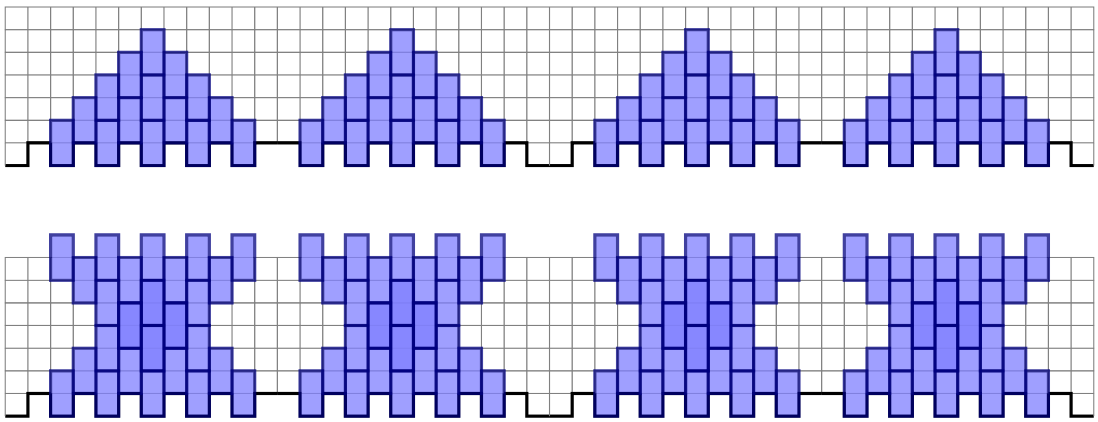
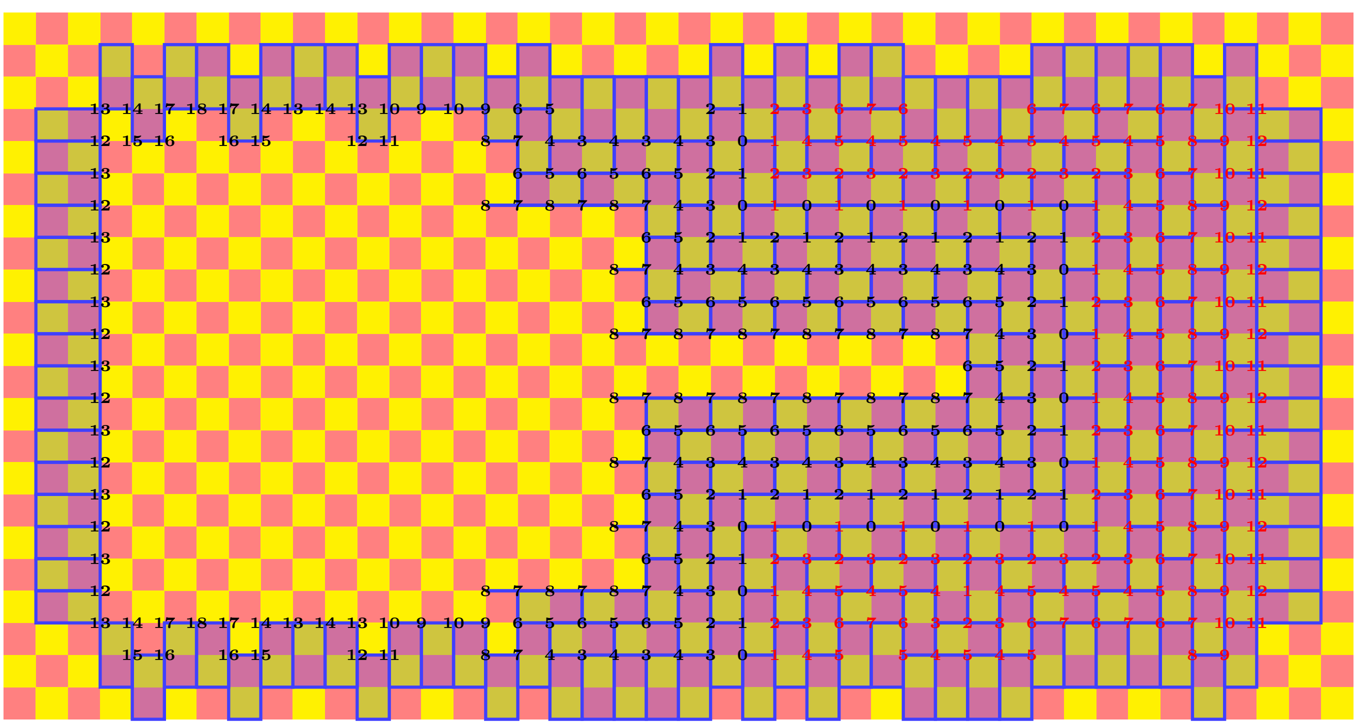


