom{m}{m-j}}_{\substack{= \binom{m}{j} \\ \text{(by symmetry of BCs)}}} \cdot \operatorname{sur}\left(\underbrace{m - (m-j)}_{=j}, n-1 \right) \\
&\qquad \text{(here, we have substituted } m-j \text{ for } j) \\
&= \sum_{j=0}^{m-1} \binom{m}{j} \operatorname{sur}(j, n-1).
\end{aligned}$$

Thus, we have proved the following:

Proposition 2.4.12. Let $m \in \mathbb{N}$ and $n > 0$. Then,

$$\operatorname{sur}(m,n) = \sum_{j=1}^{m} \binom{m}{j} \cdot \operatorname{sur}(m-j, n-1) = \sum_{j=0}^{m-1} \binom{m}{j} \operatorname{sur}(j, n-1).$$

Using this proposition and the previous one, it is easy to compute $\operatorname{sur}(m,n)$ recursively.

2nd approach: Fix $m > 0$ and $n > 0$. Classify the surjections $[m] \to [n]$ according to the image of m.

A surjection $f : [m] \to [n]$ is called

- **red** if $f(m) = f(i)$ for some $i < m$;
- **green** if it is not red (i.e., if $f(m) \neq f(i)$ for all $i < m$).

[*Examples:* If $m = 4$ and $n = 3$, then

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 2 \end{pmatrix} \qquad \text{is red,}$$

while

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 2 & 3 \end{pmatrix} \qquad \text{is green.}$$

]

Class of 2019-09-23

Thus, if a surjection ("surjection" = "surjective map") $f : [m] \to [n]$ is red, then $f|_{[m-1]}$ is "still" a surjection $[m-1] \to [n]$. Thus, a red surjection $f : [m] \to [n]$ can be viewed as a surjection $[m-1] \to [n]$. Thus, we get the following algorithm for constructing a red surjection $f : [m] \to [n]$:

- first, we choose $f(m)$ (there are n choices);
- then, we choose $f(1), f(2), \ldots, f(m-1)$; in other words, we choose the restriction $f|_{[m-1]}$ (there are $\operatorname{sur}(m-1,n)$ choices).

Hence,

$$(\#\text{ of red surjections } f : [m] \to [n]) = n \cdot \operatorname{sur}(m-1,n).$$

On the other hand, if a surjection $f : [m] \to [n]$ is green, then the restriction $f|_{[m-1]}$ has image $[n] \setminus \{f(m)\}$, so it can be viewed as a surjection $[m-1] \to [n] \setminus \{f(m)\}$. Hence, we get the following algorithm for constructing a green surjection $f : [m] \to [n]$:

- first, we choose $f(m)$ (there are n choices);
- then, we choose $f(1), f(2), \ldots, f(m-1)$; in other words, we choose the restriction $f|_{[m-1]}$ (there are $\operatorname{sur}(m-1, n-1)$ choices, because $f|_{[m-1]}$ needs to be a surjection $[m-1] \to [n] \setminus \{f(m)\}$).

Hence,

$$(\#\text{ of green surjections } f : [m] \to [n]) = n \cdot \operatorname{sur}(m-1, n-1).$$

Hence,

$$\begin{aligned}
&\operatorname{sur}(m, n) \\
&= (\#\text{ of surjections } f : [m] \to [n]) \\
&= \underbrace{(\#\text{ of red surjections } f : [m] \to [n])}_{= n \cdot \operatorname{sur}(m-1, n)} \\
&\qquad + \underbrace{(\#\text{ of green surjections } f : [m] \to [n])}_{= n \cdot \operatorname{sur}(m-1, n-1)} \\
&= n \cdot \operatorname{sur}(m-1, n) + n \cdot \operatorname{sur}(m-1, n-1) \\
&= n \cdot (\operatorname{sur}(m-1, n) + \operatorname{sur}(m-1, n-1)).
\end{aligned}$$

Thus, we have proved:

Proposition 2.4.13. Let m and n be positive integers. Then,

$$\operatorname{sur}(m, n) = n \cdot (\operatorname{sur}(m-1, n) + \operatorname{sur}(m-1, n-1)).$$

This is an even better recursion than the previous one; it is almost as simple as that of Pascal's triangle.

Corollary 2.4.14. **(a)** We have $\operatorname{sur}(n, n) = n!$ for all $n \in \mathbb{N}$.
(b) $\operatorname{sur}(m, n)$ is a multiple of $n!$ for all $n \in \mathbb{N}$ and $m \in \mathbb{N}$.

First proof. Both parts are easily shown by induction using the previous proposition.

[Second proof.] **(a)** The surjections $[n] \to [n]$ are bijections (by the Pigeonhole Principle for Surjections). Thus, they are precisely the permutations of $[n]$. Hence, their number is $n!$.

(b) Rough idea (we are, so far, missing the language for formalizing it): Each surjection $f : [m] \to [n]$ gives a way of "grouping" the elements of $[m]$ into n nonempty (disjoint) groups (= subsets). For example, the surjection $[6] \to [3]$ given in two-line notation as

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 2 & 3 & 3 & 2 \end{pmatrix}$$

yields the groups

$$\underbrace{\{2\}}_{\substack{\text{the elements}\\ \text{sent to 1}}}, \qquad \underbrace{\{1,3,6\}}_{\substack{\text{the elements}\\ \text{sent to 2}}}, \qquad \underbrace{\{4,5\}}_{\substack{\text{the elements}\\ \text{sent to 3}}}.$$

If we disregard which group correspond to which value (so we forget the values $f(i)$, but only keep track of which i and j satisfy $f(i) = f(j)$), then the # of all possible such groupings is $\operatorname{sur}(m,n)/n!$. Thus, $\operatorname{sur}(m,n)/n! \in \mathbb{N}$.

For the details of this argument, see [18s-hw3s] section §0.3 on "set partitions". □

Remark 2.4.15. The number $\operatorname{sur}(m,n)/n!$ is often denoted $\left\{ {m \atop n} \right\}$ and is called a **Stirling number of the 2nd kind**.

Here is the most explicit formula for $\operatorname{sur}(m,n)$ known:

Theorem 2.4.16. Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Then,

$$\operatorname{sur}(m,n) = \sum_{i=0}^{n} (-1)^{n-i} \binom{n}{i} i^m.$$

1st proof. Use induction & the first recursion we found for $\operatorname{sur}(m,n)$. For details, see [17f-hw2s] Exercise 4.

[2nd proof.] Use the Principle of Inclusion and Exclusion (see later). For details, see [18s-hw3s] Exercise 2 **(b)**. □

Applying this theorem to $n = 3$, we find

$$\begin{aligned}\operatorname{sur}(m,3) &= -\underbrace{0^m}_{=[m=0]} + 3 \cdot \underbrace{1^m}_{=1} - 3 \cdot 2^m + 3^m \\ &= 3^m - 3 \cdot 2^m + 3 - [m=0].\end{aligned}$$

2.5. $1^m + 2^m + ... + n^m$

Next goal: prove our theorem we stated previously about $1^m + 2^m + \cdots + n^m$. First step:

Theorem 2.5.1. Let $k \in \mathbb{N}$ and $m \in \mathbb{N}$. Then,

$$k^m = \sum_{i=0}^{m} \operatorname{sur}(m,i) \binom{k}{i}.$$

Proof. Double counting. How many maps $f : [m] \to [k]$ are there?

1st way: k^m.

2nd way: We construct a map $f : [m] \to [k]$ by the following method:

- First, we choose the number $|f([m])|$ (this is the size of the image of f, i.e., the # of distinct values of f). This is an integer in $\{0, 1, \ldots, m\}$. Call this integer i.
- Then, we choose the set $f([m])$ (this is the set of all values of f). There are $\binom{k}{i}$ many choices for this (since $f([m])$ must be an i-element subset of $[k]$).
- Finally, we choose $f(1), f(2), \ldots, f(m)$. These m numbers must be chosen from the (already determined) set $f([m])$, and must cover this set. Thus, there are $\operatorname{sur}(m, i)$ many choices here (since we are just choosing a surjection from $[m]$ to the already chosen i-element set $f([m])$).

Thus, the total # of ways is $\sum_{i=0}^{m} \binom{k}{i} \operatorname{sur}(m, i)$.

Since our two answers answer the same question, we obtain

$$k^m = \sum_{i=0}^{m} \binom{k}{i} \operatorname{sur}(m, i) = \sum_{i=0}^{m} \operatorname{sur}(m, i) \binom{k}{i}.$$

□

Theorem 2.5.2. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Then,

$$\sum_{k=0}^{n} k^m = \sum_{i=0}^{m} \operatorname{sur}(m, i) \cdot \binom{n+1}{i+1}.$$

Proof. We have

$$\sum_{k=0}^{n} \underbrace{k^m}_{\substack{=\sum_{i=0}^{k} \operatorname{sur}(m,i)\binom{k}{i} \\ \text{(by the previous theorem)}}} = \underbrace{\sum_{k=0}^{n}\sum_{i=0}^{m}}_{\substack{=\sum_{i=0}^{m}\sum_{k=0}^{n} \\ \text{(by Fubini)}}} \operatorname{sur}(m, i) \binom{k}{i}$$

$$= \sum_{i=0}^{m} \sum_{k=0}^{n} \operatorname{sur}(m, i) \binom{k}{i} = \sum_{i=0}^{m} \operatorname{sur}(m, i) \underbrace{\sum_{k=0}^{n} \binom{k}{i}}_{\substack{=\binom{n+1}{i+1} \\ \text{(by the hockey-stick identity)}}}$$

$$= \sum_{i=0}^{m} \operatorname{sur}(m, i) \cdot \binom{n+1}{i+1}.$$

□

Proof of Theorem 1.2.10. Rename k as m. Thus, we must prove

$$\sum_{i=1}^{n} i^m = \sum_{i=0}^{m} \operatorname{sur}(m,i) \cdot \binom{n+1}{i+1}.$$

But this is exactly what the previous theorem claims, since

$$\sum_{k=0}^{n} k^m = \underbrace{0^m}_{\substack{=0 \\ (\text{since } m>0)}} + 1^m + 2^m + \cdots + n^m = 1^m + 2^m + \cdots + n^m = \sum_{i=1}^{n} i^m.$$

□

2.6. The Vandermonde convolution

Class of 2019-09-25

Theorem 2.6.1. (The Vandermonde convolution, or the Chu–Vandermonde identity.)

Let $n \in \mathbb{N}$ and $x, y \in \mathbb{R}$. Then,

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

Remark 2.6.2. The "$\sum_k$" on the right hand side here means a sum over all $k \in \mathbb{Z}$. Note that $\binom{x}{k}\binom{y}{n-k} = 0$ whenever $k \notin \{0,1,\ldots,n\}$ (indeed, $\binom{y}{n-k} = 0$ if $k > n$, and $\binom{x}{k} = 0$ if $k < 0$). Thus, the 2nd equality sign in the Theorem is clear.

In the future, I will use the symbol $\overset{0}{=}$ for "equal because of zero addends". For example,

$$\sum_{k=0}^{n} k \overset{0}{=} \sum_{k=1}^{n} k \qquad \text{and} \qquad \sum_{k=0}^{m} \binom{n}{k} \overset{0}{=} \sum_{k=n}^{m} \binom{n}{k} \qquad \text{and}$$
$$\sum_{k=0}^{n} \binom{x}{k}\binom{y}{n-k} \overset{0}{=} \sum_{k} \binom{x}{k}\binom{y}{n-k}.$$

We will give two proofs of the theorem and cite another, but none of the proofs will be complete yet.

Proof. (First proof of Vandermonde convolution for $x \in \mathbb{N}$.)
For any $u \in \mathbb{R}$ and $v \in \mathbb{N}$, we have

$$\begin{aligned}
&\binom{u}{n} \\
&= \underbrace{\binom{u-1}{n-1}}_{=\binom{u-2}{n-2}+\binom{u-2}{n-1}} + \underbrace{\binom{u-1}{n}}_{=\binom{u-2}{n-1}+\binom{u-1}{n}} \\
&= \underbrace{\binom{u-2}{n-2}}_{=\binom{u-3}{n-3}+\binom{u-3}{n-2}} + 2\underbrace{\binom{u-2}{n-1}}_{=\binom{u-3}{n-2}+\binom{u-3}{n-1}} + \underbrace{\binom{u-1}{n}}_{=\binom{u-3}{n-1}+\binom{u-3}{n}} \\
&= \binom{u-3}{n-3} + 3\binom{u-3}{n-2} + 3\binom{u-3}{n-1} + \binom{u-3}{n} \\
&= \binom{u-4}{n-4} + 4\binom{u-4}{n-3} + 6\binom{u-4}{n-2} + 4\binom{u-4}{n-1} + \binom{u-4}{n} \\
&= \cdots\cdots \qquad \text{(keep applying the recurrence like this)} \\
&= \binom{u-v}{n-v} + \cdots + \binom{v}{k}\binom{u-v}{n-k} + \cdots + \binom{u-v}{n} \qquad \text{(this is what we get after } v \text{ steps)} \\
&\qquad \left(\begin{array}{c} \text{convince yourself that the coefficients follow the same} \\ \text{recursion as Pascal's triangle, and thus are } \binom{v}{k} \end{array} \right) \\
&= \sum_{k=0}^{v} \binom{v}{k}\binom{u-v}{n-k}.
\end{aligned}$$

Applying this to $u = x + y$ and $v = x$, we get

$$\begin{aligned}
\binom{x+y}{n} &= \sum_{k=0}^{x} \binom{x}{k}\binom{(x+y)-x}{n-k} = \sum_{k=0}^{x} \binom{x}{k}\binom{y}{n-k} \\
&\overset{0}{=} \sum_{k\in\mathbb{N}} \binom{x}{k}\binom{y}{n-k} \overset{0}{=} \sum_{k=0}^{n} \binom{x}{k}\binom{y}{n-k}.
\end{aligned}$$

This proves the Vandermonde convolution for $x \in \mathbb{N}$.
(This was informal, but formally it's just an induction on v.) □

Proof. (2nd proof of Vandermonde convolution for $x, y \in \mathbb{N}$.)
Double counting. How many ways are there to choose an n-element subset of $\{1, 2, \ldots, x\} \cup \{-1, -2, \ldots, -y\}$? Let us answer this in two ways:
1st way: $\binom{x+y}{n}$.

2nd way: First, decide how many positive elements our subset will have. Let's say it will have k positive elements (with $k \in \{0, 1, \ldots, n\}$). Then, choose these k positive elements (there are $\binom{x}{k}$ choices for them). Then, choose the remaining $n-k$ negative elements (there are $\binom{y}{n-k}$ choices for them). Thus, the answer is $\sum_{k=0}^{n} \binom{x}{k}\binom{y}{n-k}$.

Now, compare the two answers. This proves the Vandermonde convolution when $x, y \in \mathbb{N}$. □

Proof. (3rd proof of the Vandermonde convolution, for all $x, y \in \mathbb{R}$.) Induction on n, using the absorption identity $\binom{y}{n} = \frac{y}{n}\binom{y-1}{n-1}$. See [detnotes, first proof of Theorem 3.29]. □

Before I explain how the first two proofs above can be extended to the general case ($x, y \in \mathbb{R}$), let me draw a couple conclusions from the theorem:

Corollary 2.6.3. Let $x \in \mathbb{R}$ and $y \in \mathbb{N}$. Then,

$$\sum_{k=0}^{y} \binom{x}{k}\binom{y}{k} = \binom{x+y}{y}.$$

Proof. We have

$$\sum_{k=0}^{y} \binom{x}{k} \underbrace{\binom{y}{k}}_{=\binom{y}{y-k} \text{ (by symmetry)}} = \sum_{k=0}^{y} \binom{x}{k}\binom{y}{y-k} = \binom{x+y}{y}$$

(by Vandermonde convolution, applied to $n = y$). □

Corollary 2.6.4. Let $n \in \mathbb{N}$. Then,

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

Proof. Apply the previous corollary to $x = n$ and $y = n$. □

Remark 2.6.5. No formula for $\sum_{k=0}^{n} \binom{n}{k}^3$ is known.

However, there are such formulas for $\sum_{k=0}^{n} (-1)^k \binom{n}{k}^i$ with $i = 1, 2, 3$. We have already seen the formula for $i = 1$, and we will soon see the $i = 2$ one. (The formula for $i = 3$ is a special case of Dixon's identity.)

2.6.1. The polynomial identity trick

Next, I claim that just one trick suffices to complete the first two proofs of the Vandermonde convolution (i.e., to prove the theorem in full generality, not just for $x, y \in \mathbb{N}$). This is the "*polynomial identity trick*".

A reminder on polynomials: I will not formally define polynomials here (I might get to it later, in the chapter on generating functions). Polynomials are not functions, but you can substitute numbers into them. Informally, a polynomial (with real coefficients, in 1 variable X) is a "formal expression" of the form

$$\alpha X^a + \beta X^b + \gamma X^c + \cdots + \omega X^z \qquad \text{with } \alpha, \beta, \gamma, \ldots, \omega \in \mathbb{R} \text{ and } a, b, c, \ldots, z \in \mathbb{N}.$$

These expressions are understood to obey rules:

$$\varphi X^n + \psi X^n = (\varphi + \psi) X^n \qquad (\text{"combining like terms"});$$
$$0X^a \text{ can be removed;}$$
$$\text{terms can be swapped.}$$

We write X^0 as 1 (so αX^0 is written as α), and we write X^1 as X.

Addition of polynomials is defined by, e.g.,

$$\left(\alpha X^a + \beta X^b\right) + \left(\gamma X^c + \delta X^d\right) = \alpha X^a + \beta X^b + \gamma X^c + \delta X^d.$$

Subtraction is defined by

$$\left(\alpha X^a + \beta X^b\right) - \left(\gamma X^c + \delta X^d\right) = \alpha X^a + \beta X^b + (-\gamma) X^c + (-\delta) X^d.$$

Multiplication is defined by distributivity and $(\alpha X^a)\left(\beta X^b\right) = (\alpha\beta) X^{a+b}$. For example,

$$\begin{aligned}\left(\alpha X^a + \beta X^b\right)\left(\gamma X^c + \delta X^d\right) &= (\alpha X^a)(\gamma X^c) + (\alpha X^a)\left(\delta X^d\right) + \left(\beta X^b\right)(\gamma X^c) + \left(\beta X^b\right)\left(\delta X^d\right) \\ &= \alpha\gamma X^{a+c} + \alpha\delta X^{a+d} + \beta\gamma X^{b+c} + \beta\delta X^{b+d}.\end{aligned}$$

The **degree** of a polynomial is the largest exponent appearing in it with coefficient $\neq 0$. For example, the degree of $2X^2 + 7X^5 - X$ is 5. (We say that the degree of the polynomial 0 is $-\infty$.)

Substituting a number (or square matrix, or another polynomial) x into a polynomial $P = \alpha X^a + \beta X^b + \gamma X^c + \cdots$ yields $\alpha x^a + \beta x^b + \gamma x^c + \cdots$. This result is called $P(x)$.

A number x (say, a rational or real or complex number) is a **root** of a polynomial P if and only if $P(x) = 0$.

For a formal definition of polynomials, see [detnotes, §1.5] or [Loehr, Chapter 7] (most recommended) or most good algebra textbooks or the generating functions chapter of this course.

Theorem 2.6.6. (The "polynomial identity trick")

(a) A polynomial (with real coefficients, in 1 variable X) of degree $\leq n$ (for a given $n \in \mathbb{N}$) has $\leq n$ roots (in $\mathbb{Q}$, in $\mathbb{R}$, or in $\mathbb{C}$), unless it is the 0 polynomial (i.e., its coefficients are all 0).

(b) If a polynomial P has infinitely many roots, then P is the 0 polynomial.

(c) Let P and Q be polynomials. If

$$P(x) = Q(x) \qquad \text{for all } x \in \mathbb{N},$$

then $P = Q$.

Proof. See [Goodman, "Algebra: Abstract and Concrete", Corollary 1.8.24] for a proof of part **(a)**.

Part **(b)** follows from **(a)**.

Part **(c)** follows from **(b)**, applied to $P - Q$ instead of P. Indeed, if $P(x) = Q(x)$ for all $x \in \mathbb{N}$, then all nonnegative integers are roots of $P - Q$, and thus $P - Q$ has infinitely many roots, so that part **(b)** shows that $P - Q$ is the 0 polynomial. □

How does this help us prove the Vandermonde convolution?

Let us salvage our first proof of the Vandermonde convolution.

Fix $y \in \mathbb{R}$ and $n \in \mathbb{N}$. We have already proven

$$\binom{x+y}{n} = \sum_{k=0}^{n} \binom{x}{k}\binom{y}{n-k} \tag{23}$$

for all $x \in \mathbb{N}$. We want to prove this equality for all $x \in \mathbb{R}$.

Define two polynomials P and Q by

$$P = \binom{X+y}{n} \qquad \text{and} \qquad Q = \sum_{k=0}^{n} \binom{X}{k}\binom{y}{n-k}.$$

These are well-defined polynomials, since

$$P = \binom{X+y}{n} = \frac{(X+y)(X+y-1)(X+y-2)\cdots(X+y-n+1)}{n!}$$

and

$$Q = \sum_{k=0}^{n} \binom{X}{k}\binom{y}{n-k} = \sum_{k=0}^{n} \frac{X(X-1)(X-2)\cdots(X-k+1)}{k!}\binom{y}{n-k}.$$

Thus, $P(x) = Q(x)$ for all $x \in \mathbb{N}$ (since we have proved (7) for all $x \in \mathbb{N}$). Therefore, part **(c)** of the "polynomial identity trick" theorem yields $P = Q$. Thus, $P(x) = Q(x)$ for all $x \in \mathbb{R}$. In other words, (7) holds for all $x \in \mathbb{R}$. This completes the 1st proof of Vandermonde convolution.

We can also salvage our 2nd proof of Vandermonde convolution in a similar way. We need to apply our idea twice:

Step 1: Fix $y \in \mathbb{N}$ and $n \in \mathbb{N}$. Use the same argument as before to prove that (7) holds for all $x \in \mathbb{R}$.

Step 2: Fix $x \in \mathbb{R}$ and $n \in \mathbb{N}$. Use an analogous argument (using y instead of x) to prove that (7) holds for all $y \in \mathbb{R}$.

See [detnotes, Chapter 3] for details.

Remark. The polynomial identity trick can be applied to several other identities. For example:

- The trinomial revision identity says that

$$\binom{n}{a}\binom{a}{b} = \binom{n}{b}\binom{n-b}{a-b} \qquad \text{for all } n, a, b \in \mathbb{R}.$$

  We gave a combinatorial proof that required $n \in \mathbb{N}$. Now, using the polynomial identity trick, we can extend this to all $n \in \mathbb{R}$.

- We had the identity

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = [n = 0] \qquad \text{for all } n \in \mathbb{N}.$$

  This cannot be generalized to $n \in \mathbb{Q}$ or $n \in \mathbb{R}$, since n appears as the upper bound of the sum (and thus cannot be replaced by an X in that position).

  We can rewrite the above identity as

$$\sum_{k \in \mathbb{N}} (-1)^k \binom{n}{k} = [n = 0] \qquad \text{for all } n \in \mathbb{N}.$$

  This cannot be generalized to $n \in \mathbb{Q}$ or $n \in \mathbb{R}$, since the sum $\sum_{k \in \mathbb{N}} (-1)^k \binom{X}{k}$ is an ill-defined infinite sum (unlike $\sum_{k \in \mathbb{N}} (-1)^k \binom{n}{k}$ for each particular $n \in \mathbb{N}$), and also since $[n = 0]$ is not a polynomial in n.

Class of 2019-09-28

- Remember the theorem saying that

$$k^m = \sum_{i=0}^{m} \operatorname{sur}(m, i) \cdot \binom{k}{i} \qquad \text{for all } k, m \in \mathbb{N}.$$

Can we generalize this to all $k \in \mathbb{R}$? To all $m \in \mathbb{R}$?

To the second question: no, since k^m is not a polynomial in m (and for other reasons).

To the first question: yes, because for any fixed $m \in \mathbb{N}$, both sides are polynomials in k. So the theorem can be generalized to all $k \in \mathbb{R}$, even though our proof (by double counting) no longer works when $k \notin \mathbb{N}$.

- Recall the hockey-stick formula: It says that

$$\binom{0}{k} + \binom{1}{k} + \cdots + \binom{n}{k} = \binom{n+1}{k+1}$$

for all $n, k \in \mathbb{N}$. Can this be generalized to $n \in \mathbb{R}$? to $k \in \mathbb{R}$?

No to the first question, since n appears as a (tacit) summation bound on the LHS.

No to the second question, since $\binom{\text{constant}}{k}$ is not a polynomial in k.

And in fact, you can easily find counterexamples with $k = -1$.

You can play this game with any other identity you find.

2.6.2. Mutating the Chu–Vandermonde identity

We shall now see some "mutated" forms of the Chu–Vandermonde identity. "Mutated" means that these forms are obtained from Chu–Vandermonde by transforming the binomial coefficients using some simpler identities – in particular, the upper negation and symmetry identities. Recall these identities:

- UpNeg (= upper negation):

$$\binom{-n}{k} = (-1)^k \binom{n+k-1}{k} \qquad \text{for all } n \in \mathbb{R} \text{ and } k \in \mathbb{Z}.$$

- Symm (= symmetry):

$$\binom{n}{k} = \binom{n}{n-k} \qquad \text{for all } n \in \mathbb{N} \text{ and } k \in \mathbb{R}.$$

Now, an example of a "mutated" Chu–Vandermonde identity:

Theorem 2.6.7. ("upside-down Vandermonde convolution")
Let $n, x, y \in \mathbb{N}$. Then,

$$\binom{n+1}{x+y+1} = \sum_{k=0}^{n} \binom{k}{x} \binom{n-k}{y}.$$

Compare with the original Chu–Vandermonde identity, which says

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

However, there is an additional difference: In the upside-down identity, we really need $x, y \in \mathbb{N}$. (For example, $n = 3$, $x = -1$ and $y = 2$ would be a counterexample.)

First proof of "upside-down Vandermonde convolution". If $n < x + y$, then both sides are 0 (indeed, the LHS is clearly 0; furthermore, each addend on the RHS is 0, because if $k \in \{0, 1, \ldots, n\}$ such that $\binom{k}{x}\binom{n-k}{y} \neq 0$, then $\binom{k}{x} \neq 0 \Longrightarrow k \geq x$ and $\binom{n-k}{y} \neq 0 \Longrightarrow n-k \geq y$, and thus $n = \underbrace{k}_{\geq x} + \underbrace{(n-k)}_{\geq y} \geq x+y$, contradicting $n < x + y$).

So let us WLOG assume that $n \geq x+y$. Thus,

$$\sum_{k=0}^{n} \binom{k}{x}\binom{n-k}{y}$$

$$\overset{0}{=} \sum_{k=x}^{n-y} \underbrace{\binom{k}{x}}_{\overset{\text{Symm}}{=}\binom{k}{k-x} \overset{\text{UpNeg}}{=} (-1)^{k-x}\binom{-k+(k-x)-1}{k-x} = (-1)^{k-x}\binom{-x-1}{k-x}} \underbrace{\binom{n-k}{y}}_{\overset{\text{Symm}}{=}\binom{n-k}{n-k-y} \overset{\text{UpNeg}}{=} (-1)^{n-k-y}\binom{-(n-k)+(n-k-y)-1}{n-k-y} = (-1)^{n-k-y}\binom{-y-1}{n-k-y}}$$

$$= \sum_{k=x}^{n-y} \underbrace{(-1)^{k-x}(-1)^{n-k-y}}_{=(-1)^{n-x-y}} \binom{-x-1}{k-x}\binom{-y-1}{n-k-y}$$

$$= (-1)^{n-x-y} \sum_{k=x}^{n-y} \binom{-x-1}{k-x}\binom{-y-1}{n-k-y}$$

$$= (-1)^{n-x-y} \underbrace{\sum_{k=0}^{n-x-y} \binom{-x-1}{k}\binom{-y-1}{n-x-y-k}}_{\overset{\text{Vandermonde}}{=}\binom{(-x-1)+(-y-1)}{n-x-y}}$$

(here, we substituted k for $k-x$ in the sum)

$$= (-1)^{n-x-y} \underbrace{\binom{(-x-1)+(-y-1)}{n-x-y}}_{\overset{\text{UpNeg}}{=}(-1)^{n-x-y}\binom{-((-x-1)+(-y-1))+(n-x-y)-1}{n-x-y}}$$

$$= \underbrace{(-1)^{n-x-y}(-1)^{n-x-y}}_{=1} \underbrace{\binom{-((-x-1)+(-y-1))+(n-x-y)-1}{n-x-y}}_{=\binom{n+1}{n-x-y}}$$

$$= \binom{n+1}{n-x-y} \overset{\text{Symm}}{=} \binom{n+1}{(n+1)-(n-x-y)} = \binom{n+1}{x+y+1}.$$

□

Second proof of "upside-down Vandermonde". Double-count the # of $(x+y+1)$-element subsets of $[n+1]$.

1st way: $\binom{n+1}{x+y+1}$.

2nd way: Construct such a subset as follows:

- Choose the $(x+1)$-st smallest element of this subset. Call it $k+1$. Thus, $k \in \{0, 1, \ldots, n\}$.
- Choose the x smallest elements of this subset. There are $\binom{k}{x}$ options for this (since they need to be chosen from the k-element set $\{1, 2, \ldots, k\}$).
- Choose the y remaining elements of this subset. There are $\binom{n-k}{y}$ options for this (since they need to be chosen from the $(n-k)$-element set $\{k+2, k+3, \ldots, n+1\}$).

Thus, the total # of subsets is $\sum_{k=0}^{n} \binom{k}{x}\binom{n-k}{y}$.
Comparing the two answers, we get

$$\binom{n+1}{x+y+1} = \sum_{k=0}^{n} \binom{k}{x}\binom{n-k}{y}.$$

□

There are other identities that arise from "mutating" Chu–Vandermonde; see the notes for three examples.

2.7. Counting subsets again

Recall the theorem that said that if S is an n-element set and $k \in \mathbb{R}$, then

$$\binom{n}{k} = (\text{\# of } k\text{-element subsets of } S).$$

We have proved this by induction on n. Next, we shall show a new proof, which illustrates a rather common tactic. The underlying idea of this tactic is summarized by the saying "to count sheep, first count the legs and then divide by 4". In our situation, the "sheep" will be the k-element subsets of S, whereas the "legs" will be the **injective** k-tuples of elements of S – that is, k-tuples consisting of distinct elements of S. What makes this definition of "legs" useful is that

- the legs are easier to count than the sheep;
- each sheep has the same # of legs (although not 4).

This trick for counting is sometimes called "multijection principle" or "over-counting and correcting", but it requires no new principles.
Here are the details:

Definition 2.7.1. Let S be a set. Let $k \in \mathbb{N}$.

(a) A k-tuple $(s_1, s_2, \ldots, s_k) \in S^k$ is said to be **injective** if $s_1, s_2, \ldots, s_k$ are distinct.

(b) Let S^k_{inj} be the set of all injective k-tuples in S^k.

For example, the 3-tuple $(3, 2, 5)$ is injective, but the 3-tuple $(4, 2, 4)$ is not.

Let us count the legs:

Proposition 2.7.2. Let S be a finite set. Let $k \in \mathbb{N}$. Then,

$$\left|S^k_{\text{inj}}\right| = |S|^{\underline{k}} \qquad \text{(falling factorial)}.$$

Proof. One way to see this is to apply the dependent product rule.

Alternative: There is a bijection

$$\{\text{injective maps from } [k] \text{ to } S\} \to S^k_{\text{inj}},$$
$$f \mapsto (f(1), f(2), \ldots, f(k)).$$

The bijection principle thus yields

$$\left|S^k_{\text{inj}}\right| = (\#\text{ of injective maps from } [k] \text{ to } S) = |S|^{\underline{k}}$$

(by our theorem about counting injective maps). □

[**Aside:** Injective k-tuples need to have **all** their entries distinct. What about **non-stuttering** k-tuples, which only need to have **adjacent** entries distinct? I.e., we call a k-tuple $(s_1, s_2, \ldots, s_k)$ **non-stuttering** if $s_i \neq s_{i+1}$ for all $i \in [k-1]$. Then,

$$\left(\#\text{ of non-stuttering } k\text{-tuples in } S^k\right) = |S| \cdot (|S|-1)^{k-1} \qquad \text{for } k > 0.$$

These non-stuttering k-tuples are also known as Smirnov words or Carlitz words or ...]

Proof. (2nd proof of the theorem saying that

$$\binom{n}{k} = (\#\text{ of } k\text{-element subsets of } S).$$

)

WLOG assume $k \in \mathbb{N}$. Also, $|S| = n$, so the previous proposition yields

$$\left|S^k_{\text{inj}}\right| = |S|^{\underline{k}} = n^{\underline{k}}.$$

But

$$\left|S_{\text{inj}}^k\right| = \left(\#\text{ of injective } k\text{-tuples } \overrightarrow{s} \in S^k\right)$$
$$= \sum_{\substack{W\subseteq S;\\ |W|=k}} \left(\#\text{ of injective } k\text{-tuples } \overrightarrow{s} \in S^k \text{ such that the set of entries of } \overrightarrow{s} \text{ is } W\right)$$
$$\text{(by the sum rule)}.$$

Now:

> *Observation 1:* Let W be a k-element subset of S. Then, the injective k-tuples $\overrightarrow{s} \in S^k$ such that the set of entries of $\overrightarrow{s}$ is W are precisely the injective k-tuples in W^k.

[*Proof* uses the fact that any k-element subset of W must be W itself. For details, see notes.]

Using Observation 1, we can rewrite our above computation as

$$\left|S_{\text{inj}}^k\right| = \sum_{\substack{W\subseteq S;\\ |W|=k}} \underbrace{\left(\#\text{ of injective } k\text{-tuples in } W^k\right)}_{=\left|W_{\text{inj}}^k\right| = |W|^{\underline{k}}} \quad \text{(by the previous proposition)}$$
$$= \sum_{\substack{W\subseteq S;\\ |W|=k}} \underbrace{|W|^{\underline{k}}}_{\substack{=k^{\underline{k}}=k(k-1)(k-2)\cdots 1\\ =k!}} = \sum_{\substack{W\subseteq S;\\ |W|=k}} k!$$
$$= \left(\#\text{ of } k\text{-element subsets of } S\right) \cdot k!.$$

Comparing this with $\left|S_{\text{inj}}^k\right| = n^{\underline{k}}$, we obtain

$$n^{\underline{k}} = \left(\#\text{ of } k\text{-element subsets of } S\right) \cdot k!.$$

Solving this for $\left(\#\text{ of } k\text{-element subsets of } S\right)$, we obtain

$$\left(\#\text{ of } k\text{-element subsets of } S\right) = \frac{n^{\underline{k}}}{k!} = \frac{n\left(n-1\right)\left(n-2\right)\cdots\left(n-k+1\right)}{k!} = \binom{n}{k}.$$

□

Class of 2019-09-30

2.8. Another use for polynomials

Let us see how else polynomials can be applied. Our first illustration will be yet another proof of the Chu–Vandermonde identity.

Lemma 2.8.1. Let $p \in \mathbb{N}$. Then,

$$(1+X)^p = \sum_{m \in \mathbb{N}} \binom{p}{m} X^m.$$

Proof. We proved the binomial formula for real numbers, but the same argument works for polynomials. Thus,

$$(1+X)^p = (X+1)^p = \sum_{m=0}^{p} \binom{p}{m} X^m \underbrace{1^{p-m}}_{=1} = \sum_{m=0}^{p} \binom{p}{m} X^m \overset{0}{=} \sum_{m \in \mathbb{N}} \binom{p}{m} X^m.$$

□

Another proof of Chu–Vandermonde convolution for $x, y \in \mathbb{N}$. Rename x and y as a and b. Thus, we have $a, b, n \in \mathbb{N}$ and we must prove

$$\binom{a+b}{n} = \sum_{k=0}^{n} \binom{a}{k} \binom{b}{n-k}.$$

Consider the polynomial

$$(1+X)^{a+b} = \sum_{m \in \mathbb{N}} \binom{a+b}{m} X^m.$$

Compare this with

$$\begin{aligned}
(1+X)^{a+b} &= \underbrace{(1+X)^a}_{\substack{= \sum_{m \in \mathbb{N}} \binom{a}{m} X^m \\ = \sum_{i \in \mathbb{N}} \binom{a}{i} X^i}} \cdot \underbrace{(1+X)^b}_{\substack{= \sum_{m \in \mathbb{N}} \binom{b}{m} X^m \\ = \sum_{j \in \mathbb{N}} \binom{b}{j} X^j}} \qquad \left(\text{since } u^{a+b} = u^a \cdot u^b\right) \\
&= \left(\sum_{i \in \mathbb{N}} \binom{a}{i} X^i \right) \cdot \left(\sum_{j \in \mathbb{N}} \binom{b}{j} X^j \right) = \sum_{i \in \mathbb{N}} \binom{a}{i} X^i \cdot \sum_{j \in \mathbb{N}} \binom{b}{j} X^j \\
&= \underbrace{\sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}}}_{= \sum_{(i,j) \in \mathbb{N} \times \mathbb{N}}} \binom{a}{i} X^i \cdot \binom{b}{j} X^j = \sum_{(i,j) \in \mathbb{N} \times \mathbb{N}} \underbrace{\binom{a}{i} X^i \cdot \binom{b}{j} X^j}_{= \binom{a}{i}\binom{b}{j} X^{i+j}} \\
&= \sum_{(i,j) \in \mathbb{N} \times \mathbb{N}} \binom{a}{i} \binom{b}{j} X^{i+j} = \sum_{m \in \mathbb{N}} \sum_{\substack{(i,j) \in \mathbb{N} \times \mathbb{N}; \\ i+j=m}} \binom{a}{i} \binom{b}{j} \underbrace{X^{i+j}}_{=X^m} \\
&= \sum_{m \in \mathbb{N}} \left(\sum_{\substack{(i,j) \in \mathbb{N} \times \mathbb{N}; \\ i+j=m}} \binom{a}{i} \binom{b}{j} \right) X^m,
\end{aligned}$$

we obtain

$$\sum_{m\in\mathbb{N}} \binom{a+b}{m} X^m = \sum_{m\in\mathbb{N}} \left(\sum_{\substack{(i,j)\in\mathbb{N}\times\mathbb{N}; \\ i+j=m}} \binom{a}{i}\binom{b}{j} \right) X^m.$$

Now, this is an equality between two polynomials in X. The coefficient of X^n on the LHS is $\binom{a+b}{n}$. The coefficient of X^n on the RHS is $\sum_{\substack{(i,j)\in\mathbb{N}\times\mathbb{N}; \\ i+j=n}} \binom{a}{i}\binom{b}{j}$. But since the LHS and the RHS are equal, the coefficients of X^n on them must be equal, too. Thus,

$$\binom{a+b}{n} = \sum_{\substack{(i,j)\in\mathbb{N}\times\mathbb{N}; \\ i+j=n}} \binom{a}{i}\binom{b}{j} = \sum_{k=0}^{n} \binom{a}{k}\binom{b}{n-k}$$

(since each pair $(i,j) \in \mathbb{N}\times\mathbb{N}$ satisfying $i+j=n$ can be written as $(k, n-k)$ for a unique $k \in \{0,1,\ldots,n\}$, and conversely, each $(k,n-k)$ with $k \in \{0,1,\ldots,n\}$ is a pair $(i,j) \in \mathbb{N}\times\mathbb{N}$ satisfying $i+j=n$). Thus, we have proved Chu–Vandermonde for $a, b \in \mathbb{N}$. □

Note: If we plug $p=-1$ in the lemma above, we get

$$\begin{aligned}(1+X)^{-1} &= \sum_{m\in\mathbb{N}} \underbrace{\binom{-1}{m}}_{=(-1)^m} X^m = \sum_{m\in\mathbb{N}} (-1)^m X^m \\ &= 1 - X + X^2 - X^3 + X^4 - X^5 \pm \cdots,\end{aligned}$$

which is not a polynomial anymore. We will later make sense of it as a formal power series, and then the above equality will actually make sense and be true. More generally, the lemma holds for any number p, integer or not, but this requires a reasonable definition of what $(1+X)^p$ means when $p \notin \mathbb{N}$, and also a reasonable definition of infinite sums with infinitely many nonzero addends.

Remark 2.8.2. A similar argument can be used to prove the formula

$$\sum_{k=0}^{m} (-1)^k \binom{n}{k}\binom{n}{m-k} = \begin{cases} (-1)^{m/2}\binom{n}{m/2}, & \text{if } m \text{ is even;} \\ 0, & \text{if } m \text{ is odd} \end{cases}$$

for all $n, m \in \mathbb{N}$. (To prove this, proceed as above, but comparing coefficients in the equality $(1-X)^n \cdot (1+X)^n = \left(1-X^2\right)^n$.) You can extend this to all $n \in \mathbb{R}$ via the polynomial identity trick.

As a particular case, if we set $m=n$, then we get

$$\sum_{k=0}^{m} (-1)^k \binom{n}{k}^2 = \begin{cases} (-1)^{n/2}\binom{n}{n/2}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd} \end{cases}.$$

2.9. The principle of inclusion and exclusion

2.9.1. The principles

Let us begin with some easy facts about finite sets:

- If A_1 and A_2 are two finite sets, then
$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2| .$$
- If A_1, A_2 and A_3 are three finite sets, then
$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3| .$$

These are particular cases of a general fact:

Theorem 2.9.1. (Principle of Inclusion and Exclusion (union form)).
Let $n \in \mathbb{N}$, and let $A_1, A_2, \ldots, A_n$ be n finite sets. Then,

$$\begin{aligned}
|A_1 \cup A_2 \cup \cdots \cup A_n| &= \sum_{m=1}^{n} (-1)^{m-1} \sum_{\substack{(i_1,i_2,\ldots,i_m) \in [n]^m; \\ i_1 < i_2 < \cdots < i_m}} \left|A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m}\right| \\
&= |A_1| + |A_2| + \cdots + |A_n| \\
&\quad - |A_1 \cap A_2| - |A_1 \cap A_3| - \cdots - |A_{n-1} \cap A_n| \\
&\quad + |A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + \cdots + |A_{n-2} \cap A_{n-1} \cap A_n| \\
&\quad \pm \cdots \\
&\quad + (-1)^{n-1} |A_1 \cap A_2 \cap \cdots \cap A_n| .
\end{aligned}$$

We are going to restate this in a somewhat more convenient language before we prove this.

Definition 2.9.2. Let I be a nonempty set. For each $i \in I$, let A_i be a set. Then, we set
$$\bigcap_{i \in I} A_i = \{x \mid x \in A_i \text{ for each } i \in I\} .$$
This set $\bigcap_{i \in I} A_i$ is called the **intersection** of all the A_i.

Thus, if $I = \{i_1, i_2, \ldots, i_k\}$, then $\bigcap_{i \in I} A_i = A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}$.

This notation $\bigcap_{i \in I} A_i$ is similar to $\sum_{i \in I} a_i$, but I has to be nonempty and I can be infinite.

Now, we can restate the above Principle as follows:

Theorem 2.9.3. (Principle of Inclusion and Exclusion (union form)).
Let $n \in \mathbb{N}$, and let $A_1, A_2, \ldots, A_n$ be n finite sets. Then,

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_{\substack{I \subseteq [n]; \\ I \neq \varnothing}} (-1)^{|I|-1} \left| \bigcap_{i \in I} A_i \right|.$$

(See the notes for why this is equivalent to the Theorem above.)
Let us state yet another equivalent version of this Principle:

Theorem 2.9.4. (Principle of Inclusion and Exclusion (complement form)).
Let U be a finite set. Let $n \in \mathbb{N}$, and let $A_1, A_2, \ldots, A_n$ be n subsets of U. Then,

$$|U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n)| = \sum_{I \subseteq [n]} (-1)^{|I|} \left| \bigcap_{i \in I} A_i \right|.$$

Here, the "empty intersection" $\bigcap\limits_{i \in \varnothing} A_i$ is understood to mean U.

For example, for $n = 2$, we obtain

$$\begin{aligned}
&|U \setminus (A_1 \cup A_2)| \\
&= \sum_{I \subseteq [2]} (-1)^{|I|} \left| \bigcap_{i \in I} A_i \right| \\
&= (-1)^{|\varnothing|} \left| \bigcap_{i \in \varnothing} A_i \right| + (-1)^{|\{1\}|} \left| \bigcap_{i \in \{1\}} A_i \right| + (-1)^{|\{2\}|} \left| \bigcap_{i \in \{2\}} A_i \right| + (-1)^{|\{1,2\}|} \left| \bigcap_{i \in \{1,2\}} A_i \right| \\
&= \left| \underbrace{\bigcap_{i \in \varnothing} A_i}_{=U} \right| - \left| \underbrace{\bigcap_{i \in \{1\}} A_i}_{=A_1} \right| - \left| \underbrace{\bigcap_{i \in \{2\}} A_i}_{=A_2} \right| + \left| \underbrace{\bigcap_{i \in \{1,2\}} A_i}_{=A_1 \cap A_2} \right| \\
&= |U| - |A_1| - |A_2| + |A_1 \cap A_2|.
\end{aligned}$$

Similarly for $n = 3$ and higher.

The three theorems we have stated above are called the **Principle(s) of Inclusion and Exclusion** or the **Sylvester sieve formulas**.

We are going to prove them next time and see at least one application. Note that there are several proofs, including a pretty straightforward one by induction on n.

2.9.2. The cancellation lemma

Theorem 2.9.5. (The cancellation lemma)
Let S be a finite set. Then,

$$\sum_{I \subseteq S} (-1)^{|I|} = [S = \varnothing] .$$

Proof. (1st proof, sketched.) Let $|S| = n$. Then, splitting the sum according to $|I|$, we get

$$\begin{aligned}
\sum_{I \subseteq S} (-1)^{|I|} &= \sum_{k=0}^{n} \binom{n}{k} (-1)^k = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \\
&= \left[\underbrace{n}_{=|S|} = 0 \right] \qquad \left(\begin{array}{c} \text{by one of the properties} \\ \text{of Pascal's triangle} \end{array} \right) \\
&= [|S| = 0] = [S = \varnothing]
\end{aligned}$$

(because if two logical statements $\mathcal{A}$ and $\mathcal{B}$ are equivalent, then $[\mathcal{A}] = [\mathcal{B}]$). □

Proof. (2nd proof, idea.) If $S = \varnothing$, then this is obvious. Thus, WLOG assume that $S \neq \varnothing$. Then, there exists some $g \in S$. Choose such g.