Basic Combinatorics: Part One

June 6, 2023

Shannon Starr
Department of Mathematics
University of Alabama at Birmingham
University Hall, Room 4005
1402 Tenth Avenue South
Birmingham, AL 35294-1241
sℓstarr at *uab . edu*



Chapter 1

Introduction

In these notes (short book) we gather some basic methods of combinatorics. These include proof-by-induction, direct bijective proofs, and variants of the latter such as double-counting. Another useful method is generating functions. For us these are the four topics we want to cover in combinatorics: proof by induction, proof by bijection, double-counting methods, and generating function methods.

Digression: Often generating functions are used in conjunction with complex analysis or Fourier analysis to obtain asymptotics. If one has a sequence $a_0, a_1, \ldots$ and one can determine a function $f(x)$ such that

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n + \ldots, \tag{1.1}$$

then Cauchy's integral formula gives a way to recover the numbers a_n from the values $f(x)$. Cauchy's integral formula says

$$a_n = \frac{1}{2\pi} \int_0^{2\pi} \frac{f(re^{i\theta})}{r^n e^{in\theta}}\, d\theta\,, \tag{1.2}$$

for any choice of $r > 0$ such that the integral makes sense, where $i = \sqrt{-1}$ is the imaginary number and an important fact is Euler's formula that $e^{i\theta} = \cos(\theta) + i\sin(\theta)$. The derivation and analysis of formulas such as (1.2) could form the subject for an entire course on its own with many valuable applications[1].

We will not consider formulas such as (1.2) any further in these notes because we will not assume familiarity with complex numbers. We do not need to use them, for this basic course. The necessary prerequisite for this class is just calculus of derivatives and integrals, as one usually sees in Calculus I and II.

Aside from our first goal, which is to introduce some key facts from combinatorics, a secondary important function is that it gives us a chance to become familiar with the level of rigor in proofs

[1]For example the function $\exp(x)$ defined to be e^x can be written as $1 + x + \cdots + x^n/n! + \ldots$. So $1/n!$ is equal to $(2\pi)^{-1} \int_0^{2\pi} \exp(re^{i\theta}) r^{-n} e^{-in\theta}\, d\theta$ according to (1.2). Applying Euler's formula and using some symmetry and algebra we can rewrite this as $(2\pi)^{-1} \int_{-\pi}^{\pi} e^{r\cos(\theta)} r^{-n} \cos\big(n\theta - r\sin(\theta)\big)\, d\theta$. Choosing $r = n$ and changing variables to $t = \theta\sqrt{n}$, we get $n! = n^n e^{-n}\, C_n \sqrt{n}$, where $\lim_{n\to\infty} C_n$ is the reciprocal of $(2\pi)^{-1} \int_{-\infty}^{\infty} e^{-t^2/2}\, dt$ (by direct analysis or DCT). If only we knew that $\int_{-\infty}^{\infty} e^{-t^2/2}\, dt = \sqrt{2\pi}$, this would be a derivation of Stirling's formula!

which will also be used in other topics, such as graph theory. Many basic results in graph theory are more closely related to proofs rather than to calculations.

These notes are written to be used in conjunction with a second set of notes, about graph theory. We will start with combinatorics. But then, in the same course, we will cover basic graph theory. We will follow our own set of notes: this set of notes, and a corresponding set of notes on graph theory.

But there are many other great references. One reference that we will consider frequently is written by Dr. Peter O'Neill, Professor Emeritus, also in the department of mathematics at UAB. We will put his notes in Canvas, just for our students.

An important fact to know is that combinatorics and graph theory appear as tools in most other subjects in mathematics, such as algebra and analysis. They also have applications to topics in science outside mathematics, such as physics or computer science. A good example of a subject that uses a lot of combinatorics is probability theory.

For example, one may think of the well-known central limit theorem in probability theory as a topic that can be seen by starting with combinatorics, asking how many "permutations" and "combinations" there are. But to get the actual central limit theorem, one then changes perspective to asymptotics, which is the study of limits of analytic functions. If the functions under consideration converge to 0 or ∞, then more refined asymptotic methods seek to study the order of convergence.

Stirling's formula is a famous example in the asymptotics of combinatorial sequences. It is closely connected to the de Moivre, Laplace limit theorem. Indeed, Abraham de Moivre asked his friend James Stirling to study the asymptotics of $n! = n(n-1)(n-2)\cdots(3)(2)(1)$, to help him with his own result, the first version of the central limit theorem.