Now,

$$\begin{aligned}
\sum_{I \subseteq S} (-1)^{|I|} &= \sum_{\substack{I \subseteq S; \\ g \in I}} (-1)^{|I|} + \sum_{\substack{I \subseteq S; \\ g \notin I}} (-1)^{|I|} \\
&= \sum_{\substack{J \subseteq S; \\ g \notin J}} (-1)^{|J \cup \{g\}|} + \sum_{\substack{I \subseteq S; \\ g \notin I}} (-1)^{|I|} \\
&\qquad \left(\begin{array}{c} \text{here, we have noticed that the map} \\ \{J \subseteq S \mid g \notin J\} \to \{I \subseteq S \mid g \in I\}, \\ J \mapsto J \cup \{g\} \text{ is a bijection,} \\ \text{and thus we have substituted } J \cup \{g\} \text{ for } I \text{ in the first sum} \end{array} \right) \\
&= \sum_{\substack{I \subseteq S; \\ g \notin I}} (-1)^{|I \cup \{g\}|} + \sum_{\substack{I \subseteq S; \\ g \notin I}} (-1)^{|I|} = \sum_{\substack{I \subseteq S; \\ g \notin I}} \underbrace{\left((-1)^{|I \cup \{g\}|} + (-1)^{|I|} \right)}_{=0} = 0 = [S = \varnothing]
\end{aligned}$$

(since $S \neq \varnothing$). □

Example: For $S = \{1, 2\}$, the cancellation lemma says that

$$\underbrace{(-1)^{|\varnothing|}}_{=1} + \underbrace{(-1)^{|\{1\}|}}_{=-1} + \underbrace{(-1)^{|\{2\}|}}_{=-1} + \underbrace{(-1)^{|\{1,2\}|}}_{=1} = \underbrace{[\{1,2\} = \varnothing]}_{=0} .$$

2.9.3. The proofs

We shall use the following proposition ("counting by roll-call"):

Proposition 2.9.6. Let U be a finite set. Let T be a subset of U. Then,

$$|T| = \sum_{s \in U} [s \in T] .$$

Class of 2019-10-30

We are now going to prove the last version of the Principle of Inclusion & Exclusion:

Theorem 2.9.7. (Principle of Inclusion and Exclusion (complement form)).
Let U be a finite set. Let $n \in \mathbb{N}$, and let $A_1, A_2, \ldots, A_n$ be n subsets of U. Then,

$$|U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n)| = \sum_{I \subseteq [n]} (-1)^{|I|} \left| \bigcap_{i \in I} A_i \right| .$$

Here, the "empty intersection" $\bigcap_{i \in \varnothing} A_i$ is understood to mean U.

Proof. The idea is to fix a particular $s \in U$ and see how often it is counted on both sides of the equality. Once we know that it is counted equally often, we conclude that the two sides are equal.

Here are the details: Fix $s \in U$. Define a subset S of $[n]$ by

$$S = \{ i \in [n] \mid s \in A_i \} .$$

Now,

$$\sum_{I\subseteq[n]} (-1)^{|I|} \underbrace{\left[s \in \bigcap_{i\in I} A_i\right]}_{\substack{=[s\in A_i \text{ for each } i\in I] \\ =[i\in S \text{ for each } i\in I] \\ =[I\subseteq S]}}$$

$$= \sum_{I\subseteq[n]} (-1)^{|I|} [I \subseteq S] = \sum_{\substack{I\subseteq[n]; \\ I\subseteq S}} (-1)^{|I|} \underbrace{[I \subseteq S]}_{\substack{=1 \\ (\text{since } I\subseteq S)}} + \sum_{\substack{I\subseteq[n]; \\ I\not\subseteq S}} (-1)^{|I|} \underbrace{[I \subseteq S]}_{\substack{=0 \\ (\text{since } I\not\subseteq S)}}$$

$$= \sum_{\substack{I\subseteq[n]; \\ I\subseteq S}} (-1)^{|I|} + \underbrace{\sum_{\substack{I\subseteq[n]; \\ I\not\subseteq S}} (-1)^{|I|} 0}_{=0} = \sum_{\substack{I\subseteq[n]; \\ I\subseteq S}} (-1)^{|I|} = \sum_{I\subseteq S} (-1)^{|I|}$$

$$= [S = \varnothing] \qquad \text{(by the cancellation lemma)}$$

$$= [\text{there exist no } i \in [n] \text{ such that } s \in A_i]$$

$$= [s \notin A_i \text{ for all } i \in [n]]$$

$$= [s \in U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n)]\,.$$

So we have proved this for each $s \in U$. Now,

$$\sum_{I\subseteq[n]} (-1)^{|I|} \underbrace{\left|\bigcap_{i\in I} A_i\right|}_{\substack{= \sum_{s\in U}\left[s\in \bigcap_{i\in I} A_i\right] \\ \text{(by "counting by roll call")}}}$$

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

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

$$= \sum_{s\in U} \underbrace{\sum_{I\subseteq[n]} (-1)^{|I|} \left[s \in \bigcap_{i\in I} A_i\right]}_{\substack{=[s\in U\setminus(A_1\cup A_2\cup\cdots\cup A_n)] \\ \text{(by our previous computation)}}}$$

$$= \sum_{s\in U} [s \in U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n)] = |U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n)|$$

(by "counting by roll call"). So the complement form of the Principle of Inclusion and Exclusion is proved. □

Proof of the union form:

Theorem 2.9.8. (Principle of Inclusion and Exclusion (union form)).
Let $n \in \mathbb{N}$, and let $A_1, A_2, \ldots, A_n$ be n finite sets. Then,

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_{\substack{I \subseteq [n]; \\ I \neq \varnothing}} (-1)^{|I|-1} \left| \bigcap_{i \in I} A_i \right|.$$

Proof. Set $U = A_1 \cup A_2 \cup \cdots \cup A_n$. Now, $A_1, A_2, \ldots, A_n$ are subsets of U. From $U = A_1 \cup A_2 \cup \cdots \cup A_n$, we obtain $|U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n)| = 0$. Now, the Principle of Inclusion and Exclusion (in complement form) yields

$$|U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n)| = \sum_{I \subseteq [n]} (-1)^{|I|} \left| \bigcap_{i \in I} A_i \right|.$$

Comparing, we obtain

$$\begin{aligned} 0 &= \sum_{I \subseteq [n]} (-1)^{|I|} \left| \bigcap_{i \in I} A_i \right| \\ &= \underbrace{(-1)^{|\varnothing|}}_{=1} \left| \underbrace{\bigcap_{i \in \varnothing} A_i}_{=U} \right| + \sum_{\substack{I \subseteq [n]; \\ I \neq \varnothing}} \underbrace{(-1)^{|I|}}_{=-(-1)^{|I|-1}} \left| \bigcap_{i \in I} A_i \right| \\ &= |U| - \sum_{\substack{I \subseteq [n]; \\ I \neq \varnothing}} (-1)^{|I|-1} \left| \bigcap_{i \in I} A_i \right|. \end{aligned}$$

Thus,

$$\sum_{\substack{I \subseteq [n]; \\ I \neq \varnothing}} (-1)^{|I|-1} \left| \bigcap_{i \in I} A_i \right| = |U| = |A_1 \cup A_2 \cup \cdots \cup A_n|,$$

qed. □

2.9.4. Application: Surjections

Let us see how the Principles above are applied.

Let us compute $\operatorname{sur}(m, n)$ again. Fix $m, n \in \mathbb{N}$. Recall that $\operatorname{sur}(m, n)$ is the # of surjective maps from $[m]$ to $[n]$. In order to compute this using the PIE (= Principle of Inclusion and Exclusion), we need to find a finite set ("universe") U and n subsets $A_1, A_2, \ldots, A_n$ of U such that

$$U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n) = \{\text{surjective maps from } [m] \text{ to } [n]\}.$$

To achieve this, we set

$$U = \{\text{maps from } [m] \text{ to } [n]\};$$
$$A_i = \{\text{maps from } [m] \text{ to } [n] \text{ that don't take } i \text{ as a value}\}.$$

Now, for each subset I of $[n]$, we have

$$\begin{aligned}\bigcap_{i\in I} A_i &= \{\text{maps from } [m] \text{ to } [n] \text{ that don't take any } i \in I \text{ as a value}\} \\ &= \{\text{maps from } [m] \text{ to } [n] \text{ whose image is contained in } [n]\setminus I\}\end{aligned}$$

and thus

$$\begin{aligned}\left|\bigcap_{i\in I} A_i\right| &= |\{\text{maps from } [m] \text{ to } [n] \text{ whose image is contained in } [n]\setminus I\}| \\ &= \{\text{maps from } [m] \text{ to } [n]\setminus I\} = (n-|I|)^m.\end{aligned}$$

Now, the Principle of Inclusion and Exclusion (complement form) yields

$$\begin{aligned}&|U\setminus(A_1\cup A_2\cup\cdots\cup A_n)| \\ &= \sum_{I\subseteq[n]} (-1)^{|I|} \underbrace{\left|\bigcap_{i\in I} A_i\right|}_{=(n-|I|)^m} = \sum_{I\subseteq[n]} (-1)^{|I|}(n-|I|)^m \\ &= \sum_{k=0}^{n} \sum_{\substack{I\subseteq[n];\\ |I|=k}} \underbrace{(-1)^{|I|}(n-|I|)^m}_{=(-1)^k(n-k)^m} = \sum_{k=0}^{n} \underbrace{\sum_{\substack{I\subseteq[n];\\ |I|=k}} (-1)^k (n-k)^m}_{=\binom{n}{k}\cdot(-1)^k(n-k)^m} \\ &= \sum_{k=0}^{n} \binom{n}{k}\cdot(-1)^k(n-k)^m = \sum_{k=0}^{n}(-1)^k\binom{n}{k}(n-k)^m \\ &= \sum_{i=0}^{n}(-1)^{n-i}\binom{n}{n-i} i^m \qquad \text{(here, we substituted } n-i \text{ for } k\text{)} \\ &= \sum_{i=0}^{n}(-1)^{n-i}\binom{n}{i} i^m.\end{aligned}$$

Since $|U\setminus(A_1\cup A_2\cup\cdots\cup A_n)| = \operatorname{sur}(m,n)$ (since $U\setminus(A_1\cup A_2\cup\cdots\cup A_n)$ is the set of all surjective maps from $[m]$ to $[n]$). Thus, we get

$$\operatorname{sur}(m,n) = \sum_{i=0}^{n}(-1)^{n-i}\binom{n}{i} i^m.$$

This is Theorem 2.4.17 in the notes.

2.9.5. Application: Derangements

Let $n \in \mathbb{N}$. Let us count derangements of $[n]$. Recall that their # is called D_n.

Set

$$U = \{\text{permutations of } [n]\};$$
$$A_i = \{\text{permutations of } [n] \text{ that send } i \text{ to } i\},$$

so that

$$U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n) = \{\text{derangements of } [n]\}.$$

Now, for every subset I of $[n]$, we have

$$\bigcap_{i \in I} A_i = \{\text{permutations of } [n] \text{ that send } i \text{ to } i \text{ for each } i \in I\}$$

and thus

$$\begin{aligned}
\left| \bigcap_{i \in I} A_i \right| &= |\{\text{permutations of } [n] \text{ that send } i \text{ to } i \text{ for each } i \in I\}| \\
&= |\{\text{permutations of } [n] \setminus I\}| \\
&\qquad \left(\begin{array}{c} \text{since a permutation of } [n] \text{ that sends } i \text{ to } i \text{ for each } i \in I \\ \text{is ``the same as'' a permutation of } [n] \setminus I \end{array} \right) \\
&= |[n] \setminus I|! = (n - |I|)!.
\end{aligned}$$

So the Principle of Inclusion and Exclusion (in the complement form) yields

$$\begin{aligned}
&|U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n)| \\
&= \sum_{I \subseteq [n]} (-1)^{|I|} \underbrace{\left| \bigcap_{i \in I} A_i \right|}_{=(n-|I|)!} = \sum_{I \subseteq [n]} (-1)^{|I|} (n - |I|)! \\
&= \text{.....} \\
&= \sum_{k=0}^{n} (-1)^k \underbrace{\binom{n}{k}}_{= \frac{n!}{k! \cdot (n-k)!}} (n-k)! = \sum_{k=0}^{n} (-1)^k \frac{n!}{k!}.
\end{aligned}$$

Since $|U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n)| = D_n$ (because $U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n)$ is the set of all derangements of $[n]$), this becomes

$$D_n = \sum_{k=0}^{n} (-1)^k \frac{n!}{k!}.$$

This is Theorem 1.7.9 **(d)**.

2.9.6. Application: Euler's totient function

Reminder: Two integers a and b are **coprime** (or **relatively prime**) if $\gcd(a,b) = 1$.

Definition 2.9.9. The function $\phi : \{1,2,3,\ldots\} \to \mathbb{N}$ (called **Euler's totient function** or **Euler's ϕ-function**) is defined by

$$\phi(u) = (\#\text{ of all } m \in [u] \text{ that are coprime to } u).$$

Theorem 2.9.10. If u is a positive integer, and if $p_1, p_2, \ldots, p_n$ are the distinct primes that divide u, then

$$\phi(u) = u \prod_{i=1}^{n} \left(1 - \frac{1}{p_i}\right).$$

This can be proved using the PIE. Here,

$$U = [u];$$
$$A_i = \{m \in [u] \mid p_i \text{ divides } m\}.$$

Thus, $U \setminus (A_1 \cup A_2 \cup \cdots \cup A_n) = \{m \in [u] \mid m \text{ is coprime to } u\}$.

More details will appear in the notes.

Class of 2019-11-04

2.10. Compositions, weak compositions and multisets

2.10.1. Compositions

How many ways are there to write 5 as a sum of 3 positive integers, if the order matters?

$$5 = 1+1+3 = 1+3+1 = 3+1+1$$
$$= 1+2+2 = 2+1+2 = 2+2+1.$$

Theorem 2.10.1. Let $\mathbb{P} = \{1,2,3,\ldots\}$. Let $n, k \in \mathbb{N}$. Then,

$$\left(\#\text{ of } (x_1, x_2, \ldots, x_k) \in \mathbb{P}^k \text{ satisfying } x_1 + x_2 + \cdots + x_k = n\right)$$
$$= \binom{n-1}{n-k}$$
$$= \begin{cases} \dbinom{n-1}{k-1}, & \text{if } n > 0; \\ [k=0], & \text{if } n = 0 \end{cases}.$$

Remark 2.10.2. Some terminology:
A **composition** is a tuple (= finite list) of positive integers.
A **composition into k parts** is a k-tuple of positive integers.
A **composition of n** is a tuple of positive integers whose sum is n.
Thus, the Theorem above says

$$\begin{aligned}&(\#\text{ of compositions of } n \text{ into } k \text{ parts})\\ &= \binom{n-1}{n-k}\\ &= \begin{cases}\binom{n-1}{k-1}, & \text{if } n>0;\\ [k=0], & \text{if } n=0\end{cases}.\end{aligned}$$

Proof. (Proof of the Theorem.)
WLOG assume $n \geq 1$ (else, the claim is straightforward: there is only composition of 0, namely $()$).
For each composition $(x_1, x_2, \ldots, x_k)$ of n into k parts, we set

$$\begin{aligned}D(x_1, x_2, \ldots, x_k) &= \left\{x_1 + x_2 + \cdots + x_j \mid j \in [k-1]\right\}\\ &= \{x_1,\ x_1+x_2,\ x_1+x_2+x_3,\ \ldots,\ x_1+x_2+\cdots+x_{k-1}\}\\ &= \{x_1 < x_1+x_2 < x_1+x_2+x_3 < \cdots < x_1+x_2+\cdots+x_{k-1}\}\\ &\qquad (\text{the inequalities hold since } x_1, x_2, \ldots, x_k \text{ are positive});\end{aligned}$$

this is a subset of $[n-1]$ and has $k-1$ elements.
(For example, the composition $(3,5,2)$ of 10 has $D(3,5,2) = \{3,8\}$.)
The map

$$\begin{aligned}\{\text{compositions of } n \text{ into } k \text{ parts}\} &\to \{(k-1)\text{-element subsets of } [n-1]\},\\ (x_1, x_2, \ldots, x_k) &\mapsto D(x_1, x_2, \ldots, x_k)\end{aligned}$$

is a bijection (the inverse map sends any $(k-1)$-element subset $\{s_1 < s_2 < \cdots < s_{k-1}\}$ of $[n-1]$ to

$$(s_1 - s_0, s_2 - s_1, s_3 - s_2, \ldots, s_k - s_{k-1}),$$

where we set $s_0 = 0$ and $s_k = n$). See HW0 Exercise 1 **(b)** for details. (The difference is that now, k is fixed.)

Thus, the bijection principle yields

$$\begin{aligned}
&(\#\text{ of compositions of } n \text{ into } k \text{ parts}) \\
&= (\#\text{ of } (k-1)\text{-element subsets of } [n-1]) \\
&= \binom{n-1}{k-1} \qquad \text{(by the combinatorial interpretation of BCs)} \\
&= \binom{n-1}{(n-1)-(k-1)} \qquad \text{(by the symmetry of Pascal's triangle, since } n-1 \geq 0) \\
&= \binom{n-1}{n-k}.
\end{aligned}$$

This proves the theorem. □

2.10.2. Binary compositions

Theorem 2.10.3. Let $n, k \in \mathbb{N}$. Then,

$$\begin{aligned}
&\left(\#\text{ of } (x_1, x_2, \ldots, x_k) \in \{0,1\}^k \text{ satisfying } x_1 + x_2 + \cdots + x_k = n\right) \\
&= \binom{k}{n}.
\end{aligned}$$

Proof. To construct a k-tuple $(x_1, x_2, \ldots, x_k) \in \{0,1\}^k$ satisfying $x_1 + x_2 + \cdots + x_k = n$, we just need to choose which n indices $i \in [k]$ will have $x_i = 1$. There are $\binom{k}{n}$ many options for this, since this boils down to choosing an n-element subset of $[k]$. □

2.10.3. Weak compositions

Theorem 2.10.4. Let $n, k \in \mathbb{N}$. Then,

$$\begin{aligned}
&\left(\#\text{ of } (x_1, x_2, \ldots, x_k) \in \mathbb{N}^k \text{ satisfying } x_1 + x_2 + \cdots + x_k = n\right) \\
&= \binom{n+k-1}{n} \\
&= \begin{cases} \dbinom{n+k-1}{k-1}, & \text{if } k > 0; \\ [n=0], & \text{if } k = 0 \end{cases}.
\end{aligned}$$

Remark 2.10.5. Tuples of nonnegative integers are called **weak compositions**.

Example 2.10.6. For $n = 2$ and $k = 3$, the k-tuples $(x_1, x_2, \ldots, x_k) \in \mathbb{N}^k$ satisfying $x_1 + x_2 + \cdots + x_k = n$ are

$$\begin{aligned} &2 = 0+0+2 = 0+2+0 = 2+0+0 \\ &= 1+1+0 = 1+0+1 = 0+1+1. \end{aligned}$$

Thus, their # is 6. This is exactly what the theorem yields, since $\dbinom{2+3-1}{2} = \dbinom{4}{2} = 6$.

Proof. (Proof of the theorem.) The map

$$\begin{aligned} &\left\{(x_1, x_2, \ldots, x_k) \in \mathbb{N}^k \mid x_1 + x_2 + \cdots + x_k = n\right\} \\ &\to \left\{(x_1, x_2, \ldots, x_k) \in \mathbb{P}^k \mid x_1 + x_2 + \cdots + x_k = n + k\right\}, \\ &(x_1, x_2, \ldots, x_k) \mapsto (x_1 + 1, x_2 + 1, \ldots, x_k + 1) \end{aligned}$$

is a bijection. Thus, the bijection principle yields

$$\begin{aligned} &\left(\#\text{ of } (x_1, x_2, \ldots, x_k) \in \mathbb{N}^k \text{ satisfying } x_1 + x_2 + \cdots + x_k = n\right) \\ &= \left(\#\text{ of } (x_1, x_2, \ldots, x_k) \in \mathbb{P}^k \text{ satisfying } x_1 + x_2 + \cdots + x_k = n + k\right) \\ &= \binom{n+k-1}{(n+k)-k} \qquad \left(\begin{array}{c}\text{by the second-to-last theorem, applied} \\ \text{to } n+k \text{ instead of } n\end{array}\right) \\ &= \binom{n+k-1}{n} = \begin{cases} \dbinom{n+k-1}{k-1}, & \text{if } k > 0; \\ [n = 0], & \text{if } k = 0 \end{cases} \end{aligned}$$

(by the symmetry of Pascal's triangle and an easy distinction of cases). □

2.10.4. Multisubsets

Definition 2.10.7. Let S be a set.

A **finite multisubset of** S is a map from S to $\mathbb{N}$.

We regard such a map $f : S \to \mathbb{N}$ as a "set with multiplicities", in which each $s \in S$ appears $f(s)$ many times.

For example, "the multisubset $\{1, 4, 4, 5, 7, 7, 7\}_{\text{multi}}$ of $[8]$" is encoded as the map from $[8]$ to $\mathbb{N}$ given in two-line notation as

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 0 & 0 & 2 & 1 & 0 & 3 & 0 \end{pmatrix}.$$

The **size** of a multisubset $f : S \to \mathbb{N}$ is defined to be the nonnegative integer $\sum_{s \in S} f(s)$.

Corollary 2.10.8. Let $n, k \in \mathbb{N}$. Let S be a k-element set. Then,

$$(\#\text{ of multisubsets of } S \text{ having size } n) = \binom{n+k-1}{n}.$$

Example 2.10.9. The multisubsets of $[3]$ having size 2 are

$$\{1,1\}_{\text{multi}}, \quad \{1,2\}_{\text{multi}}, \quad \{1,3\}_{\text{multi}}, \quad \{2,2\}_{\text{multi}}, \quad \{2,3\}_{\text{multi}}, \quad \{3,3\}_{\text{multi}}.$$

So there are 6 of them, and this is what the corollary predicts.

Proof. Let $s_1, s_2, \ldots, s_k$ be the k elements of S. Then, the map

$$\{\text{multisubsets of } S \text{ having size } n\} \to \left\{(x_1, x_2, \ldots, x_k) \in \mathbb{N}^k \mid x_1 + x_2 + \cdots + x_k = n\right\},$$
$$f \mapsto (f(s_1), f(s_2), \ldots, f(s_k))$$

is a bijection. Now, use the bijection principle and the previous theorem. □

Exercise 2.10.1. Let $m \in \mathbb{N}$ and $a, b \in \{0, 1, \ldots, m\}$. Prove that

$$(\#\text{ of lacunar subsets of } [2m] \text{ with exactly } a \text{ even and } b \text{ odd elements}) = \binom{m-a}{b} \cdot \binom{m-b}{a}.$$

Proof. (Solution sketch.) The dependent product principle does not apply here: If we (say) first choose the b odd elements, then the # of ways to choose the remaining a even elements depends on our chosen b odd elements. In [18s-hw2s], I show two solutions. Here is a simpler one:

Let S be a lacunar subset of $[2m]$ with exactly a even and b odd elements. Write S as $S = \{s_1 < s_2 < \cdots < s_{a+b}\}$. Then, since S is lacunar, we have

$$s_1 - 0 < s_2 - 1 < s_3 - 2 < \cdots < s_{a+b} - (a+b-1),$$

thus

$$s_1 - 0 \leq s_2 - 2 \leq s_4 - 4 \leq \cdots \leq s_{a+b} - 2(a+b-1).$$

Consider the multisubset

$$M_S := \{s_1 - 0 \leq s_2 - 2 \leq s_4 - 4 \leq \cdots \leq s_{a+b} - 2(a+b-1)\}_{\text{multi}}$$

of $[2m-2(a+b-1)]$. It has exactly a even and b odd elements. Thus, we can split M_S into a "multiset union" $M_{S,\text{even}} \cup M_{S,\text{odd}}$ (like a union of sets, but multiplicities get added), where

$$M_{S,\text{even}} \text{ is a size-}a \text{ multisubset of } \{2,4,6,\ldots,2m-2(a+b-1)\},$$

and where

$$M_{S,\text{odd}} \text{ is a size-}b \text{ multisubset of } \{1,3,5,\ldots,2m-2(a+b-1)-1\}.$$

Moreover, this encoding

$$\begin{aligned}
&\{\text{lacunar subsets of } [2m] \text{ with exactly } a \text{ even and } b \text{ odd elements}\} \\
&\to \{\text{size-}a \text{ multisubsets of } \{2,4,6,\ldots,2m-2(a+b-1)\}\} \\
&\qquad \times \{\text{size-}b \text{ multisubsets of } \{1,3,5,\ldots,2m-2(a+b-1)-1\}\}, \\
&S \mapsto (M_{S,\text{even}}, M_{S,\text{odd}})
\end{aligned}$$

is a bijection. Hence, the bijection principle and the product rule yield

$$\begin{aligned}
&(\#\text{ of lacunar subsets of } [2m] \text{ with exactly } a \text{ even and } b \text{ odd elements}) \\
&= \underbrace{(\#\text{ of size-}a \text{ multisubsets of } \{2,4,6,\ldots,2m-2(a+b-1)\})}_{=\binom{a+(m-a-b+1)-1}{a}=\binom{m-b}{a}} \\
&\qquad \cdot \underbrace{(\#\text{ of size-}b \text{ multisubsets of } \{1,3,5,\ldots,2m-2(a+b-1)-1\})}_{=\binom{b+(m-a-b+1)-1}{b}=\binom{m-a}{b}} \\
&= \binom{m-b}{a} \cdot \binom{m-a}{b}.
\end{aligned}$$

□

Note that, as a consequence, you get the following formula for Fibonacci numbers:

$$f_{2m+1} = \sum_{a=0}^{m} \sum_{b=0}^{m} \binom{m-a}{b} \binom{m-b}{a}.$$

2.11. Multinomial coefficients

Definition 2.11.1. Let $n, n_1, n_2, \ldots, n_k \in \mathbb{N}$ be such that $n_1 + n_2 + \cdots + n_k = n$. Then,

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}.$$

Remark 2.11.2. **(a)** We are not defining this for negative or non-integer n.

(b) If $m \in \{0, 1, \ldots, n\}$, then $\dbinom{n}{m} = \dbinom{n}{m, n-m}$.

Proposition 2.11.3. Let $n, n_1, n_2, \ldots, n_k \in \mathbb{N}$ be such that $n_1 + n_2 + \cdots + n_k = n$. Then:

(a) We have

$$\begin{aligned}
\binom{n}{n_1, n_2, \ldots, n_k} &= \prod_{i=1}^{k} \binom{n - n_1 - n_2 - \cdots - n_{i-1}}{n_i} \\
&= \binom{n}{n_1} \binom{n - n_1}{n_2} \binom{n - n_1 - n_2}{n_3} \cdots \underbrace{\binom{n - n_1 - n_2 - \cdots - n_{k-1}}{n_k}}_{=1} \\
&= \prod_{i=1}^{k-1} \binom{n - n_1 - n_2 - \cdots - n_{i-1}}{n_i}.
\end{aligned}$$

(b) $\dbinom{n}{n_1, n_2, \ldots, n_k} \in \mathbb{N}$.

(c) The # of maps $f : [n] \to [k]$ satisfying

$$(\#\text{ of } a \in [n] \text{ such that } f(a) = i) = n_i \qquad \text{for each } i \in [k]$$

is $\dbinom{n}{n_1, n_2, \ldots, n_k}$.

(d) Let α be the n-tuple

$$\left(\underbrace{1, 1, \ldots, 1}_{n_1 \text{ times}}, \underbrace{2, 2, \ldots, 2}_{n_2 \text{ times}}, \ldots, \underbrace{k, k, \ldots, k}_{n_k \text{ times}} \right).$$

Then, the # of distinct anagrams of α (that is, n-tuples obtained from α by permuting its entries) is $\dbinom{n}{n_1, n_2, \ldots, n_k}$.

"Anagrams" are often called "permutations".

Example 2.11.4. How many anagrams does the word "anagram" have?
Equivalently, how many anagrams does the word "aaagmnr" have?
Equivalently, how many anagrams does the 7-tuple "$(1,1,1,2,3,4,5)$" have?
Part **(d)** of the above proposition says that the answer is $\binom{7}{3,1,1,1,1} = \dfrac{7!}{3! \cdot 1! \cdot 1! \cdot 1! \cdot 1!} = \dfrac{7!}{3!} = 7 \cdot 6 \cdot 5 \cdot 4 = 840$.

Class of 2019-11-06

Proof. (Proof of the above Proposition.) **(a)** Easy (telescoping product).
(b) follows from **(a)**.
(c) Here is a way to construct any map $f : [n] \to [k]$ satisfying

$$(\#\text{ of } a \in [n] \text{ such that } f(a) = i) = n_i \qquad \text{for each } i \in [k]:$$

- Choose $\{a \in [n] \mid f(a) = 1\}$. (Here, we have $\binom{n}{n_1}$ choices.)
- Choose $\{a \in [n] \mid f(a) = 2\}$. (Here, we have $\binom{n-n_1}{n_2}$ choices.)
- Choose $\{a \in [n] \mid f(a) = 3\}$. (Here, we have $\binom{n-n_1-n_2}{n_3}$ choices.)
- etc.

Altogether, the total # of f's is therefore

$$\prod_{i=1}^{k} \binom{n - n_1 - n_2 - \cdots - n_{i-1}}{n_i} = \binom{n}{n_1, n_2, \ldots, n_k} \qquad (\text{by part } \textbf{(a)}).$$

(d) The map

$$\{\text{anagram of } \alpha\} \to \{\text{map } f : [n] \to [k] \text{ as in part } \textbf{(c)}\},$$
$$(\beta_1, \beta_2, \ldots, \beta_n) \mapsto (\text{the map } f : [n] \to [k] \text{ sending } j \text{ to } \beta_j)$$

is a bijection. [Rigorously proving this would take us a while.] Part **(c)** concludes the argument. □

Note that multinomial coefficients generalize the binomial coefficients that you see in Pascal's triangle:

Proposition 2.11.5. Let $n \in \mathbb{N}$ and $k \in \{0, 1, \ldots, n\}$, then

$$\underbrace{\binom{n}{k}}_{\text{binomial coeff.}} = \underbrace{\binom{n}{k, n-k}}_{\text{multinomial coeff.}}.$$

Proof. This is just saying $\binom{n}{k} = \dfrac{n!}{k!\,(n-k)!}$, but we proved this long ago. □

Proposition 2.11.6. Let $n \in \mathbb{N}$ and $n_1, n_2, \ldots, n_k \in \mathbb{N}$ such that $n_1 + n_2 + \cdots + n_k = n$. Then,

$$\binom{n}{n_1, n_2, \ldots, n_k} = \binom{n}{n_{\sigma(1)}, n_{\sigma(2)}, \ldots, n_{\sigma(k)}}$$

for any permutation σ of $[k]$.

Proof. The definition of multinomial coefficients yields

$$\begin{aligned}\binom{n}{n_{\sigma(1)}, n_{\sigma(2)}, \ldots, n_{\sigma(k)}} &= \frac{n!}{n_{\sigma(1)}! n_{\sigma(2)}! \cdots n_{\sigma(k)}!} = \frac{n!}{n_1! n_2! \cdots n_k!} \\ &\qquad \left(\text{since } n_{\sigma(1)}! n_{\sigma(2)}! \cdots n_{\sigma(k)}! = n_1! n_2! \cdots n_k!\right) \\ &= \binom{n}{n_1, n_2, \ldots, n_k}.\end{aligned}$$

□

Theorem 2.11.7. (Recurrence relation for multinomial coefficients.)
Let $n, n_1, n_2, \ldots, n_k \in \mathbb{N}$ be such that $n_1 + n_2 + \cdots + n_k = n > 0$. Then,

$$\binom{n}{n_1, n_2, \ldots, n_k} = \sum_{i=1}^{k} \underbrace{\binom{n-1}{n_1, \ldots, n_{i-1}, n_i - 1, n_{i+1}, \ldots, n_k}}_{\text{This should be interpreted as 0 if } n_i = 0}.$$

Proof. LTTR. □

($\Longrightarrow$ Pascal's pyramid.)

Theorem 2.11.8. (Multinomial formula.)
Let $x_1, x_2, \ldots, x_k$ be k numbers. Let $n \in \mathbb{N}$. Then,

$$(x_1 + x_2 + \cdots + x_k)^n = \sum_{\substack{(n_1, n_2, \ldots, n_k) \in \mathbb{N}; \\ n_1 + n_2 + \cdots + n_k = n}} \binom{n}{n_1, n_2, \ldots, n_k} x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}.$$

Proof. LTTR. Either use part **(c)** of yesterday's last proposition (see [Galvin, Thm. 11.7]); or use induction on n; or use induction on k. □

3. The twelvefold way

3.1. What is it?

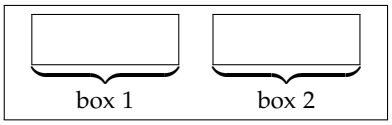
The *twelvefold way* is a table of $4 \cdot 3 = 12$ standard counting problems that tend to appear often. Let us first introduce it informally, and then define it formally and study it in detail.

Informal description: Given a set A of balls and a set X of boxes. A **placement** means a way to distribute the balls into the boxes. Rigorously, we can think of a placement as a map from A to X; more precisely, this is what we will call an "$L \to L$ **placement**".

How many such placements are there? Of course, $|X|^{|A|}$ (by Theorem **??**).

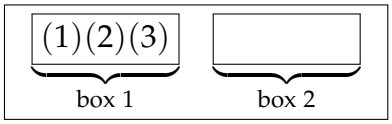
Example 3.1.1. Let $X = [2]$ and $A = [3]$, so that $|X| = 2$ and $|A| = 3$. Thus, we are trying to place 3 balls (called $1, 2, 3$) into 2 boxes (called $1, 2$).

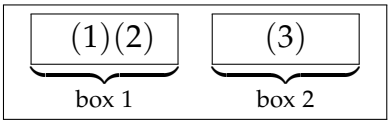
We will draw such placements as follows: We will always draw the boxes in increasing order:



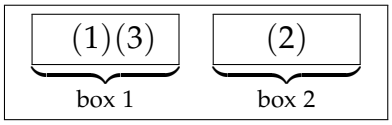
We draw each ball as a pair of parentheses with a number inside – so, for example, the ball 3 will be drawn as "(3)".

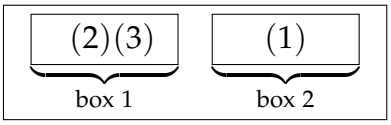
So the $L \to L$ placements of the balls $1, 2, 3 \in A$ in the boxes $1, 2 \in A$ are



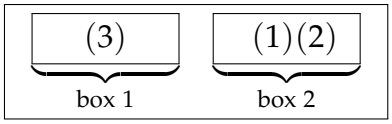
,


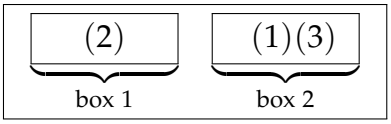
,



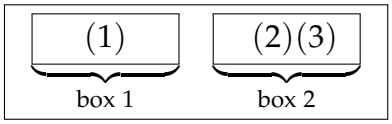
,


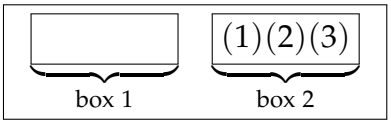
,



,


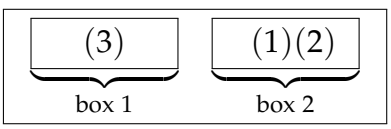
,

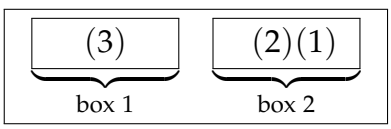


,


.

The order of balls **within** a single box does not matter:



=


.

This suggests the following variations:

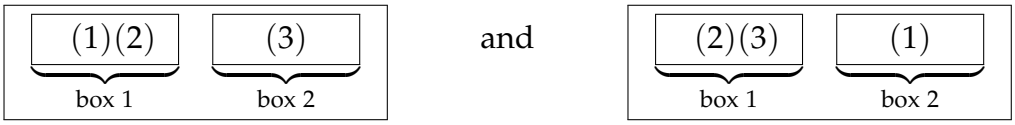
- What if we require $f : A \to X$ to be injective (i.e., each box contains ≤ 1 ball) or surjective (i.e., each box contains ≥ 1 ball)?

- What if the balls are unlabelled (i.e., indistinguishable)?

  To make this rigorous, we will use equivalence classes:

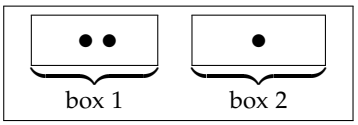
  We will say that two maps $f : A \to X$ are "ball-equivalent" if there is a permutation of the balls that transforms one into the other. Then, "$U \to L$ **placements**" (= placements of unlabelled balls in labelled boxes) will be equivalence classes of ball-equivalence.

  E.g.: the two $L \to L$ placements

  are ball-equivalent, and therefore contribute to only 1 ball-equivalence class – i.e., they count as one and the same $U \to L$ placement.

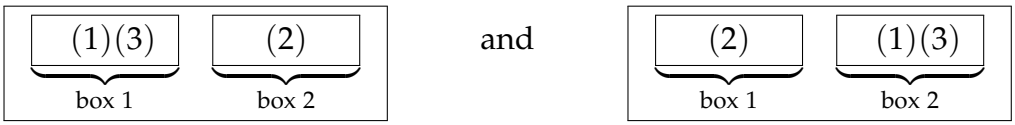
  We shall draw $U \to L$ placements in the same way as we draw $L \to L$ placements, but representing balls by "•" symbols (rather than by numbers in parentheses). Thus, the $U \to L$ placement that we have just discussed becomes

  .

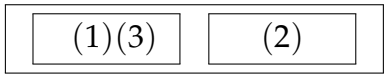
- What if the boxes are unlabelled (i.e., indistinguishable)?

  This can be made rigorous in a similar way, but we now have to permute boxes. This gives "$L \to U$ **placements**". For example, the two $L \to L$ placements

  are box-equivalent (i.e., can be transformed into one another by permuting the boxes) and thus contribute to only 1 box-equivalence class – i.e., they count as one and the same $L \to U$ placement.

  We shall draw $L \to U$ placements in the same way as we draw $L \to L$ placements, but without marking the boxes as "box 1", "box 2" etc.. Thus, the $U \to L$ placement that we have just discussed becomes

  .

- What if both the boxes and the balls are indistinguishable? This kind of placements will be called "$U \to U$ **placements**". We shall draw them in the same way as we draw $L \to L$ placements, but representing balls by "$\bullet$" symbols and without marking the boxes.

In total, we get $3 \cdot 4 = 12$ many different counting problems. We list them in a table:

	arbitrary	injective	surjective
$L \to L$	$\lvert X\rvert^{\lvert A\rvert}$		
$U \to L$			
$L \to U$			
$U \to U$			

Here:

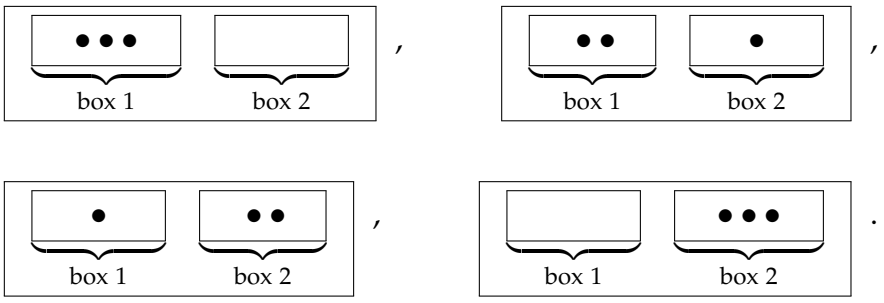
- $L \to L$ means "balls are labelled, boxes are labelled", so we are just counting maps $f : A \to X$.
- $U \to L$ means "balls are unlabelled, boxes are labelled".
- $L \to U$ means "balls are labelled, boxes are unlabelled".
- $U \to U$ means "balls are unlabelled, boxes are unlabelled".

Example 3.1.2. Let $X = [2]$ and $A = [3]$. Then, let us count how many placements of each kind we have:

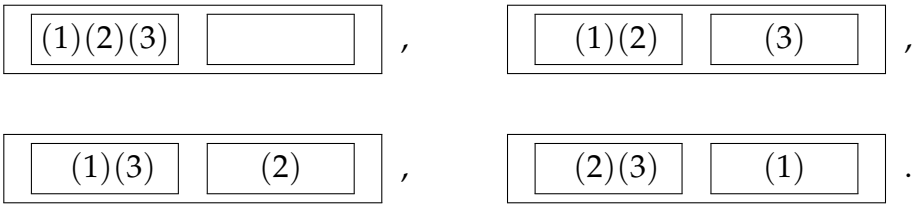
	arbitrary	injective	surjective
$L \to L$	8	0	6
$U \to L$	4	0	2
$L \to U$	4	0	3
$U \to U$	2	0	1

In fact:

- The $U \to L$ placements are



- The $L \to U$ placements are



- The $U \to U$ placements are

In general, not every among the 12 questions has a closed-form answer. But there are, at least, good recursions.

3.2. $L \to L$

Definition 3.2.1. An $L \to L$ **placement** from A to X is just a map from A to X. The value of this map at some $a \in A$ is called the **box in which ball a is placed**.

Proposition 3.2.2.

$$\begin{aligned}&(\#\text{ of } L \to L \text{ placements } A \to X)\\ &= (\#\text{ of maps from } A \text{ to } X) = |X|^{|A|}.\end{aligned}$$

Proof. This is Theorem **??**. □

Proposition 3.2.3.

$$\begin{aligned}&(\#\text{ of injective } L \to L \text{ placements } A \to X)\\ &= (\#\text{ of injective maps from } A \text{ to } X) = |X|^{\underline{|A|}}.\end{aligned}$$

Proof. This is Theorem **??**. □

Proposition 3.2.4.

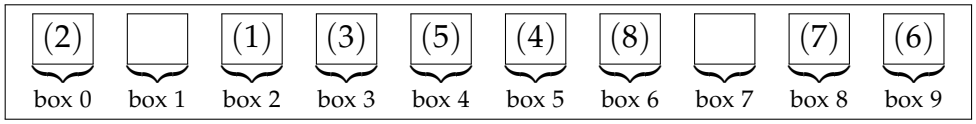
$$\begin{aligned}&(\#\text{ of surjective } L \to L \text{ placements } A \to X)\\ &= (\#\text{ of surjective maps from } A \text{ to } X) = \operatorname{sur}(|A|, |X|),\end{aligned}$$

Proof. This is Proposition **??**. □

Here are some examples of counting problems that boil down to counting $L \to L$ placements:

- assigning grades (from a finite set X) to students (from a finite set A): these are arbitrary $L \to L$ placements.
- assigning IP addresses to computers: these are injective $L \to L$ placements.
- How many 8-digit telephone numbers are there with no 2 equal digits?

  These correspond to injective $L \to L$ placements with $A = [8]$ and $X = \{0, 1, \ldots, 9\}$; for example, the telephone number 20354986 corresponds to the $L \to L$ placement



  .

  $\Longrightarrow$ The total # of such numbers is $10^{\underline{8}} = 10 \cdot 9 \cdot \cdots \cdot 3 = 10!/2!$.

3.3. Equivalence relations

What does it mean for balls, or boxes, to be unlabelled?

Rigorously, it means that we are counting **not** the maps $f : A \to X$, but rather their **equivalence classes** wrt some relation. (The abbreviation "wrt" means "with respect to".)

Before we make sense of this, let us introduce (or recall) the notions of *relations* and, in particular, *equivalence relations*.

Class of 2019-11-08

3.3.1. Relations

I will follow [**?**, Chapter 3].

Definition 3.3.1. Let S be a set. A **(binary) relation** on S is formally defined as a subset of $S \times S$. Informally, it is a statement $x \ R \ y$ defined for every pair of elements $(x, y) \in S \times S$. For each pair $(x, y) \in S \times S$, this statement $x \ R \ y$ is either true or false.

The informal and the formal definitions are equivalent, since:

- Given a statement $x \ R \ y$ defined for every pair $(x, y) \in S \times S$, we can encode it as a subset of $S \times S$, namely as the subset

$$\{(x, y) \in S \times S \mid x \ R \ y\}.$$

- Conversely, every subset T can be decoded into a statement (namely, "$(x,y) \in T$").

In the following, we will say "**relation**" for "**binary relation**".

Example 3.3.2. Let $S = \mathbb{Z}$.

(a) The relation $=$ is a binary relation on S. As a subset of $S \times S$, this relation is

$$\{(a,b) \in S \times S \mid a = b\}$$
$$= \{(c,c) \mid c \in S\} = \{\ldots, (-2,-2), (-1,-1), (0,0), (1,1), (2,2), \ldots\}.$$

(b) The relation $<$ is a binary relation on S. As a subset of $S \times S$, this relation is

$$\{(a,b) \in S \times S \mid a < b\}.$$

(c) The relation $\leq$ is a binary relation on S.

(d) The relation $\neq$ is a binary relation on S.

(e) Fix $n \in \mathbb{Z}$. Define a binary relation $\underset{n}{\equiv}$ on $S = \mathbb{Z}$ by

$$\left(a \underset{n}{\equiv} b\right) \iff (a \equiv b \bmod n) \iff (n \mid a-b)$$
$$\iff (a = b + kn \text{ for some } k \in \mathbb{Z}).$$

This is called **congruence modulo** n. Note that the relation $\underset{0}{\equiv}$ is $=$.

When n is positive, $a \equiv b \bmod n$ if and only if $a\%n = b\%n$.

For $n = 1$, we have

$$\left(a \underset{1}{\equiv} b\right) \iff (1 \mid a-b) \iff \left(\frac{a-b}{1} \in \mathbb{Z}\right) \iff (a-b \in \mathbb{Z}) \iff (\text{true}).$$

Thus, the relation $\underset{1}{\equiv}$ always holds (i.e., we have $a \underset{1}{\equiv} b$ for all $(a,b) \in S \times S$). Thus, it equals the whole set $S \times S$ (when considered as a subset of $S \times S$).

(f) Define a binary relation N on S by

$$(a \ N \ b) \iff (\text{false}).$$

That is, there is no pair $(a,b) \in S \times S$ such that $a \ N \ b$. As a subset of $S \times S$, this relation is $\varnothing$.

(g) Define a binary relation A on S by

$$(a \ A \ b) \iff (\text{true}).$$

That is, every pair $(a,b) \in S \times S$ satisfies $a \ N \ b$. As a subset of $S \times S$, this relation is $S \times S$. This is the above relation $\underset{1}{\equiv}$, but the way we just defined it, it makes sense for every set S (not just for $\mathbb{Z}$).

(h) The relation $\mid$ is a binary relation on S.

(i) The relation "is coprime to" is a binary relation on S.

There are other examples on other sets. For example, "parallel" is a relation between lines in plane geometry.

3.3.2. Equivalence relations

Definition 3.3.3. Let R be a binary relation on a set S.
(a) We say that R is **reflexive** if every $a \in S$ satisfies $a \ R \ a$.
(b) We say that R is **symmetric** if every $a, b \in S$ satisfying $a \ R \ b$ satisfy $b \ R \ a$.
(c) We say that R is **transitive** if every $a, b, c \in S$ satisfying $a \ R \ b$ and $b \ R \ c$ satisfy $a \ R \ c$.

Example 3.3.4. Let S be the set $\mathbb{Z}$.
(a) The relation $=$ is reflexive, symmetric and transitive.
(b) The relation $<$ is not reflexive, not symmetric but transitive.
(c) The relation $\leq$ is reflexive, not symmetric but transitive.
(d) The relation $\neq$ is not reflexive, but symmetric. It is not transitive (e.g., $2 \neq 3$ and $3 \neq 2$, but $2 = 2$).
(e) For every $n \in \mathbb{Z}$, the congruence relation $\equiv_n$ is reflexive (since $n \mid a - a$), symmetric (since $n \mid a - b$ implies $n \mid b - a$) and transitive (since $n \mid a - b$ and $n \mid b - c$ implies $n \mid a - c$).
(f) The relation N is symmetric (vacuously) and transitive (vacuously), but not reflexive.
(g) The relation A is reflexive, symmetric and transitive.
(h) The divisibility relation $\mid$ is reflexive but not symmetric. It is transitive (if $a \mid b$ and $b \mid c$, then $a \mid c$).
(i) The coprimality relation is symmetric but neither reflexive nor transitive.

Definition 3.3.5. An **equivalence relation** on a set S means a relation on S that is reflexive, symmetric and transitive.

Example 3.3.6. Let S be any set. The relation $=$ on S is an equivalence relation.

Example 3.3.7. Let $n \in \mathbb{Z}$. Then, the congruence relation $\equiv_n$ on $\mathbb{Z}$ is an equivalence relation.

Example 3.3.8. The relation "parallel" on $\{$lines in the plane$\}$ is an equivalence relation.