Graph theory gets used in more advanced courses in computation and electrical engineering, such as network analysis. We will consider that topic in a second set of notes. But there is an interesting intersection of combinatorics and graph theory, which is why these two topics often are taught together in one course. For example, counting the number of spanning trees of a connected graph can be done using Kirchoff's formula, called the matrix-tree theorem, which is also closely connected to the Cauchy-Binet theorem (a generalization of the basic formula that guarantees that the determinant is a multiplicative function on matrices). So there are some topics at the end of the graph theory section that will use the tools we develop in these notes on combinatorics.

1 Induction and the summation notation

An important first formula in combinatorics is the formula

$$1+2+3+\cdots+n\,=\,\frac{n(n+1)}{2}\,. \tag{1.3}$$

We will consider alternative proofs of this fact. In order to have a picture to consider, we have drawn an example for $n=6$, which is Figure 1.1.

It is relatively easy to see that this number is $n(n+1)/2$. To prove this we use induction. Let us define $a_0, a_1, a_2, \ldots$ such that we have

$$a_1\,=\,1\,,\qquad a_2\,=\,1+2\,,\qquad a_3\,=\,1+2+3\,,\ \text{ and generally }\ a_n\,=\,1+2+\cdots+n\,. \tag{1.4}$$

In order to match convention, let us also define

$$a_0\,=\,0\,. \tag{1.5}$$

1 + 2 + 3 + 4 + 5 + 6

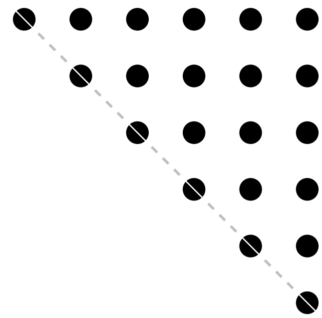


Figure 1.1: We include a picture for the sum of $1 + 2 + \cdots + n$ for $n = 6$. Counting, it is equal to $n(n + 1) = (6)(7)/2 = 21$. This is also $1/2$ of $n^2 + n$.

Suppose that for some $n = 0, 1, \ldots$ we have already proved that for that particular choice of n we do have $a_n = n(n + 1)/2$. This is called the "induction hypothesis." Then for $n + 1$ we have

$$a_{n+1} = 1 + 2 + \cdots + n + (n + 1)\,, \quad \text{which equals} \quad a_n + (n + 1)\,. \tag{1.6}$$

Then using the formula we have assumed for that choice of n, that $a_n = n(n + 1)/2$ we get

$$a_{n+1} = a_n + (n + 1) = \frac{n(n + 1)}{2} + (n + 1)\,. \tag{1.7}$$

Combining the two fractions with a common denominator, we get

$$\begin{aligned} a_{n+1} &= \frac{n(n + 1)}{2} + (n + 1) \\ &= \frac{n(n + 1)}{2} + \frac{2(n + 1)}{2} \\ &= \frac{n(n + 1) + 2(n + 1)}{2}\,, \end{aligned} \tag{1.8}$$

So factoring the common term $(n + 1)$ in the numerator, we get

$$a_{n+1} = \frac{(n + 2)(n + 1)}{2}\,. \tag{1.9}$$

If we define $m = n + 1$, then we get $a_m = m(m + 1)/2$, again. So we have obtained a piece of a proof, called the "induction step."

Now note that a_0, which was defined to be 0 is equal to $0(1)/2$ because that is also equal to 0. This is called the "initial step," of the proof. (We do not have to prove the steps in order, as long as we do not use *circular logic*.) So we have the following situation:

- We proved that $a_0 = 0$ which means $a_n = n(n + 1)/2$ for the special case $n = 0$. That is the initial step.

- We proved that for any $n = 0, 1, 2, \ldots,$ we have that if it has already been established that $a_n = n(n + 1)/2$ for that choice of n, then it is also true that $a_m = m(m + 1)/2$ for $m = n + 1$, as the induction step.