Example 3.3.9. The relation "similar" on $\{$triangles in the plane$\}$ is an equivalence relation.

Example 3.3.10. The relation "directly similar" (= similar with the same orientation) on $\{$triangles in the plane$\}$ is an equivalence relation.

Example 3.3.11. The relation "indirectly similar" (= similar with opposite orientation) on {triangles in the plane} is not an equivalence relation, because it is not reflexive.

Example 3.3.12. Let S and T be two sets. Let $f : S \to T$ be a map. Define a relation $\underset{f}{\equiv}$ on S by

$$\left(a \underset{f}{\equiv} b\right) \Longleftrightarrow \left(f\left(a\right) = f\left(b\right)\right).$$

This relation $\underset{f}{\equiv}$ is an equivalence relation.

Example 3.3.13. Let S be {all points on the landmass of the Earth}, and define a relation $\sim$ by

$$\left(a \sim b\right) \Longleftrightarrow \left(\text{there is a land route from } a \text{ to } b\right).$$

Then, $\sim$ is an equivalence relation.

Example 3.3.14. Let

$$S = \mathbb{Z} \times \left(\mathbb{Z} \setminus \{0\}\right) = \left\{\left(a_1, a_2\right) \in \mathbb{Z}^2 \ \mid \ a_1 \in \mathbb{Z} \text{ and } a_2 \in \mathbb{Z} \setminus \{0\}\right\}.$$

Define a relation $\underset{*}{\sim}$ on S by

$$\left(\left(a_1, a_2\right) \underset{*}{\sim} \left(b_1, b_2\right)\right) \iff \left(a_1 b_2 = a_2 b_1\right).$$

This relation $\underset{*}{\sim}$ is an equivalence relation.

3.3.3. Equivalence classes

Definition 3.3.15. Let $\sim$ be an equivalence relation on a set S.

(a) For each $a \in S$, we define a subset $\left[a\right]_{\sim}$ of S by

$$\left[a\right]_{\sim} = \left\{b \in S \ \mid \ b \sim a\right\}.$$

This subset $\left[a\right]_{\sim}$ is called the **equivalence class of** a (for the relation $\sim$), or the **$\sim$-equivalence class of** a.

(b) The **equivalence classes of** $\sim$ are defined to be the sets $\left[a\right]_{\sim}$ for $a \in S$. They are also known as the **$\sim$-equivalence classes**.

Example 3.3.16. Consider the relation $\underset{3}{\equiv}$ on $\mathbb{Z}$. We have

$$\begin{aligned}[5]_{\underset{3}{\equiv}} &= \left\{ b \in \mathbb{Z} \mid b \underset{3}{\equiv} 5 \right\} = \{ b \in \mathbb{Z} \mid b \equiv 5 \bmod 3 \} \\ &= \{5 + 3k \mid k \in \mathbb{Z}\} = \{\ldots, -4, -1, 2, 5, 8, 11, 14, \ldots\}\end{aligned}$$

and

$$\begin{aligned}[3]_{\underset{3}{\equiv}} &= \left\{ b \in \mathbb{Z} \mid b \underset{3}{\equiv} 3 \right\} = \{ b \in \mathbb{Z} \mid b \equiv 3 \bmod 3 \} \\ &= \{3 + 3k \mid k \in \mathbb{Z}\} = \{\ldots, -6, -3, 0, 3, 6, 9, \ldots\}\end{aligned}$$

and

$$\begin{aligned}[2]_{\underset{3}{\equiv}} &= \left\{ b \in \mathbb{Z} \mid b \underset{3}{\equiv} 2 \right\} = \{ b \in \mathbb{Z} \mid b \equiv 2 \bmod 3 \} \\ &= \{2 + 3k \mid k \in \mathbb{Z}\} = \{\ldots, -4, -1, 2, 5, 8, 11, 14, \ldots\}.\end{aligned}$$

Note that $[2]_{\underset{3}{\equiv}} = [5]_{\underset{3}{\equiv}}$.

Let us state some basic properties of equivalence classes. For their proofs, see [**?**, §3.3.2] or any sufficiently thorough introduction to proofs.

Proposition 3.3.17. Let $\sim$ be an equivalence relation on a set S. Let $a \in S$. Then,

$$[a]_{\sim} = \{ b \in S \mid a \sim b \}.$$

Proposition 3.3.18. Let $\sim$ be an equivalence relation on a set S. Let $a \in S$. Then, $a \in [a]_{\sim}$.

Proposition 3.3.19. Let $\sim$ be an equivalence relation on a set S. Let $x, y \in S$.
(a) If $x \sim y$, then $[x]_{\sim} = [y]_{\sim}$.
(b) If not $x \sim y$, then $[x]_{\sim}$ and $[y]_{\sim}$ are disjoint.
(c) We have $x \sim y$ if and only if $x \in [y]_{\sim}$.
(d) We have $x \sim y$ if and only if $y \in [x]_{\sim}$.
(e) We have $x \sim y$ if and only if $[x]_{\sim} = [y]_{\sim}$.

As a consequence, any two $\sim$-equivalence classes are either identical or disjoint.

Example 3.3.20. Consider the relation $\underset{*}{\sim}$ defined on the set

$$S = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\}) = \left\{ (a_1, a_2) \in \mathbb{Z}^2 \mid a_1 \in \mathbb{Z} \text{ and } a_2 \in \mathbb{Z} \setminus \{0\} \right\}$$

by

$$\Big((a_1, a_2) \underset{*}{\sim} (b_1, b_2)\Big) \iff (a_1 b_2 = a_2 b_1)\,.$$

Then, the equivalence classes of $\underset{*}{\sim}$ are the rational numbers. This is how the rational numbers are rigorously defined! So the relation $\underset{*}{\sim}$ tells us which pairs of integers give the same rational number. The equivalence class $[(a_1, a_2)]_{\underset{*}{\sim}}$ is denoted by $\dfrac{a_1}{a_2}$.

Example 3.3.21. Let S be $\{$all points on the landmass of the Earth$\}$, and consider the relation $\sim$ defined by

$$(a \sim b) \iff (\text{there is a land route from } a \text{ to } b)\,.$$

Then, the $\sim$-equivalence classes are the continents/islands.

Example 3.3.22. Let A be a set, and $k \in \mathbb{N}$. The relation $\underset{\text{perm}}{\sim}$ on A^k is defined as follows:

$$\Big(\mathbf{p} \underset{\text{perm}}{\sim} \mathbf{q}\Big) \iff (\mathbf{p} \text{ is an anagram of } \mathbf{q})$$
$$\iff (\mathbf{p} \text{ can be obtained from } \mathbf{q} \text{ by permuting the entries})\,.$$

For example, $(3,8,8,2) \underset{\text{perm}}{\sim} (8,3,2,8)$.

The equivalence classes of $\underset{\text{perm}}{\sim}$ are called **unordered k-tuples** of elements of A. They are in bijection with size-k multisubsets of A.

Example 3.3.23. Let $n \in \mathbb{Z}$. The equivalence classes of the relation $\underset{n}{\equiv}$ are called "integers modulo n", and their set is denoted by $\mathbb{Z}/n\mathbb{Z}$ or $\mathbb{Z}_n$ or $\mathbb{Z}/n$ or Z_n or C_n (depending on author and context). In particular, the sets $\mathbb{Z}/12\mathbb{Z}$ and $\mathbb{Z}/24\mathbb{Z}$ have a fairly known meaning: they stand for the hours on a clock. For example, when you say "3 hours after 11 is 2", you are really saying that $11 + 3 \equiv 2 \bmod 12$ or, equivalently, $[11]_{\underset{12}{\equiv}} + [3]_{\underset{12}{\equiv}} = [2]_{\underset{12}{\equiv}}$ for an appropriate definition of addition of equivalence classes.

Next time, we will define box-equivalence and ball-equivalence on placements of balls into boxes. Then, by taking equivalence classes, we will be able to make sense of "unlabelled balls" and "unlabelled boxes".

Class of 2019-11-08

3.3.4. Defining unlabelled boxes and balls

What does it mean for balls, or boxes, to be unlabelled (in the twelvefold way)?

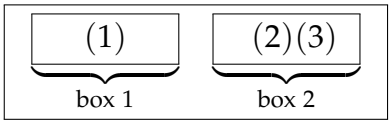
Rigorously, it means that we are counting
not the maps $f : A \to X$,
but their equivalence classes with respect to certain equivalence relations.
What relations?

Definition 3.3.24. Let $f, g : A \to X$. Then, we say that

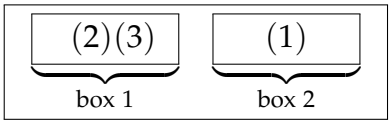
- f is **box-equivalent** to g (written $f \overset{\text{box}}{\sim} g$) if and only if there is a permutation σ of X such that $f = \sigma \circ g$ (in other words, f can be obtained from g by permuting boxes).
- f is **ball-equivalent** to g (written $f \overset{\text{ball}}{\sim} g$) if and only if there is a permutation τ of A such that $f = g \circ \tau$ (in other words, f can be obtained from g by permuting balls).
- f is **box-ball-equivalent** to g (written $f \overset{\text{box}}{\underset{\text{ball}}{\sim}} g$) if and only if there is a permutation σ of X and a permutation τ of A such that $f = \sigma \circ g \circ \tau$.

Example 3.3.25. Let $A = [3]$ and $X = [2]$.

(a) The two $L \to L$ placements

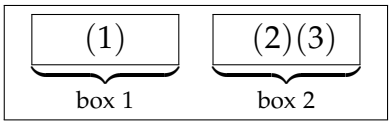


and

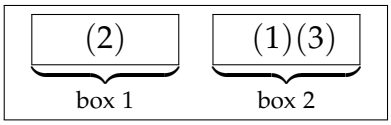


are box-equivalent to one another (since they can be obtained from one another by swapping the two boxes), but they are not ball-equivalent to one another (since a permutation of balls cannot change the fact that the first placement has 1 ball in the first box, whereas the second placement has 2).

(b) The two $L \to L$ placements

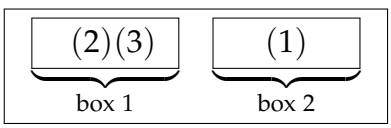


and

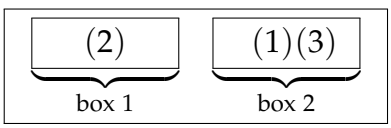


are ball-equivalent to one another (since they can be obtained from one another by swapping the balls 1 and 2), but they are not box-equivalent to one another (since a permutation of boxes cannot change the fact that balls 2 and 3 are together in a single box in the first placement, but not in the second).

(c) The two $L \to L$ placements

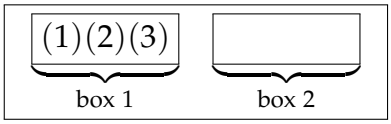


and

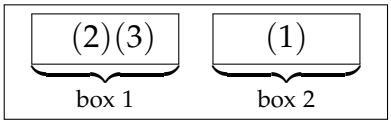


are box-ball-equivalent to one another (since they can be obtained from one another by swapping the two boxes and swapping balls 1 and 2). But they are neither box-equivalent nor ball-equivalent.

(d) The two $L \to L$ placements



and



are not box-ball-equivalent to one another (since no permutation of boxes or of balls can change the fact that the first placement has an empty box, but the second does not).

Proposition 3.3.26. All of $\overset{\text{box}}{\sim}$, $\overset{\text{ball}}{\sim}$ and $\underset{\text{ball}}{\overset{\text{box}}{\sim}}$ are equivalence relations on the set $\{\text{maps } A \to X\}$.

Proof. Easy. Hint:

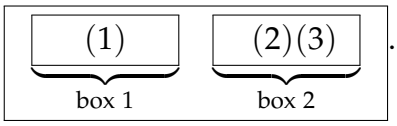
- If σ and τ are two permutations of a set S, then $\sigma \circ \tau$ is a permutation of S.
- If σ is a permutation of a set S, then σ^{-1} is a permutation of S.

□

Definition 3.3.27. Now, we define:

- $U \to L$ **placements** to be the $\overset{\text{ball}}{\sim}$-equivalence classes.
- $L \to U$ **placements** to be the $\overset{\text{box}}{\sim}$-equivalence classes.
- $U \to U$ **placements** to be the $\underset{\text{ball}}{\overset{\text{box}}{\sim}}$-equivalence classes.

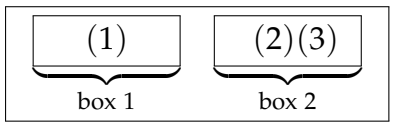
Example 3.3.28. Consider again our running example, with $X = [2]$ and $A = [3]$. Let g be the $L \to L$ placement

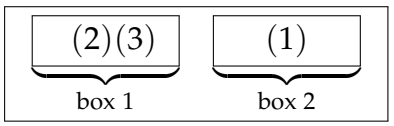


.

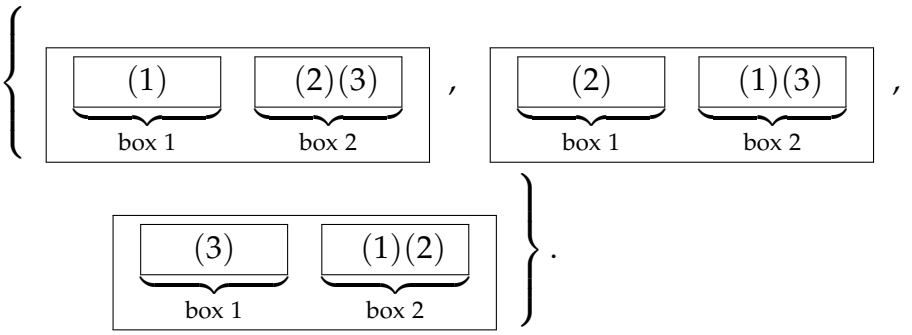
(a) The $\overset{\text{box}}{\sim}$-equivalence class of g is

$$\left\{ \quad , \quad \right\}.$$

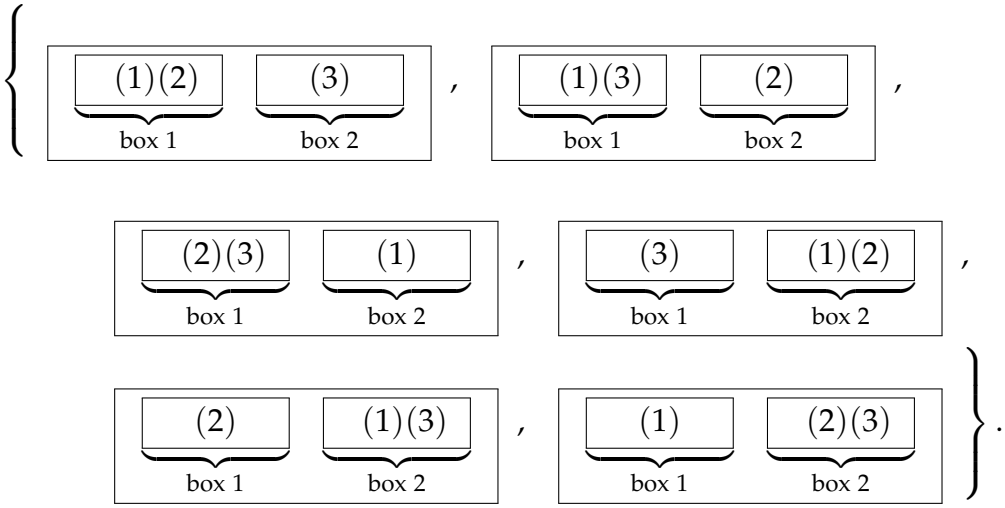




(b) The $\overset{\text{ball}}{\sim}$-equivalence class of g is



(c) The $\overset{\text{box}}{\underset{\text{ball}}{\sim}}$-equivalence class of g is



Recall the following crucial general fact about equivalence classes:

Proposition 3.3.29. Let $\sim$ be an equivalence relation on a set S. Let $x \in S$ and $y \in S$. Then, $x \sim y$ if and only if the $\sim$-equivalence classes of x and y are identical (that is, $[x]_\sim = [y]_\sim$).

Thus, "counting elements of S up to equivalence" means counting $\sim$-equivalence classes. Thus,

- by counting $\overset{\text{ball}}{\sim}$-equivalence classes, we will be counting placements of boxes in balls where the boxes are unlabelled;
- similarly for $\overset{\text{box}}{\sim}$ and $\overset{\text{box}}{\underset{\text{ball}}{\sim}}$.

3.4. $U \to L$

Recall: The $U \to L$ placements are the $\overset{\text{ball}}{\sim}$-equivalence classes.

Example 3.4.1. For $X = [2]$ and $A = [3]$, there are 4 different $U \to L$ placements:

TODO: picture.

Notation: When visualizing a $U \to L$ placement, we will just draw the balls as circles, without putting any numbers in them. So the 4 $U \to L$ placements for $X = [2]$ and $A = [3]$ are [TODO: picture].

Now, we are ready to count the $U \to L$ placements in the general setting:

Proposition 3.4.2. We have

$$\begin{aligned}&(\#\text{ of } U \to L \text{ placements } A \to X)\\ &= \left(\#\text{ of } \left(x_1, x_2, \ldots, x_{|X|}\right) \in \mathbb{N}^{|X|} \text{ satisfying } x_1 + x_2 + \cdots + x_{|X|} = |A|\right)\\ &= \binom{|A| + |X| - 1}{|A|}.\end{aligned}$$

Proof. The second equality sign follows from a theorem in the previous chapter (Theorem 2.10.4, more precisely the first equality in that theorem).

It remains to prove the first equality sign. First, we WLOG assume that $X = [|X|]$ (so the boxes are labelled $1, 2, \ldots, |X|$). Then, consider the bijection

$$\{U \to L \text{ placements } A \to X\} \to \left\{\left(x_1, x_2, \ldots, x_{|X|}\right) \in \mathbb{N}^{|X|} \mid x_1 + x_2 + \cdots + x_{|X|} = |A|\right\},$$

$$\boxed{a_1 \text{ balls}}\ \boxed{a_2 \text{ balls}}\ \cdots\ \boxed{a_{|X|} \text{ balls}} \mapsto \left(a_1, a_2, \ldots, a_{|X|}\right).$$

(More rigorously: The $\overset{\text{ball}}{\sim}$-equivalence class of a map $f : A \to X$ is sent to the weak composition

$$\left(\left|f^{-1}(1)\right|, \left|f^{-1}(2)\right|, \ldots, \left|f^{-1}(|X|)\right|\right),$$

where

$$f^{-1}(x) := \{a \in A \mid f(a) = x\}.$$

To see that this is a bijection, we need to show that if $f, g : A \to X$ are two maps such that

$$\left|f^{-1}(x)\right| = \left|g^{-1}(x)\right| \qquad \text{for all } x \in X,$$

then $f \overset{\text{ball}}{\sim} g$.)

Thus, the bijection principle yields

$$\begin{aligned}&(\#\text{ of } U \to L \text{ placements } A \to X)\\ &= \left(\#\text{ of } \left(x_1, x_2, \ldots, x_{|X|}\right) \in \mathbb{N}^{|X|} \text{ satisfying } x_1 + x_2 + \cdots + x_{|X|} = |A|\right).\end{aligned}$$

This is exactly the first equality sign of the proposition. □

Recall: An $L \to L$ placement is surjective if and only if each box has at least one ball in it.

We define surjectivity of $U \to L$ placements in the same way. Thus, a $\overset{\text{ball}}{\sim}$-equivalence class is a surjective $U \to L$ placement if and only if all its elements are surjective maps.

Proposition 3.4.3. We have

$$\begin{aligned}&(\#\text{ of surjective } U \to L \text{ placements } A \to X)\\ &= \left(\#\text{ of } \left(x_1, x_2, \ldots, x_{|X|}\right) \in \mathbb{P}^{|X|} \text{ satisfying } x_1 + x_2 + \cdots + x_{|X|} = |A|\right)\\ &= \binom{|A|-1}{|A|-|X|}.\end{aligned}$$

Proof. The first equality sign is proved similarly to the previous proposition.

The second equality sign follows from a theorem in the last chapter (Theorem 2.10.1). □

Recall: An $L \to L$ placement is injective if and only if each box has at most one ball in it.

We define injectivity of $U \to L$ placements in the same way. Thus, a $\overset{\text{ball}}{\sim}$-equivalence class is an injective $U \to L$ placement if and only if all its elements are injective maps.

Proposition 3.4.4. We have

$$\begin{aligned}&(\#\text{ of injective } U \to L \text{ placements } A \to X)\\ &= \left(\#\text{ of } \left(x_1, x_2, \ldots, x_{|X|}\right) \in \{0,1\}^{|X|} \text{ satisfying } x_1 + x_2 + \cdots + x_{|X|} = |A|\right)\\ &= \binom{|X|}{|A|}.\end{aligned}$$

Proof. The first equality is proven similarly to above.

The 2nd equality follows from Theorem 2.10.3. □

Thus, our table now looks as follows:

	arbitrary	injective	surjective
$L \to L$	$\lvert X\rvert^{\lvert A\rvert}$	$\lvert X\rvert^{\underline{\lvert A\rvert}}$	$\operatorname{sur}(\lvert X\rvert, \lvert A\rvert)$
$U \to L$	$\binom{\lvert A\rvert + \lvert X\rvert - 1}{\lvert A\rvert}$	$\binom{\lvert X\rvert}{\lvert A\rvert}$	$\binom{\lvert A\rvert - 1}{\lvert A\rvert - \lvert X\rvert}$
$L \to U$			
$U \to U$			

.

3.5. $L \to U$

Recall: The $L \to U$ placements are the $\overset{\text{box}}{\sim}$-equivalence classes.

Example 3.5.1. TODO: picture.

Proposition 3.5.2. We have

$$(\#\text{ of injective } L \to U \text{ placements}) = [\lvert A\rvert \leq \lvert X\rvert].$$

Proof. If $\lvert A\rvert > \lvert X\rvert$, then the Pigeonhole Principle shows that no such placements exist, and thus (# of injective $L \to U$ placements) $= 0$.

Now assume that $\lvert A\rvert \leq \lvert X\rvert$. Then, injective $L \to U$ placements do exist, and are identical; e.g.,

TODO: picture.

Thus, their # is $1 = [\lvert A\rvert \leq \lvert X\rvert]$. □

Recall: If $n \in \mathbb{N}$ and $k \in \mathbb{N}$, then $\left\{ n \atop k \right\} := \operatorname{sur}(n,k)/k!$ is called a **Stirling number of the 2nd kind**.

Proposition 3.5.3. We have

$$(\#\text{ of surjective } L \to U \text{ placements}) = \left\{ \lvert A\rvert \atop \lvert X\rvert \right\} = \frac{\operatorname{sur}(\lvert A\rvert, \lvert X\rvert)}{\lvert X\rvert!}.$$

Proof. The crux of the proof is the following claim:

> *Claim 1:* Each $\overset{\text{box}}{\sim}$-equivalence class of surjective maps (i.e., each surjective $L \to U$ placement) contains exactly $\lvert X\rvert!$ many maps $A \to X$.

[*Proof of Claim 1:* Consider the $\overset{\text{box}}{\sim}$-equivalence class of some surjective map $g : A \to X$. The elements of this class are all maps of the form $\sigma \circ g$ with σ a permutation of X. There are $|X|!$ many permutations of X, and they all lead to **distinct** maps $\sigma \circ g$ (because [TODO: picture] you can tell what $\sigma(x)$ is by looking at a ball placed in box x by g, and checking which box contains this ball in $\sigma \circ g$). Thus, there are exactly $|X|!$ many distinct maps $\sigma \circ g$ in the class of g. This proves Claim 1.]

Now,

$$\begin{aligned}
&\operatorname{sur}(|A|,|X|) \\
&= (\#\text{ of surjections } A \to X) \\
&= \sum_{\substack{C \text{ is a } \overset{\text{box}}{\sim}\text{-equivalence} \\ \text{class of surjections}}} \underbrace{|C|}_{\substack{=|X|! \\ \text{(by Claim 1)}}} \qquad \left(\begin{array}{c}\text{here, we have split the sum} \\ \text{according to the } \overset{\text{box}}{\sim}\text{-equivalence class}\end{array}\right) \\
&= \sum_{\substack{C \text{ is a } \overset{\text{box}}{\sim}\text{-equivalence} \\ \text{class of surjections}}} |X|! = \left(\#\text{ of } \overset{\text{box}}{\sim}\text{-equivalence classes of surjections}\right) \cdot |X|!.
\end{aligned}$$

□

Class of 2019-11-10

Proof. Thus, dividing by $|X|!$, we obtain

$$\begin{aligned}
&\left(\#\text{ of } \overset{\text{box}}{\sim}\text{-equivalence classes of surjections}\right) \\
&= \operatorname{sur}(|A|,|X|) / |X|! = \left\{ {|A| \atop |X|} \right\}
\end{aligned}$$

(since $\left\{ {|A| \atop |X|} \right\}$ was defined to be $\operatorname{sur}(|A|,|X|) / |X|!$). □

Proposition 3.5.4. We have

$$(\#\text{ of } L \to U \text{ placements } A \to X) = \left\{ {|A| \atop 0} \right\} + \left\{ {|A| \atop 1} \right\} + \cdots + \left\{ {|A| \atop |X|} \right\}.$$

Proof. We can prove a better claim: For each $k \in \{0,1,\ldots,|X|\}$, we have

$$(\#\text{ of } L \to U \text{ placements with exactly } k \text{ nonempty boxes}) = \left\{ {|A| \atop k} \right\}.$$

(This is not hard to see, because in an $L \to U$ placement, we can WLOG assume that all empty boxes are at the end, and thus we can simply ignore the empty boxes.)

Adding these equalities up for all $k \in \{0,1,\ldots,|X|\}$, we obtain precisely the claim of the proposition. (Details LTTR.) □

Thus, our table now looks as follows:

	arbitrary	injective	surjective
$L \to L$	$\lvert X\rvert^{\lvert A\rvert}$	$\lvert X\rvert^{\underline{\lvert A\rvert}}$	$\operatorname{sur}(\lvert X\rvert, \lvert A\rvert)$
$U \to L$	$\binom{\lvert A\rvert + \lvert X\rvert - 1}{\lvert A\rvert}$	$\binom{\lvert X\rvert}{\lvert A\rvert}$	$\binom{\lvert A\rvert - 1}{\lvert A\rvert - \lvert X\rvert}$
$L \to U$	$\left\{ {\lvert A\rvert \atop 0} \right\} + \left\{ {\lvert A\rvert \atop 1} \right\} + \cdots + \left\{ {\lvert A\rvert \atop \lvert X\rvert} \right\}$	$[\lvert A\rvert \leq \lvert X\rvert]$	$\left\{ {\lvert A\rvert \atop \lvert X\rvert} \right\}$
$U \to U$			

.

Let us say a few words about an equivalent version of surjective $L \to U$ placements: the set partitions.

Definition 3.5.5. Let S be a set.

(a) A **set partition** of S is a set $\mathcal{F}$ of disjoint nonempty subsets of S such that the union of these subsets is S.

In other words, a set partition of S is a set $\{S_1, S_2, \ldots, S_k\}$ of nonempty subsets of S such that each element of S lies in exactly one S_i. (Here, we are assuming that S is finite.)

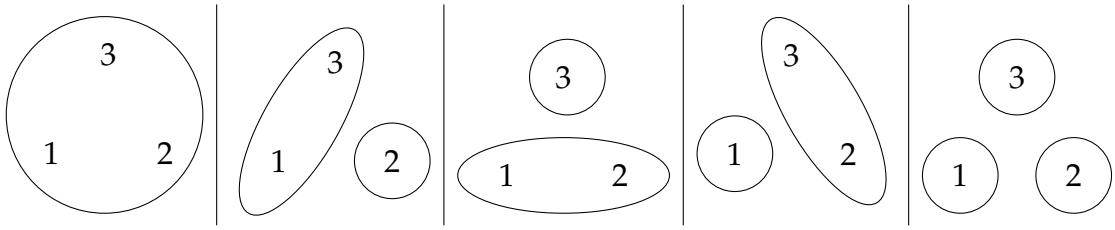
(b) If $\mathcal{F}$ is a set partition of S, then the elements of $\mathcal{F}$ are called the **parts** (or **blocks**) of $\mathcal{F}$. Keep in mind that they are subsets of S.

(c) If a set partition $\mathcal{F}$ of S has k parts, then we say that $\mathcal{F}$ is a **set partition of S into k parts**.

Example 3.5.6. Here are all set partitions of the set $[3] = \{1,2,3\}$:

$$\{\{1,2,3\}\}, \quad \{\{1,2\},\{3\}\}, \quad \{\{1,3\},\{2\}\}, \quad \{\{2,3\},\{1\}\}, \quad \{\{1\},\{2\},\{3\}\}.$$

And here are the same set partitions, drawn as pictures (each part of the set partition corresponds to a blob):



Proposition 3.5.7. Let A be an n-element set. Let $k \in \mathbb{N}$. Then,

$$(\#\text{ of set partitions of } A \text{ into } k \text{ parts}) = \left\{ {n \atop k} \right\}.$$

Proof. Let $X = [k]$. Then, there is a bijection

$$\{\text{set partitions of } A \text{ into } k \text{ parts}\} \to \{\text{surjective } L \to U \text{ placements } A \to X\},$$
$$\{S_1, S_2, \ldots, S_k\} \mapsto (\text{all elements in } S_i \text{ go into box } i).$$

(This is well-defined, because the resulting $L \to U$ placement does not depend on the order in which we have listed the blocks of our set partition; any two orders lead to $\overset{\text{box}}{\sim}$-equivalent maps.) □

Recall that we have shown a bunch of properties of $\operatorname{sur}(m,n)$. Since $\left\{ {m \atop n} \right\} = \operatorname{sur}(m,n)/n!$, we can translate them into properties of $\left\{ {m \atop n} \right\}$:

Proposition 3.5.8. Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$.

(a) We have $\left\{ {n \atop 0} \right\} = [n = 0]$.

(b) We have $\left\{ {0 \atop k} \right\} = [k = 0]$.

(c) We have $\left\{ {n \atop k} \right\} = 0$ if $k > n$.

(d) We have $\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k\left\{ {n-1 \atop k} \right\}$ if $n > 0$ and $k > 0$.

(e) We have $\left\{ {n \atop k} \right\} = \sum_{j=0}^{n-1} \binom{n}{j} \left\{ {j \atop k-1} \right\} / k$.

(f) We have $\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} i^n$.

Proof. **(a)** Follows from Proposition **?? (a)**.
(b) Follows from Proposition **?? (d)**.
(c) Follows from Proposition **?? (f)**.
(d) Follows from Proposition **??**.
(e) Follows from Proposition **??**.
(f) Follows from Theorem **??**. □

Remark 3.5.9. Let $n \in \mathbb{N}$. The **n-th Bell number** $B(n)$ is defined as the # of all set partitions of $[n]$. Thus,

$$B(n) = \left\{ {n \atop 0} \right\} + \left\{ {n \atop 1} \right\} + \cdots + \left\{ {n \atop n} \right\}.$$

For example, $B(3) = 5$.

There is no explicit formula for $B(n)$, but there is a recursion:

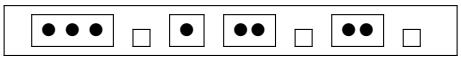
$$B(n+1) = \sum_{i=0}^{n} \binom{n}{i} B(i).$$

See [**?**, §0.3.2] for more about these.

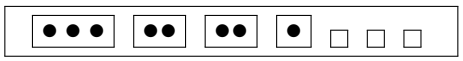
3.6. $U \to U$ and integer partitions

It remains to count $U \to U$ placements.

A $U \to U$ placement looks like this:



but the boxes, too, are interchangeable. Thus, we can order the boxes by decreasing number of balls:

 .

You can encode this $U \to U$ placement by a sequence of numbers, which say how many balls lie in each box:

$$(3,2,2,1,0,0,0)\,.$$

The decreasing order makes this sequence unique.

Let us introduce a name for such sequences, more precisely, for such sequences that don't contain zeroes:

Definition 3.6.1. A **partition** of an integer n is a weakly decreasing list $(a_1, a_2, \ldots, a_k)$ of positive integers whose sum is n (that is, $a_1 \geq a_2 \geq \cdots \geq a_k > 0$ and $a_1 + a_2 + \cdots + a_k = n$).

Instead of "partition", we can also say "integer partition".

The integers $a_1, a_2, \ldots, a_k$ are called the **parts** of the partition.

If a partition of n has k parts, then we say that it is a **partition of n into k parts**.

Example 3.6.2. The partitions of 5 are

$$(5)\,, \quad (4,1)\,, \quad (3,2)\,, \quad (3,1,1)\,, \quad (2,2,1)\,, \quad (2,1,1,1)\,, \quad (1,1,1,1,1)\,.$$

Remark 3.6.3. A partition of n is the same as a weakly decreasing composition of n.

Definition 3.6.4. Let $n \in \mathbb{Z}$ and $k \in \mathbb{N}$. Then, we set

$$p_k(n) = (\#\text{ of partitions of } n \text{ into } k \text{ parts})\,.$$

Example 3.6.5.

$$p_0(5) = 0, \quad p_1(5) = 1, \quad p_2(5) = 2, \quad p_3(5) = 2,$$
$$p_4(5) = 1, \quad p_5(5) = 1, \quad p_k(5) = 0 \text{ for } k > 5.$$

Proposition 3.6.6. **(a)** $p_k(n) = 0$ when $n < 0$.
(b) $p_k(n) = 0$ when $k > n$.
(c) $p_0(n) = [n = 0]$.
(d) $p_1(n) = [n > 0]$.
(e) $p_k(n) = p_k(n-k) + p_{k-1}(n-1)$ for all $n \in \mathbb{Z}$ and $k \geq 1$.
(f) $p_2(n) = \lfloor n/2 \rfloor$ if $n \geq 0$.

Proof. **(a)** A sum of positive integers is never negative.
(b) If $(a_1, a_2, \ldots, a_k)$ is a partition of n into k parts, then

$$n = \underbrace{a_1}_{\geq 1} + \underbrace{a_2}_{\geq 1} + \cdots + \underbrace{a_k}_{\geq 1} \geq \underbrace{1+1+\cdots+1}_{k \text{ times}} = k.$$

Thus, no such partition exists if $k > n$.
(c) A partition of n into 0 parts is a 0-tuple of positive integers whose sum is n. But the sum of a 0-tuple is always 0. So such a partition exists only for $n = 0$, and is unique. Thus, $p_0(n) = 1$ if $n = 0$ and is 0 otherwise.
(d) If $n > 0$, then there is only one partition of n into 1 part, namely the 1-tuple (n). If $n = 0$, then there is no partition of n into 1 part.
(e) Let us call a partition of n

- **red** if 1 is a part of it;
- **green** if 1 is not a part of it.

Any red partition of n must end with a 1 (since it contains a 1 but is weakly decreasing). Thus, there is a bijection

$$\{\text{red partitions of } n \text{ into } k \text{ parts}\} \to \{\text{partitions of } n-1 \text{ into } k-1 \text{ parts}\},$$
$$(a_1, a_2, \ldots, a_{k-1}, 1) \mapsto (a_1, a_2, \ldots, a_{k-1}).$$

Hence, the bijection principle yields

$$(\#\text{ of red partitions of } n \text{ into } k \text{ parts})$$
$$= (\#\text{ of partitions of } n-1 \text{ into } k-1 \text{ parts}) = p_{k-1}(n-1).$$

All entries of a green partition of n are > 1 (since they are positive integers and $\neq 1$). Thus, there is a bijection

$$\{\text{green partitions of } n \text{ into } k \text{ parts}\} \to \{\text{partition of } n-k \text{ into } k \text{ parts}\}$$
$$(a_1, a_2, \ldots, a_k) \mapsto (a_1 - 1, a_2 - 1, \ldots, a_k - 1).$$

Hence, the bijection principle yields

$$(\#\text{ of green partitions of } n \text{ into } k \text{ parts})$$
$$= (\#\text{ of partitions of } n-k \text{ into } k \text{ parts}) = p_k(n-k).$$

Adding these two equalities together, we get the claim of **(e)**.
(f) The partitions of n into 2 parts are

$$(n-1,1), \qquad (n-2,2), \qquad (n-3,3), \qquad \ldots, \qquad (\lceil n/2 \rceil, \lfloor n/2 \rfloor)$$

(where $\lceil x \rceil$ denotes the ceiling of the real number x). So there are $\lfloor n/2 \rfloor$ many of them. □

Proposition 3.6.7. We have

$$(\#\text{ of surjective } U \to U \text{ placements } A \to X) = p_{|X|}(|A|).$$

Proof. Proof idea: Encode a surjective $U \to U$ placement as a partition of $|A|$ into $|X|$ parts: namely the partition $\left(a_1, a_2, \ldots, a_{|X|}\right)$, where

$$a_i = (\#\text{ of balls in the box with the } i\text{-th largest \# of balls}).$$

This is a bijection. □

Proposition 3.6.8. We have

$$(\#\text{ of injective } U \to U \text{ placements } A \to X) = [|A| \leq |X|].$$

Proof. This follows from the corresponding fact about $L \to U$ placements. □

Proposition 3.6.9. We have

$$(\#\text{ of } U \to U \text{ placements } A \to X) = p_0(|A|) + p_1(|A|) + \cdots + p_{|X|}(|A|).$$

Proof. Similar to the proof for # of $L \to U$ placements. □

Thus, we have obtained a full table:

	arbitrary	injective	surjective
$L \to L$	$\lvert X\rvert^{\lvert A\rvert}$	$\lvert X\rvert^{\underline{\lvert A\rvert}}$	$\operatorname{sur}(\lvert X\rvert, \lvert A\rvert)$
$U \to L$	$\binom{\lvert A\rvert + \lvert X\rvert - 1}{\lvert A\rvert}$	$\binom{\lvert X\rvert}{\lvert A\rvert}$	$\binom{\lvert A\rvert - 1}{\lvert A\rvert - \lvert X\rvert}$
$L \to U$	$\left\{ {\lvert A\rvert \atop 0} \right\} + \left\{ {\lvert A\rvert \atop 1} \right\} + \cdots + \left\{ {\lvert A\rvert \atop \lvert X\rvert} \right\}$	$[\lvert A\rvert \leq \lvert X\rvert]$	$\left\{ {\lvert A\rvert \atop \lvert X\rvert} \right\}$
$U \to U$	$p_0(\lvert A\rvert) + p_1(\lvert A\rvert) + \cdots + p_{\lvert X\rvert}(\lvert A\rvert)$	$[\lvert A\rvert \leq \lvert X\rvert]$	$p_{\lvert X\rvert}(\lvert A\rvert)$

.

By the way, what if we add an extra column for "bijective"?

	arbitrary	injective	surjective	bijective
$L \to L$	$\lvert X\rvert^{\lvert A\rvert}$	$\lvert X\rvert^{\underline{\lvert A\rvert}}$	$\operatorname{sur}(\lvert X\rvert, \lvert A\rvert)$	$[\lvert A\rvert = \lvert X\rvert] \cdot \lvert X\rvert!$
$U \to L$	$\binom{\lvert A\rvert + \lvert X\rvert - 1}{\lvert A\rvert}$	$\binom{\lvert X\rvert}{\lvert A\rvert}$	$\binom{\lvert A\rvert - 1}{\lvert A\rvert - \lvert X\rvert}$	$[\lvert A\rvert = \lvert X\rvert]$
$L \to U$	$\left\{ {\lvert A\rvert \atop 0} \right\} + \left\{ {\lvert A\rvert \atop 1} \right\} + \cdots + \left\{ {\lvert A\rvert \atop \lvert X\rvert} \right\}$	$[\lvert A\rvert \leq \lvert X\rvert]$	$\left\{ {\lvert A\rvert \atop \lvert X\rvert} \right\}$	$[\lvert A\rvert = \lvert X\rvert]$
$U \to U$	$p_0(\lvert A\rvert) + p_1(\lvert A\rvert) + \cdots + p_{\lvert X\rvert}(\lvert A\rvert)$	$[\lvert A\rvert \leq \lvert X\rvert]$	$p_{\lvert X\rvert}(\lvert A\rvert)$	$[\lvert A\rvert = \lvert X\rvert]$

.

So that's not a very interesting column.

Class of 2019-11-15

3.7. Integer partitions (an introduction)

Recall the following notations we introduced (for $n \in \mathbb{Z}$ and $k \in \mathbb{N}$):

- $p(n) = (\#$ of partitions of $n)$.
- $p_k(n) := (\#$ of partitions of n into k parts$)$.

One of the propositions from last time says that

$$p_k(n) = p_k(n-k) + p_{k-1}(n-k) \qquad \text{for all } n \in \mathbb{Z} \text{ and } k \geq 1.$$

Also, it is obvious that

$$p(n) = p_0(n) + p_1(n) + \cdots + p_n(n).$$

Let us count partitions with some more special properties.

Definition 3.7.1. Let $n \in \mathbb{Z}$.
(a) Let

$$\begin{aligned} p_{\text{odd}}(n) &= (\#\text{ of partitions of } n \text{ into odd parts}) \\ &= (\#\text{ of partitions } (a_1, a_2, \ldots, a_k) \text{ of } n \text{ such that all } a_i \text{ are odd}). \end{aligned}$$

(b) Let

$$\begin{aligned} p_{\text{dist}}(n) &= (\#\text{ of partitions of } n \text{ into distinct parts}) \\ &= (\#\text{ of partitions } (a_1, a_2, \ldots, a_k) \text{ of } n \text{ such that } a_1 > a_2 > \cdots > a_k). \end{aligned}$$

Example 3.7.2. **(a)** We have

$$p_{\text{odd}}(7) = |\{(7), (5,1,1), (3,3,1), (3,1,1,1,1), (1,1,1,1,1,1,1)\}| = 5.$$

(b) We have

$$p_{\text{dist}}(7) = |\{(7), (6,1), (5,2), (4,3), (4,2,1)\}| = 5.$$

Theorem 3.7.3. (Euler) Let $n \in \mathbb{Z}$. Then, $p_{\text{odd}}(n) = p_{\text{dist}}(n)$.

Proof. Here is a rough outline. See **[?]** for a more detailed version (albeit with a slightly different version of the bijection).

We construct a map

$$A : \{\text{partitions of } n \text{ into odd parts}\} \to \{\text{partitions of } n \text{ into distinct parts}\}$$

which transforms a partition as follows: Repeatedly merge two equal parts until no more equal parts can be found. "Merging two equal parts" means replacing two equal parts a, a by the single part $2a$, and (if necessarily) rearranging the resulting tuple back into weakly decreasing order.

(*Examples:* Let us compute $A(5,5,3,1,1,1)$:

$$(\underline{5,5},3,1,1,1) \to (10,3,\underline{1,1},1) \to (10,3,2,1)$$

(where we underline equal entries that are about to get merged). Thus, $A(5,5,3,1,1,1) = (10,3,2,1)$.

Let us compute $A(5,3,1,1,1,1)$:

$$(5,3,\underline{1,1},1,1) \to (5,3,2,\underline{1,1}) \to (5,3,\underline{2,2}) \to (5,4,3).$$

Thus, $A(5,3,1,1,1,1) = (5,4,3)$.)

Why is this map A well-defined? This is not obvious.

Our definition of A was non-deterministic: It tells us to merge equal parts; but it does not tell us which equal parts to choose first (and there can be several choices). Thus, we have to prove that the result of our many merges does not depend on the order in which we do the merges.

One way to prove this is using something called the **diamond lemma**. Another way is by writing the parts of the partitions in binary (this is what Andrews does in **[?]**; he doesn't even talk about merging).

The inverse of A transforms a partition by repeatedly splitting even parts into two equal pieces.

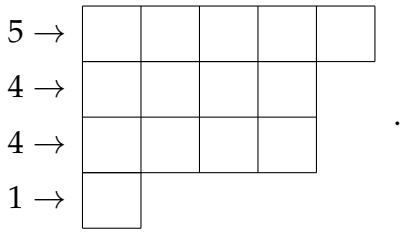
(The map A is called the Glaisher bijection; there are several other bijections that work.) □

Proposition 3.7.4. Let $n \in \mathbb{Z}$ and $k > 0$. Then,

$$p_k(n) = (\#\text{ of partitions of } n \text{ whose largest part is } k).$$

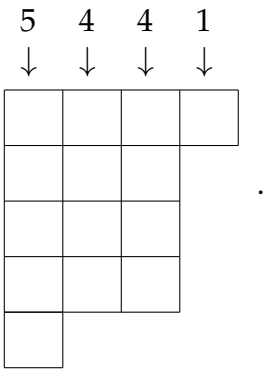
Example 3.7.5. We have $p_3(5) = (\#\text{ of partitions of } 5 \text{ whose largest part is } 3)$ (indeed, both numbers are 2). Indeed, the partitions of 5 into 3 parts are $(3,1,1)$ and $(2,2,1)$.

Proof. Picture proof: e.g., let $n = 14$ and $k = 4$. Start with the partition $\lambda = (5,4,4,1)$ of n into k parts. Draw a table of k left-aligned rows, where the length of each row equals the corresponding part of λ:



.

Now, flip the table across the main diagonal (i.e., the diagonal going from the top-left to the bottom-right), so that the rows become columns and vice versa:



.

The lengths of the rows of the resulting table again form a partition of n. The largest part of this new partition is k (because our original table had k rows, so the flipped table has k columns). In our example, this new partition is $(4,3,3,3,1)$.

Thus, we obtain a map

$$\{\text{partitions of } n \text{ into } k \text{ parts}\} \to \{\text{partitions of } n \text{ whose largest part is } k\},$$

which transforms a partition by flipping its table.

This map is a bijection; indeed, its inverse map is defined in the same way. This bijection is called **conjugation of partitions**. The bijection principle now yields the proposition.

(Note: The table that we constructed above is called the **Young diagram** of λ, or the **Ferrers diagram** of λ.) □

Definition 3.7.6. For any $k \in \mathbb{Z}$, define $w_k \in \mathbb{N}$ by

$$w_k = \frac{(3k-1)\,k}{2}.$$

This is called a **pentagonal number**.

Theorem 3.7.7. (Euler's recursion for the partition numbers)
For each $n > 0$, we have

$$\begin{aligned} p(n) &= p(n-1) + p(n-2) - p(n-5) - p(n-7) + p(n-12) + p(n-15) \pm \cdots \\ &= \sum_{\substack{k \in \mathbb{Z}; \\ k \neq 0}} (-1)^{k-1} p(n - w_k). \end{aligned}$$

We might get to prove this later on, using the technique of **generating functions**.

There is, of course, much more to say about partitions. See [**?**] or [**?**] for two introductions.

3.8. Odds and ends

Here are some random counting exercises.

Exercise 3.8.1. Given n persons ($n > 0$) and k tasks ($k > 0$).

(a) What is the # of ways to assign a task to each person such that each task has at least 1 person working on it?

(b) What if we additionally want to choose a leader for each task (among the people assigned to this task)?

(c) What if, instead, we want to choose a vertical hierarchy (between all people working on the task) for each task? (A "vertical hierachy" means a ranking of all people working on the task, with no ties.)

Example: Assume $n = 8$ and $k = 3$. Let our 8 people be $1, 2, 3, 4, 5, 6, 7, 8$, and let our 3 tasks be A, B, C.

(a) One option is

task	people working on it
A	$1, 2, 5$
B	3
C	$4, 6, 7, 8$

(b) One option is

task	people working on it
A	$1, 2, 5$ with leader 2
B	3 with leader 3
C	$4, 6, 7, 8$ with leader 7

(c) One option is

task	people working on it
A	$1 > 5 > 2$
B	3
C	$7 > 8 > 4 > 6$

(where the "$>$" signs stand for "is ranked above").

Proof. (Solution sketch.)

(a) $\operatorname{sur}(n,k)$.

Proof. Choosing such an arrangement is tantamount to choosing a surjection $\{\text{people}\} \to \{\text{tasks}\}$.

(b) $n^{\underline{k}} \cdot k^{n-k}$.

Proof. First, choose a leader for each task. There are $n^{\underline{k}}$ options for this. Then, every of the remaining $n-k$ people joins one of the k leaders. There are k^{n-k} options for this.

(c) $n! \cdot \dbinom{n-1}{k-1}$.

Proof. First, order all the n people in some way. There are $n!$ options for this. Then, split this ordering into k nonempty chunks. There are $\dbinom{n-1}{k-1}$ options for this, since we need to put $k-1$ separators into the $n-1$ positions between two consecutive people in our ordering. □

Exercise 3.8.2. Let $n \in \mathbb{N}$. How many compositions of n have the property that all entries of the composition belong to $\{1,2\}$?

(For example, for $n = 5$, these compositions are

$$(1,1,1,1,1)\,, \quad (1,1,1,2)\,, \quad (1,1,2,1)\,, \quad (1,2,1,1)\,,$$
$$(2,1,1,1)\,, \quad (2,2,1)\,, \quad (2,1,2)\,, \quad (1,1,2)\,.$$

Proof. Answer: the Fibonacci number f_{n+1}.

Proof 1: strong induction on n.

Proof 2: bijection to lacunar subsets.

Proof 3: bijection to domino tilings. □

Class of 2019-11-18

4. Permutations

4.1. Introduction

We will now talk about permutations in more detail. For deeper treatments, see [Bóna: Combinatorics of Permutations] and [Sagan: The symmetric group] and [Stanley: Enumerative Combinatorics, vol. 1. Ch. 1].

Recall: A **permutation** of a set X is a bijection from X to X.

4.2. Definitions

Definition 4.2.1. Let $n \in \mathbb{N}$. Let S_n be the set of all permutations of $[n]$. This set S_n is called the **n-th symmetric group**. It is closed under composition (i.e., for any $\alpha \in S_n$ and $\beta \in S_n$, we have $\alpha \circ \beta \in S_n$), and under inverses (i.e., for any $\sigma \in S_n$, we have $\sigma^{-1} \in S_n$) and contains $\mathrm{id}_{[n]}$.

Definition 4.2.2. Let $n \in \mathbb{N}$ and $\sigma \in S_n$. We introduce two notations for σ:

(a) The **one-line notation** of σ is the n-tuple $(\sigma(1), \sigma(2), \ldots, \sigma(n))$. (Conventionally, authors use square brackets for it, but we use parentheses.)

(b) The **cycle digraph** of σ is defined (informally) as follows:

For each $i \in [n]$, draw a point ("node") labelled i.

For each $i \in [n]$, draw an arrow ("arc") from the node labelled i to the node labelled $\sigma(i)$.

The result is called the **cycle digraph** of σ.

4.3. Transpositions and cycles

Definition 4.3.1. (a) Let i and j be two distinct elements of a set X.

Then, the **transposition** $t_{i,j}$ is the permutation of X that sends i to j, sends j to i, and leaves all other elements in their places.

If $X = [n]$ for some $n \in \mathbb{N}$, and if $i < j$, then the one-line notation of $t_{i,j}$ is

$$\left(\underbrace{1, 2, \ldots, i-1}_{\text{numbers from 1 to } i-1}, j, \underbrace{i+1, i+2, \ldots, j-1}_{\text{numbers from } i+1 \text{ to } j-1}, i, \underbrace{j+1, j+2, \ldots, n}_{\text{numbers from } j+1 \text{ to } n} \right).$$

(b) Let $n \in \mathbb{N}$ and $i \in [n-1]$. Then, the **simple transposition** s_i is defined by $s_i = t_{i,i+1} \in S_n$. So a simple transposition is a transposition that swaps two consecutive integers.

Convention 4.3.2. If α and β are two permutations of a set X, then we write $\alpha\beta$ for $\alpha \circ \beta$.

Also, $\alpha^i := \underbrace{\alpha \circ \alpha \circ \cdots \circ \alpha}_{i \text{ times}}$. If $i = 0$, this is understood to be id_X.

Proposition 4.3.3. Let $n \in \mathbb{N}$.
(a) We have $s_i^2 = \mathrm{id}$ for all $i \in [n-1]$. (Recall: $s_i^2 = s_i \circ s_i$.)
(b) We have $s_i s_j = s_j s_i$ for all $i, j \in [n-1]$ with $|i-j| > 1$.
(c) We have $s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1} = t_{i,i+2}$ for all $i \in [n-2]$. (This is known as "the braid relation for permutations".)

Proof. Straightforward verification that both sides send each $k \in [n]$ to the same value. □

Definition 4.3.4. Let $n \in \mathbb{N}$. Let w_0 be the permutation in S_n that sends each $i \in [n]$ to $n+1-i$.

In other words, it "reflects" all numbers from 1 to n across the middle of $[n]$. It is the unique strictly decreasing permutation of $[n]$.

Definition 4.3.5. Let X be a set. Let $i_1, i_2, \ldots, i_k$ be k distinct elements of X. Then,

$$\mathrm{cyc}_{i_1,i_2,\ldots,i_k}$$

means the permutation of X that sends $i_1 \mapsto i_2$, $i_2 \mapsto i_3$, $i_3 \mapsto i_4$, …, $i_{k-1} \mapsto i_k$, $i_k \mapsto i_1$ and leaves all other elements of X unchanged. This is called a **k-cycle**.

Remark 4.3.6. People often write $(i_1, i_2, \ldots, i_k)$ for $\mathrm{cyc}_{i_1,i_2,\ldots,i_k}$.

Remark 4.3.7. A permutation α is called an **involution** if $\alpha^2 = \mathrm{id}$. Both $t_{i,j}$ and w_0 are involutions. But k-cycles with $k > 2$ are not involutions.

Proposition 4.3.8. Let $n \in \mathbb{N}$.
(a) For any k distinct elements $i_1, i_2, \ldots, i_k$ of $[n]$, we have

$$\mathrm{cyc}_{i_1,i_2,\ldots,i_k} = \underbrace{t_{i_1,i_2} t_{i_2,i_3} \cdots t_{i_{k-1},i_k}}_{k-1 \text{ transpositions}}.$$

(b) For any $i \in [n]$ and $k \in \mathbb{N}$ such that $i + k - 1 \leq n$, we have

$$\mathrm{cyc}_{i,i+1,\ldots,i+k-1} = s_i s_{i+1} \cdots s_{i+k-2}.$$

(c) For any $i \in [n]$, we have $\mathrm{cyc}_i = \mathrm{id}$.
(d) For any distinct $i, j \in [n]$, we have $\mathrm{cyc}_{i,j} = t_{i,j}$.
(e) For any k distinct elements $i_1, i_2, \ldots, i_k$ of $[n]$, we have

$$\mathrm{cyc}_{i_1,i_2,\ldots,i_k} = \mathrm{cyc}_{i_k,i_1,i_2,\ldots,i_{k-1}}.$$

(f) For any k distinct elements $i_1, i_2, \ldots, i_k$ of $[n]$ and $\sigma \in S_n$, then

$$\sigma \, \mathrm{cyc}_{i_1,i_2,\ldots,i_k} \, \sigma^{-1} = \mathrm{cyc}_{\sigma(i_1),\sigma(i_2),\ldots,\sigma(i_k)}.$$

(g) If $1 \leq i < j \leq n$, then

$$\begin{aligned} t_{i,j} &= s_i s_{i+1} \cdots s_{j-1} \cdots s_{i+1} s_i \\ &\qquad \left(\begin{array}{c} \text{this is a product starting at } s_i\text{, then walking up to } s_{j-1} \\ \text{and back down to } s_i \end{array} \right) \\ &= s_{j-1} s_{j-2} \cdots s_i \cdots s_{j-2} s_{j-1} \\ &\qquad \left(\begin{array}{c} \text{this is a product starting at } s_{j-1}\text{, then walking down to } s_i \\ \text{and back up to } s_{j-1} \end{array} \right). \end{aligned}$$

(h) We have

$$\begin{aligned} w_0 &= \operatorname{cyc}_{1,2,\ldots,n} \operatorname{cyc}_{1,2,\ldots,n-1} \cdots \operatorname{cyc}_1 \\ &= \left(s_1 s_2 \cdots s_{n-1}\right) \left(s_1 s_2 \cdots s_{n-2}\right) \cdots \left(s_1 s_2\right) s_1 \\ &= \operatorname{cyc}_1 \operatorname{cyc}_{2,1} \operatorname{cyc}_{3,2,1} \cdots \operatorname{cyc}_{n,n-1,\ldots,1} \\ &= s_1 \left(s_2 s_1\right) \left(s_3 s_2 s_1\right) \cdots \left(s_{n-1} s_{n-2} \cdots s_1\right). \end{aligned}$$

Proof. All of these are doable with some bookkeeping and induction. See [detnotes, Chapter 5] for some of the proofs. □

4.4. Inversions and lengths

Definition 4.4.1. Let $n \in \mathbb{N}$ and $\sigma \in S_n$.

(a) An **inversion** of σ is a pair (i, j) of elements of $[n]$ such that $i < j$ and $\sigma(i) > \sigma(j)$.

(b) The **length** (or **Coxeter length**) of σ is the # of inversions of σ. It is called $\ell(\sigma)$. (In LaTeX: \ell)

Example 4.4.2. Let $\pi = (3,1,4,2) \in S_4$ (in one-line notation).

The inversions of π are

$$(1,4) \qquad \left(\text{since } 1 < 4 \text{ and } \underbrace{\pi(1)}_{=3} > \underbrace{\pi(4)}_{=2} \right) \qquad \text{and}$$

$$(3,4) \qquad \left(\text{since } 3 < 4 \text{ and } \underbrace{\pi(3)}_{=4} > \underbrace{\pi(4)}_{=2} \right) \qquad \text{and}$$

$$(1,2) \qquad \left(\text{since } 1 < 2 \text{ and } \underbrace{\pi(1)}_{=3} > \underbrace{\pi(2)}_{=1} \right).$$

So the length of π is 3.

Remark 4.4.3. If $\sigma \in S_n$, then $0 \leq \ell(\sigma) \leq \binom{n}{2}$.

The only $\sigma \in S_n$ with $\ell(\sigma) = 0$ is id.

The only $\sigma \in S_n$ with $\ell(\sigma) = \binom{n}{2}$ is w_0.

Inbetween, there are many permutations with a given $\ell(\sigma)$.

2019-11-20 notes

Remark 4.4.4. If $n \in \mathbb{N}$, then the permutations of $[n]$ can be represented as the vertices of a $(n-1)$-dimensional polyhedron in n-dimensional space. Namely, each permutation σ of $[n]$ gives rise to the point $(\sigma(1), \sigma(2), \ldots, \sigma(n)) \in \mathbb{R}^n$, and the convex hull of all these points is the polyhedron. This is called the **permutahedron**.

Proposition 4.4.5. For every $\sigma \in S_n$, we have $\ell\left(\sigma^{-1}\right) = \ell(\sigma)$.

Proof. Recall: An inversion of σ is a pair $(i,j) \in [n] \times [n]$ such that $i < j$ and $\sigma(i) > \sigma(j)$. An inversion of σ^{-1} is a pair $(u,v) \in [n] \times [n]$ such that $u < v$ and $\sigma^{-1}(u) > \sigma^{-1}(v)$.

The map

$$\{\text{inversions of } \sigma\} \to \left\{\text{inversions of } \sigma^{-1}\right\},$$
$$(i,j) \mapsto (\sigma(j), \sigma(i))$$

is well-defined and bijective (its inverse map sends each $(u,v) \in \left\{\text{inversions of } \sigma^{-1}\right\}$ to $\left(\sigma^{-1}(v), \sigma^{-1}(u)\right)$). So the bijection principle yields $\ell(\sigma) = \ell\left(\sigma^{-1}\right)$.

(For details, see [detnotes, Exercise 5.2 **(f)**].) □

Recall from last time

- the transpositions $t_{i,j}$ (swapping i with j while leaving all other numbers unchanged), and
- the simple transpositions $s_i = t_{i,i+1}$.

Proposition 4.4.6. Let $n \in \mathbb{N}$, $\sigma \in S_n$ and $k \in [n-1]$.

(a) We have

$$\ell(\sigma \circ s_k) = \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma(k) < \sigma(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma(k) > \sigma(k+1). \end{cases}$$

(b) We have

$$\ell(s_k \circ \sigma) = \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma^{-1}(k) < \sigma^{-1}(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma^{-1}(k) > \sigma^{-1}(k+1). \end{cases}$$

[**Note:** $\sigma^{-1}(i)$ is the **position** in which i appears in the one-line notation of σ. For example, if $\sigma = (5,1,2,3,6,4)$, then $\sigma^{-1}(6) = 5$.]

Proof. **(b)** How do the inversions of $s_k \circ \sigma$ differ from the inversions of σ ?

I claim that they are the same, except that

- if $\sigma^{-1}(k) < \sigma^{-1}(k+1)$, then we gain a new inversion $\left(\sigma^{-1}(k), \sigma^{-1}(k+1)\right)$;
- if $\sigma^{-1}(k) > \sigma^{-1}(k+1)$, then we lose an existing inversion $\left(\sigma^{-1}(k+1), \sigma^{-1}(k)\right)$.

For example, let

$$n = 7 \qquad \text{and} \qquad \sigma = (6,2,4,1,7,3,5) \qquad \text{and} \qquad k = 4.$$

Thus, $\sigma^{-1}(k) < \sigma^{-1}(k+1)$ and $s_k \circ \sigma = (6,2,5,1,7,3,4)$. So we gain a new inversion $(3,7)$ since $5 > 4$, but all other inversions remain the same.

For another example, let

$$n = 7 \qquad \text{and} \qquad \sigma = (3,1,4,2,7,6,5) \qquad \text{and} \qquad k = 3.$$

Thus, $\sigma^{-1}(k) > \sigma^{-1}(k+1)$ and $s_k \circ \sigma = (2,1,4,3,7,6,5)$. So we lose the inversion $(1,4)$ since $3 > 2$, but all other inversions remain the same.

Thus, when going from σ to $s_k \circ \sigma$, the # of inversions increases by 1 if $\sigma^{-1}(k) < \sigma^{-1}(k+1)$ and decreases by 1 if $\sigma^{-1}(k) > \sigma^{-1}(k+1)$. This is precisely the claim of **(b)**.

(a) Apply part **(b)** to σ^{-1} instead of σ. We get

$$\begin{aligned}
\ell\left(s_k \circ \sigma^{-1}\right) &= \begin{cases} \ell\left(\sigma^{-1}\right) + 1, & \text{if } \left(\sigma^{-1}\right)^{-1}(k) < \left(\sigma^{-1}\right)^{-1}(k+1); \\ \ell\left(\sigma^{-1}\right) - 1, & \text{if } \left(\sigma^{-1}\right)^{-1}(k) > \left(\sigma^{-1}\right)^{-1}(k+1) \end{cases} \\
&= \begin{cases} \ell(\sigma) + 1, & \text{if } \sigma(k) < \sigma(k+1); \\ \ell(\sigma) - 1, & \text{if } \sigma(k) > \sigma(k+1) \end{cases} \\
&\qquad \left(\text{since } \ell\left(\sigma^{-1}\right) = \ell(\sigma) \text{ (by previous Prop.) and } \left(\sigma^{-1}\right)^{-1} = \sigma\right).
\end{aligned}$$

On the other hand, the previous Prop. yields

$$\begin{aligned}
\ell\left(s_k \circ \sigma^{-1}\right) &= \ell\left(\underbrace{\left(s_k \circ \sigma^{-1}\right)^{-1}}_{\substack{=\left(\sigma^{-1}\right)^{-1} \circ (s_k)^{-1} \\ (\text{since } (\alpha\circ\beta)^{-1} = \beta^{-1}\circ\alpha^{-1})}}\right) = \ell\left(\underbrace{\left(\sigma^{-1}\right)^{-1}}_{=\sigma} \circ \underbrace{(s_k)^{-1}}_{\substack{=s_k \\ (\text{since } s_k \circ s_k = \mathrm{id})}}\right) \\
&= \ell(\sigma \circ s_k).
\end{aligned}$$

Comparing these equalities, we get

$$\ell(\sigma \circ s_k) = \begin{cases} \ell(\sigma)+1, & \text{if } \sigma(k) < \sigma(k+1); \\ \ell(\sigma)-1, & \text{if } \sigma(k) > \sigma(k+1). \end{cases}$$

(See [detnotes, Exercise 5.2 **(a)**] for a similar proof in more detail.) □

Remark 4.4.7. Let $n \in \mathbb{N}$ and $\sigma \in S_n$. Let $i, j \in [n]$ be such that $i < j$ and $\sigma(i) > \sigma(j)$. Is

$$\ell\left(\sigma \circ t_{i,j}\right) < \ell(\sigma) \ ?$$

It's not obvious, but the answer is "yes". See [18f-hw4s].

Recall: A **simple transposition** in S_n means one of the transpositions $s_1, s_2, \ldots, s_{n-1}$. We shall occasionally abbreviate "simple transposition" as "**simple**".

Theorem 4.4.8. Let $n \in \mathbb{N}$ and $\sigma \in S_n$.

(a) We can write σ as a composition of $\ell(\sigma)$ simples.

(b) $\ell(\sigma)$ is the smallest $p \in \mathbb{N}$ such that we can write σ as a composition of p simples.

[**Keep in mind:** The composition of 0 simples is id.]

Example 4.4.9. In S_4, we have

$$\underbrace{(4,1,3,2)}_{\text{one-line not.}} = \underbrace{s_2 \circ s_3 \circ s_2}_{=s_3 \circ s_2 \circ s_3} \circ s_1 = s_3 \circ s_2 \circ \underbrace{s_3 \circ s_1}_{=s_1 \circ s_3} = s_3 \circ s_2 \circ s_1 \circ s_3$$
$$= s_2 \circ s_1 \circ s_1 \circ s_3 \circ s_2 \circ s_1 = \cdots.$$

Proof. (Proof of Theorem.)

(a) Induction on $\ell(\sigma)$.

Induction base: If $\ell(\sigma) = 0$, then $\sigma = \mathrm{id}$, so we can write σ as a composition of 0 simples.

Induction step: Fix $h \in \mathbb{N}$. Assume (as the IH) that Theorem **(a)** holds for $\ell(\sigma) = h$.

Now, let $\sigma \in S_n$ be such that $\ell(\sigma) = h+1$.

Hence, $\ell(\sigma) = h+1 > 0$, so $\sigma \neq \mathrm{id}$.

Therefore, there exists some $k \in [n-1]$ such that $\sigma(k) > \sigma(k+1)$ (because otherwise, we we would have $\sigma(1) \leq \sigma(2) \leq \cdots \leq \sigma(n)$, and this would imply $\ell(\sigma) = 0$, whence $\sigma = \mathrm{id}$). Fix such a k.

Part **(a)** of the previous proposition yields

$$\ell(\sigma \circ s_k) = \begin{cases} \ell(\sigma)+1, & \text{if } \sigma(k) < \sigma(k+1); \\ \ell(\sigma)-1, & \text{if } \sigma(k) > \sigma(k+1) \end{cases} = \underbrace{\ell(\sigma)}_{=h+1} - 1$$
$$= (h+1) - 1 = h.$$

Thus, the IH (applied to $\sigma \circ s_k$ instead of σ) shows that we can write $\sigma \circ s_k$ as a composition of $\ell(\sigma \circ s_k) = h$ simples:

$$\sigma \circ s_k = s_{i_1} \circ s_{i_2} \circ \cdots \circ s_{i_h} \qquad \text{for some } i_1, i_2, \ldots, i_h \in [n-1].$$

Composing both sides of this equality with $(s_k)^{-1}$, we obtain

$$\begin{aligned}\sigma &= \left(s_{i_1} \circ s_{i_2} \circ \cdots \circ s_{i_h}\right) \circ \underbrace{(s_k)^{-1}}_{=s_k} \\ &= \left(s_{i_1} \circ s_{i_2} \circ \cdots \circ s_{i_h}\right) \circ s_k = s_{i_1} \circ s_{i_2} \circ \cdots \circ s_{i_h} \circ s_k.\end{aligned}$$

This shows that we can write σ as a composition of $h+1 = \ell(\sigma)$ simples. Thus, Theorem **(a)** holds for $\ell(\sigma) = h+1$. This completes the induction step, and so Theorem **(a)** is proved by induction.

[The idea behind this proof is known as "bubblesort".]

(b) Omitted.

(For details, see [detnotes, Exercise 5.2 **(g)**].) □

Corollary 4.4.10. Let $n \in \mathbb{N}$.

(a) We have $\ell(\sigma \circ \tau) \equiv \ell(\sigma) + \ell(\tau) \bmod 2$ for all $\sigma \in S_n$ and $\tau \in S_n$. (In other words, if $\ell(\sigma \circ \tau)$ is even, then $\ell(\sigma) + \ell(\tau)$ is even, and the same for "odd".)

(b) We have $\ell(\sigma \circ \tau) \leq \ell(\sigma) + \ell(\tau)$.

(c) If $\sigma = s_{k_1} \circ s_{k_2} \circ \cdots \circ s_{k_q}$, then $q \geq \ell(\sigma)$ and $q \equiv \ell(\sigma) \bmod 2$.

Example 4.4.11. Let $n = 4$ and $\sigma = (3,2,1,4)$ and $\tau = (3,1,4,2)$ in one-line notation. Then, $\ell(\sigma) = 3$ and $\ell(\tau) = 3$. Now, $\sigma \circ \tau = (1,3,4,2)$ has $\ell(\sigma \circ \tau) = 2$.

Corollary **(a)** says $\ell(\sigma \circ \tau) \equiv \ell(\sigma) + \ell(\tau) \bmod 2$. In other words, $2 \equiv 3 + 3 \bmod 2$.

Corollary **(b)** says $\ell(\sigma \circ \tau) \leq \ell(\sigma) + \ell(\tau)$. In other words, $2 \leq 3 + 3$.

Proof. (Proof of Corollary) See [detnotes, Exercises 5.2 and 5.3]. (Or do it yourself, e.g., by induction on $\ell(\sigma)$.) □

Proposition 4.4.12. Let $n \in \mathbb{N}$.

(a) We have $\ell(s_k) = 1$ for any $k \in [n-1]$.

(b) We have $\ell\left(t_{i,j}\right) = 2|i-j| - 1$ for any distinct $i, j \in [n]$.

(c) *We have* $\ell\left(\mathrm{cyc}_{i,i+1,\ldots,i+k-1}\right) = k-1$ for all i, k.

(d) We have $\ell\left(\mathrm{cyc}_{i_1,i_2,\ldots,i_k}\right) \geq k-1$ for all distinct $i_1, i_2, \ldots, i_k \in [n]$.

2019-11-20 notes

(e) We have $\ell(\mathrm{id}) = 0$ and $\ell(w_0) = \binom{n}{2}$. (Recall: w_0 is the "reflection across the middle of $[n]$".)

Proof. **(a)** follows from **(b)**.
(b) is [detnotes, Exercise 5.10], but also easy to check.
(c) and **(d)** are parts of [detnotes, Exercise 5.16].
(e) is trivial. □

Remark 4.4.13. For a given k and n, how many $\sigma \in S_n$ have length k ?

- The # of $\sigma \in S_n$ having $\ell(\sigma) = 0$ is 1 (namely, just $\sigma = \mathrm{id}$).
- The # of $\sigma \in S_n$ having $\ell(\sigma) = 1$ is $n-1$ (namely, just $\sigma = s_k$ with $k \in [n-1]$).
- The # of $\sigma \in S_n$ having $\ell(\sigma) = 2$ is $n(n+1)/2$. (Exercise.)

What about the general case?
There is no explicit formula, but there is a generating function:

Proposition 4.4.14. Let $n \in \mathbb{N}$. Then,

$$\begin{aligned}\sum_{w \in S_n} x^{\ell(w)} &= \prod_{i=1}^{n-1} \left(1 + x + x^2 + \cdots + x^i\right) \\ &= (1+x) \cdot \left(1 + x + x^2\right) \cdot \left(1 + x + x^2 + x^3\right) \cdot \cdots \cdot \left(1 + x + x^2 + \cdots + x^{n-1}\right).\end{aligned}$$

Example 4.4.15. Applying this to $n = 3$, we obtain

$$\sum_{w \in S_3} x^{\ell(w)} = (1+x) \cdot \left(1 + x + x^2\right).$$

Let us check this:

$$\begin{aligned}\sum_{w \in S_3} x^{\ell(w)} &= x^{\ell(1,2,3)} + x^{\ell(1,3,2)} + x^{\ell(2,1,3)} + x^{\ell(2,3,1)} + x^{\ell(3,1,2)} + x^{\ell(3,2,1)} \\ &\qquad \text{(where we are writing each permutation in one-line notation)} \\ &= x^0 + x^1 + x^1 + x^2 + x^2 + x^3 = x^0 + 2x^1 + 2x^2 + x^3 \\ &= (1+x) \cdot \left(1 + x + x^2\right).\end{aligned}$$

4.5. Descents

Definition 4.5.1. Let $n \in \mathbb{N}$ and $\sigma \in S_n$.
A **descent** of σ means a $k \in [n-1]$ such that $\sigma(k) > \sigma(k+1)$.
The **descent set** of σ, denoted $\operatorname{Des}\sigma$, is the set of all descents of σ.

Exercise 4.5.1. Fix $n \geq 4$.
(a) How many $\sigma \in S_n$ have 0 descents?
(b) How many $\sigma \in S_n$ have 1 descent?
(c) How many $\sigma \in S_n$ have $n-1$ descents?
(d) How many $\sigma \in S_n$ satisfy $1 \in \operatorname{Des}\sigma$ (that is, $\sigma(1) > \sigma(2)$) ?
(e) How many $\sigma \in S_n$ satisfy $1,2 \in \operatorname{Des}\sigma$ (that is, $\sigma(1) > \sigma(2) > \sigma(3)$) ?
(f) How many $\sigma \in S_n$ satisfy $1,3 \in \operatorname{Des}\sigma$ (that is, $\sigma(1) > \sigma(2)$ and $\sigma(3) > \sigma(4)$) ?

Proof. (Solution sketch.) **(a)** The answer is 1 (namely, $\sigma = \mathrm{id}$).
(d) The answer is $\dfrac{n!}{2}$.
First proof: The map

$$\{\sigma \in S_n \mid \sigma(1) > \sigma(2)\} \to \{\sigma \in S_n \mid \sigma(1) < \sigma(2)\}, \quad \sigma \mapsto \sigma \circ s_1$$

is bijective. So each of the 2 sets is half as large as S_n (because each $\sigma \in S_n$ satisfies either $\sigma(1) > \sigma(2)$ or $\sigma(1) < \sigma(2)$). But $|S_n| = n!$.
Second proof: To construct a $\sigma \in S_n$ satisfying $\sigma(1) > \sigma(2)$, we can can proceed as follows:

- Choose the **set** $\{\sigma(1), \sigma(2)\}$. There are $\dbinom{n}{2}$ options.
- Thus, $\sigma(1)$ and $\sigma(2)$ are already uniquely determined, because they have to satisfy $\sigma(1) > \sigma(2)$.
- Choose $\sigma(3), \sigma(4), \ldots, \sigma(n)$. There are $(n-2)!$ options.

$\Longrightarrow$ The total # is $\dbinom{n}{2} \cdot (n-2)! = \dfrac{n!}{2!} = \dfrac{n!}{2}$.
(e) The answer is $\dfrac{n!}{3!}$.
For the detailed proof, see [18s-mt1s], §0.2.
(f) The answer is $\dfrac{n!}{2! \cdot 2!} = \dfrac{n!}{4}$.
For the detailed proof, see [18s-mt1s], §0.2.
See [18f, some hw/mt] for the more general question how many permutations $\sigma \in S_n$ have a given bunch of numbers in their descent set.
(b) First of all, fix $i \in [n-1]$. How many $\sigma \in S_n$ have $\operatorname{Des}\sigma = \{i\}$?
In other words, how many $\sigma \in S_n$ satisfy

$$\sigma(1) < \sigma(2) < \cdots < \sigma(i) > \sigma(i+1) < \sigma(i+2) < \cdots < \sigma(n).$$

We have

$$(\#\text{ of } \sigma \in S_n \text{ such that } \sigma(1) < \sigma(2) < \cdots < \sigma(i) \text{ and } \sigma(i+1) < \sigma(i+2) < \cdots < \sigma(n)) = \binom{n}{i}$$

(because in order to construct such a σ, it suffices to choose the i-element subset $\{\sigma(1), \sigma(2), \ldots, \sigma(i)\}$ of $[n]$). Thus,

$$\begin{aligned}
&(\#\text{ of } \sigma \in S_n \text{ such that } \sigma(1) < \sigma(2) < \cdots < \sigma(i) > \sigma(i+1) < \sigma(i+2) < \cdots < \sigma(n)) \\
&= \underbrace{(\#\text{ of } \sigma \in S_n \text{ such that } \sigma(1) < \sigma(2) < \cdots < \sigma(i) \text{ and } \sigma(i+1) < \sigma(i+2) < \cdots < \sigma(n))}_{= \binom{n}{i}} \\
&- \underbrace{(\#\text{ of } \sigma \in S_n \text{ such that } \sigma(1) < \sigma(2) < \cdots < \sigma(i) < \sigma(i+1) < \sigma(i+2) < \cdots < \sigma(n))}_{=1} \\
&= \binom{n}{i} - 1.
\end{aligned}$$

Now, forget that we fixed i. Summing this over all $i \in [n-1]$, we get

$$\begin{aligned}
&(\#\text{ of } \sigma \in S_n \text{ that have exactly 1 descent}) \\
&= \sum_{i=1}^{n-1} \left(\binom{n}{i} - 1 \right) = \underbrace{\sum_{i=1}^{n-1} \binom{n}{i}}_{=2^n-2} - (n-1) = 2^n - 2 - (n-1) = 2^n - (n+1).
\end{aligned}$$

(c) Only 1 permutation $\sigma \in S_n$ has $n-1$ descents, namely w_0. □

4.6. Signs

Definition 4.6.1. Let $n \in \mathbb{N}$. The **sign** of a permutation $\sigma \in S_n$ is $(-1)^{\ell(\sigma)}$.
It is called $(-1)^\sigma$ or $\operatorname{sgn}(\sigma)$ or $\operatorname{sign}(\sigma)$ or $\varepsilon(\sigma)$.

Proposition 4.6.2. Let $n \in \mathbb{N}$.
(a) We have $(-1)^{\mathrm{id}} = 1$.
(b) We have $(-1)^{t_{i,j}} = -1$.
(c) We have $(-1)^{\mathrm{cyc}_{i_1,i_2,\ldots,i_k}} = (-1)^{k-1}$ for any distinct $i_1, i_2, \ldots, i_k \in [n]$.
(d) We have $(-1)^{\sigma\circ\tau} = (-1)^\sigma \cdot (-1)^\tau$ for any $\sigma, \tau \in S_n$. (In the lingo of abstract algebra, this is saying "The sign is a group homomorphism from S_n to $\{1, -1\}$".)
(e) We have $(-1)^{\sigma_1\circ\sigma_2\circ\cdots\circ\sigma_p} = (-1)^{\sigma_1}(-1)^{\sigma_2}\cdots(-1)^{\sigma_p}$ for any $\sigma_1, \sigma_2, \ldots, \sigma_p \in S_n$.
(f) We have $(-1)^{\sigma^{-1}} = (-1)^\sigma$ for any $\sigma \in S_n$. (The LHS has to be read as $(-1)^{(\sigma^{-1})}$.)

(g) We have

$$(-1)^{\sigma} = \prod_{1 \leq i < j \leq n} \frac{\sigma(i) - \sigma(j)}{i - j} \qquad \text{for each } \sigma \in S_n.$$

(h) If $x_1, x_2, \ldots, x_n$ are any numbers, and $\sigma \in S_n$, then

$$\prod_{1 \leq i < j \leq n} \left(x_{\sigma(i)} - x_{\sigma(j)} \right) = (-1)^{\sigma} \prod_{1 \leq i < j \leq n} (x_i - x_j).$$

Proof. See [detnotes, Chapter 5, section on signs]. Most of this follows easily from what we have proved above. □

Definition 4.6.3. Let $n \in \mathbb{N}$. A permutation $\sigma \in S_n$ is said to be

- **even** if $(-1)^{\sigma} = 1$ (that is, if $\ell(\sigma)$ is even);
- **odd** if $(-1)^{\sigma} = -1$ (that is, if $\ell(\sigma)$ is odd).

Corollary 4.6.4. Let $n \geq 2$. Then,

$$(\#\text{ of even } \sigma \in S_n) = (\#\text{ of odd } \sigma \in S_n) = n!/2.$$

Proof. The map

$$\{\text{even } \sigma \in S_n\} \to \{\text{odd } \sigma \in S_n\}, \quad \sigma \mapsto \sigma \circ s_1$$

is a bijection. □

Example 4.6.5. The 15-game

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 15 & 14 & \end{array} \quad \longrightarrow \quad \begin{array}{cccc} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & \end{array}$$

(via swaps of the empty cell with the neighboring square ("slides")) is unsolvable.

Why?

Proof sketch.

(a) Let us first consider the 3×3-analogue:

$$\begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 8 & 7 & \end{array} \quad \longrightarrow \quad \begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & \end{array}.$$

This is also impossible.

Proof: For each position P, let σ_P be the permutation of $[8]$ whose one-line notation is what you get if you read P row by row from left to right.

$$P = \begin{array}{ccc} a & b & c \\ d & e & \\ f & g & h \end{array} \mapsto \sigma_P = (a,b,c,d,e,f,g,h) \text{ (in one-line notation).}$$

Now, a slide changes σ_P **either** not at all **or** by multiplying it with a $\underbrace{\text{3-cycle}}_{\substack{\text{cyc}_{p,q,r} \\ \text{for some } p,q,r}}$ (that is, σ_P becomes $\sigma_P \circ \text{cyc}_{p,q,r}$). Thus, the sign of σ_P never changes (since 3-cycles have sign 1). Thus, the 3×3-game is unsolvable (since the initial position and the target position have different signs of σ_P).

(b) Now to the 4×4-version.

The sign of σ_P is no longer invariant. Instead, every vertical slide flips the sign of σ_P. Therefore,

$$(-1)^{\sigma_P} \cdot (-1)^{\text{which row has an empty square}}$$

is invariant. This again proves that the game is unsolvable, because the initial and target positions have different values of this invariant.

Remark 4.6.6. For any $n \in \mathbb{N}$, the set of all even permutations $\sigma \in S_n$ is called the **alternating group** A_n.

2019-12-04 notes

5. Lattice paths (brief introduction)

(See [18f], i.e., the notes from last year, for more details.)

Definition 5.0.1. The **integer lattice** (or, for short, **lattice**) is the set $\mathbb{Z}^2 = \mathbb{Z} \times \mathbb{Z}$.

Its elements are called **points**; indeed, any element $(a,b) \in \mathbb{Z}^2$ can be identified with the point with coordinates a and b on the plane.

Points can be added and subtracted entrywise: e.g.,

$$(a,b) - (c,d) = (a-c, b-d).$$

If $(a,b) \in \mathbb{Z}^2$ and $(c,d) \in \mathbb{Z}^2$ are two points, then a **lattice path** (for short, **LP**) from (a,b) to (c,d) means

- (informally) a path from (a,b) to (c,d) that uses only 2 kinds of steps:
  - "up-steps" (U) going $(p,q) \mapsto (p,q+1)$;
  - "right-steps" (R) going $(p,q) \mapsto (p+1,q)$.

- (rigorously) a tuple $(v_0, v_1, \ldots, v_n)$ of points $v_i \in \mathbb{Z}^2$ such that

$$v_0 = (a,b) \qquad \text{and} \qquad v_n = (c,d)$$

$$\text{and} \qquad v_i - v_{i-1} \in \left\{ \underbrace{(0,1)}_{\text{up-step}}, \ \underbrace{(1,0)}_{\text{right-step}} \right\} \qquad \text{for each } i \in [n].$$

Example 5.0.2. (See blackboard) is a LP from $(0,0)$ to $(5,3)$.
Formally, it is the 9-tuple

$$((0,0), (1,0), (1,1), (2,1), (3,1), \ldots).$$

It can be specified by its "step sequence" *RURRRUUR* (if its starting point is known to be $(0,0)$).

Proposition 5.0.3. Let $(a,b) \in \mathbb{Z}^2$ and $(c,d) \in \mathbb{Z}^2$ be two points. Then,

$$(\#\text{ of LPs from } (a,b) \text{ to } (c,d)) = \begin{cases} \dbinom{c+d-a-b}{c-a}, & \text{if } c+d \geq a+b; \\ 0, & \text{if } c+d < a+b \end{cases}.$$

Proof. In each LP, the x-coordinates of the points weakly increase at each step, and so do the y-coordinates. Thus, LPs from (a,b) to (c,d) can only exist when $c \geq a$ and $d \geq b$.

Furthermore, each step of a LP increases

$$(\text{x-coordinate}) + (\text{y-coordinate})$$

by exactly 1. Hence, each LP $(v_0, v_1, \ldots, v_n)$ from (a,b) to (c,d) must have $n = c+d-a-b$. Thus, if $c+d < a+b$, then the # of LPs is 0. Otherwise, the bijection

$$\{\text{LPs from } (a,b) \text{ to } (c,d)\} \to \{(c-a)\text{-element subsets of } [c+d-a-b]\},$$
$$(v_0, v_1, \ldots, v_n) \mapsto \{i \in [n] \mid v_i - v_{i-1} = (1,0)\}$$

shows that the # of LPs is $\dbinom{c+d-a-b}{c-a}$. □

Definition 5.0.4. Let $\mathbf{v} = (v_0, v_1, \ldots, v_n)$ be a LP from (a,b) to (c,d). Let $p \in \mathbb{Z}^2$. We say that $p \in \mathbf{v}$ (in words: p **lies on v**) if $p \in \{v_0, v_1, \ldots, v_n\}$.

Exercise 5.0.1. Find the # of LPs $\mathbf{v}$ from $(0,0)$ to $(6,6)$ such that $(2,2) \in \mathbf{v}$.

Proof. (Solution sketch.) Each such $\mathbf{v}$ consists of a LP from $(0,0)$ to $(2,2)$ and a LP from $(2,2)$ to $(6,6)$. Thus, the product rule yields

$$\begin{aligned}&(\#\text{ of LPs } \mathbf{v} \text{ from } (0,0) \text{ to } (6,6) \text{ such that } (2,2) \in \mathbf{v}) \\ &= \underbrace{(\#\text{ of LPs from } (0,0) \text{ to } (2,2))}_{=\binom{2+2-0-0}{2-0}} \cdot \underbrace{(\#\text{ of LPs from } (2,2) \text{ to } (6,6))}_{=\binom{6+6-2-2}{6-2}}.\end{aligned}$$

□

Definition 5.0.5. A LP $\mathbf{v}$ is said to be **Catalan** if $x \geq y$ for each $(x,y) \in \mathbf{v}$.
(Visually, this means that $\mathbf{v}$ never strays above the diagonal $x = y$.)

(In [18f], I say "legal" instead of "Catalan".)

Definition 5.0.6. If $n, m \in \mathbb{Z}$, then we set

$$L_{n,m} = (\#\text{ of Catalan LPs from } (0,0) \text{ to } (n,m)).$$

Proposition 5.0.7. **(a)** We have $L_{n,m} = L_{n-1,m} + L_{n,m-1}$ for any $n \in \mathbb{Z}$ and $m \in \mathbb{Z}$ satisfying $n \geq m$ and $(n,m) \neq (0,0)$.
(b) If $n \in \mathbb{N}$ and $m \in \mathbb{N}$ satisfy $n < m$, then $L_{n,m} = 0$.
(c) If $n \in \mathbb{N}$ and $m \in \mathbb{N}$ satisfy $n \geq m-1$, then

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

(d) If $n \in \mathbb{N}$ and $m \in \mathbb{N}$ satisfy $n \geq m-1$, then

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

(e) For any $n \in \mathbb{N}$, we have

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

Proof. Main idea: Recall the notion of "upsided tuples" from MT3 exercise 4. There is a bijection

$$\begin{aligned}\{\text{Catalan LPs from } (0,0) \text{ to } (n,m)\} &\to \{\text{upsided } (n+m)\text{-tuples } (i_1, i_2, \ldots, i_{n+m}) \text{ with } i_1, i_2, \ldots, i_{n+m} \\ (v_0, v_1, \ldots, v_n) &\mapsto (i_1, i_2, \ldots, i_{n+m}), \qquad \text{where } i_k = [v_k - v_{k-1} = (1,0)].\end{aligned}$$

Thus, the claims of **(b)** and **(c)** follow from MT3 exercise 4. The claim of **(d)** follows by simple algebra from **(c)**. The claim of **(e)** follows by applying **(d)** to $m = n$. The claim of **(a)** follows by looking at the last step of a Catalan LP. □

Definition 5.0.8. For any $n \in \mathbb{N}$, the number $L_{n,n} = \dfrac{1}{n+1}\dbinom{2n}{n} = \dbinom{2n}{n} - \dbinom{2n}{n-1}$ is called the n**-th Catalan number** and is denoted by C_n.

See [Stanley, "Catalan numbers"]. A few other places where they appear are:

- Let us say that a permutation $\sigma \in S_n$ is **123-avoiding** if there exist no $i < j < k$ in $[n]$ such that $\sigma(i) < \sigma(j) < \sigma(k)$. Then, the # of 123-avoiding permutations $\sigma \in S_n$ is C_n.
- Consider all possible ways to fully parenthesize a given expression $a_1 + a_2 + \cdots + a_n$. For example, for $n = 4$, these are

$$\begin{aligned} &(a_1 + a_2) + (a_3 + a_4), && (a_1 + (a_2 + a_3)) + a_4, \\ &a_1 + ((a_2 + a_3) + a_4), && a_1 + (a_2 + (a_3 + a_4)), \\ &((a_1 + a_2) + a_3) + a_4. \end{aligned}$$

  For any $n \in \mathbb{N}$, there are C_{n-1} ways to do this.
- Fix an $n \geq 3$, and a convex n-gon G_n. How many ways are there to triangulate G_n (i.e., to subdivide G_n into triangles whose vertices are vertices of G_n) ?

  The answer is C_{n-2}.

There are several variations on Catalan numbers and Catalan LPs, such as r**-Catalan numbers**.

6. Generating functions (introduction)

We have already seen generating functions. Here are two more examples.

Basic idea: Any sequence $(a_0, a_1, a_2, \ldots)$ of numbers gives rise to a "power series" $a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots$, called its "**generating function**". What exactly it is is a question that would take us a while (see [Loehr, Ch. 7 (1st edition)] or [Niven, "Formal power series"] or [19s]), but let us just calculate with these generating functions as if they were polynomials in x. (Of course, they are not literally polynomials in x, since they can have infinitely many nonzero coefficients.)

Example 1. Recall the Fibonacci sequence $(f_0, f_1, f_2, \ldots)$ with

$$f_0 = 0, \qquad f_1 = 1, \qquad f_n = f_{n-1} + f_{n-2}.$$

Consider its gf (= generating function)

$$\begin{aligned} F(x) &= f_0 + f_1 x + f_2 x^2 + \cdots \\ &= 0 + 1x + 1x^2 + 2x^3 + 3x^4 + 5x^5 + 8x^6 + \cdots . \end{aligned}$$

Then,

$$\begin{aligned} F(x) &= f_0 + f_1 x + f_2 x^2 + f_3 x^3 + f_4 x^4 + \cdots \\ &= \underbrace{0 + 1x}_{=x} + \underbrace{(f_1 + f_0) x^2 + (f_2 + f_1) x^3 + (f_3 + f_2) x^4 + \cdots}_{\substack{= \left(f_1 x^2 + f_2 x^3 + f_3 x^4 + \cdots\right) \\ + \left(f_0 x^2 + f_1 x^3 + f_2 x^4 + \cdots\right)}} \\ &= x + \underbrace{\left(f_1 x^2 + f_2 x^3 + f_3 x^4 + \cdots\right)}_{\substack{= f_0 x + f_1 x^2 + f_2 x^3 + f_3 x^4 + \cdots \\ \text{(since } f_0 = 0\text{)}}} + \left(f_0 x^2 + f_1 x^3 + f_2 x^4 + \cdots\right) \\ &= x + \underbrace{\left(f_0 x + f_1 x^2 + f_2 x^3 + f_3 x^4 + \cdots\right)}_{= xF(x)} + \underbrace{\left(f_0 x^2 + f_1 x^3 + f_2 x^4 + \cdots\right)}_{= x^2 F(x)} \\ &= x + xF(x) + x^2 F(x) . \end{aligned}$$

Solving this equation for $F(x)$, we get

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

where $\phi = \dfrac{1 + \sqrt{5}}{2}$ and $\psi = \dfrac{1 - \sqrt{5}}{2}$ are the "golden ratios". Applying partial fraction decomposition to the RHS, we obtain

$$F(x) = \frac{x}{(1 - \phi x)(1 - \psi x)} = \frac{1}{\sqrt{5}} \cdot \frac{1}{1 - \phi x} - \frac{1}{\sqrt{5}} \cdot \frac{1}{1 - \psi x}.$$

Now, what are the coefficients of $\dfrac{1}{1 - \alpha x}$ for $\alpha \in \mathbb{C}$?

Well:

$$\frac{1}{1 - x} = 1 + x + x^2 + x^3 + \cdots ,$$

because

$$\begin{aligned} &(1 - x)\left(1 + x + x^2 + x^3 + \cdots\right) \\ &= \left(1 + x + x^2 + x^3 + \cdots\right) - x\left(1 + x + x^2 + x^3 + \cdots\right) \\ &= \left(1 + x + x^2 + x^3 + \cdots\right) - \left(x + x^2 + x^3 + \cdots\right) = 1. \end{aligned}$$

If we substitute αx for x here, we obtain

$$\begin{aligned}\frac{1}{1-\alpha x} &= 1+\alpha x+(\alpha x)^2+(\alpha x)^3+\cdots \\ &= 1+\alpha x+\alpha^2 x^2+\alpha^3 x^3+\cdots .\end{aligned}$$

Thus, our formula for $F(x)$ becomes

$$\begin{aligned}F(x) &= \frac{1}{\sqrt{5}}\cdot\frac{1}{1-\phi x}-\frac{1}{\sqrt{5}}\cdot\frac{1}{1-\psi x} \\ &= \frac{1}{\sqrt{5}}\cdot\left(1+\phi x+\phi^2 x^2+\phi^3 x^3+\cdots\right)-\frac{1}{\sqrt{5}}\cdot\left(1+\psi x+\psi^2 x^2+\psi^3 x^3+\cdots\right) \\ &= \frac{1-1}{\sqrt{5}}+\frac{\phi-\psi}{\sqrt{5}}x+\frac{\phi^2-\psi^2}{\sqrt{5}}x^2+\frac{\phi^3-\psi^3}{\sqrt{5}}x^3+\cdots .\end{aligned}$$

Now, comparing coefficients before x^n, we get

$$f_n = \frac{\phi^n-\psi^n}{\sqrt{5}} \qquad \text{for each } n\in\mathbb{N}.$$

This is exactly Binet's formula. Unlike the first time we saw it, we now have a motivated "proof" of it.

However, of course, this is only a proof if we can explain

- what a power series is;
- what x is;
- why we can divide by power series like $1-x-x^2$;
- why we can substitute αx for x into a power series;
- why we can expand infinite sums;
- ...

This is done nicely in [Loehr, Ch. 7 (1st edition)] and [19s] and in most detailed textbooks on abstract algebra.

[18f] gives a quick overview. [Niven, "Formal power series"].

[Wilf, "generatingfunctionology"].