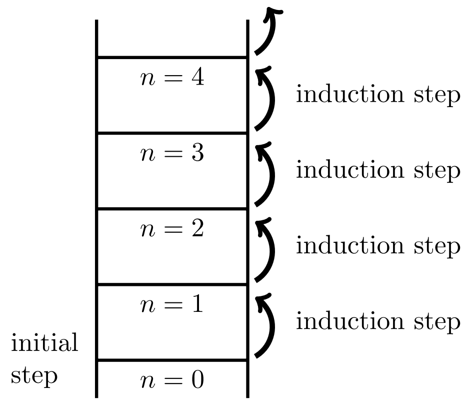


Figure 1.2: In proof-by-induction, we prove a particular formula such as $a_n = n(n+1)/2$ for an initial value, such as $n = 0$. We also prove the induction step: the formula for a_n implies the formula is also true for a_{n+1}. Then by climbing the ladder of the induction argument, applying the induction step iteratively, we conclude that the formula is then true for a_1 and then a_2, and so on. We accept that this proves it for all a_n.

- So setting $n = 0$ we do have $a_n = n(n+1)/2$. So because of the induction step, we conclude that $a_m = m(m+1)/2$ for $m = n+1 = 0+1 = 1$. In other words, if we set $n = 1$ then we get $a_n = n(n+1)/2$ (just like for $n = 0$).
- So, setting $n = 1$, we iterate the induction step. We get that $a_m = m(m+1)/2$ for $m = n+1 = 1+1 = 2$. So we also have $a_n = n(n+1)/2$ for the special choice of $n = 2$.

And we can continue that type of argument for as far as we like. For example, after $n = 2$, using the induction argument we can get to $n = 3$. Then using the induction step, we can get to $n = 4$. Then using the induction step, we can get to $n = 5$. Then we can get to $n = 6$, which is what is shown in Figure 1.1.

At this point, in our mathematical framework, we have to make a decision. We may all agree that the above type of argument implies that

$$a_n = \frac{n(n+1)}{2}, \quad \text{for all } n = 0, 1, \ldots. \tag{1.10}$$

But some mathematicians do not believe that. Those mathematicians are opposed to relegating to *mathematical analysis* any infinite operation[2].

This type of argument, called "proof-by-induction," clearly is an infinite operation. We are not saying that $a_n = n(n+1)/2$ only for all n's up to one million. We are not even saying that $a_n = n(n+1)/2$ only for all n's up to 2^{1024}, which is just a little more than the maximum floating-point number in IEEE double precision. We are saying that by climbing the induction steps arbitrarily high, we can prove that $a_n = n(n+1)/2$ for all the infinite number of choices of n equal to $0, 1, 2, \ldots$, which is never ending. See, for example, Figure 1.2 for a pictorial representation of what we are describing here, in words.

[2]One famous mathematician who was decidedly on the finitary side of things was the great analyst, Edward Nelson. See `https://mathweb.ucsd.edu/~sbuss/ResearchWeb/nelson/talk.pdf`.

1.1 Sigma notation

Another topic which is related to proof-by-induction is the summation notation. Suppose that we want to write a formula like

$$a_n \;=\; 1 + 2 + \cdots + n\,. \tag{1.11}$$

Sometimes the pattern which is implicit in the ellipses (dot-dot-dots) may be ambiguous to some readers. Therefore, to make things as explicit as possible we use the Sigma notation.

$$a_n \;=\; \sum_{i=1}^{n} i\,. \tag{1.12}$$

What we mean in general by

$$\sum_{i=a}^{b} f(i) \;=\; f(a) + f(a+1) + \cdots + f(i) + \cdots + f(b)\,, \tag{1.13}$$

assuming that a and b are numbers (usually integers, either positive or negative whole numbers or zero) such that $b - a$ is an integer. If $b < a$ then we interpret the summation above to be 0, by convention. Note that, if we define $F(0)$, $F(1)$, $F(2)$, *et cetera*, by the formula

$$F(n) \;=\; \sum_{i=1}^{n} f(i)\,, \tag{1.14}$$

then we have, by convention and induction, that

$$F(0) \;=\; 0\,, \quad \text{and for all } n = 0, 1, \dots, \quad F(n+1) \;=\; F(n) + f(n+1)\,. \tag{1.15}$$

In fact, the previous two displayed equations are logically equivalent. We define the summation notation using induction. Therefore the second formula is tautologically equivalent to the first formula. So, for example, from $a_n = \sum_{i=1}^{n} i$, we have

$$a_0 \;=\; 0 \;\text{ and, for all } n = 0, 1, \dots, \text{ we have } \; a_{n+1} \;=\; a_n + (n+1)\,. \tag{1.16}$$

1.2 A second example

Suppose that we decide to sum the a_n's that we defined from the last section. More specifically,let us define b_0, b_1, . . . such that

$$b_0 \;=\; 0\,, \tag{1.17}$$

and for each $n = 1, 2, \dots$, let us define

$$b_n \;=\; \sum_{i=1}^{n} a_i \;=\; 1 + \sum_{i=1}^{n} \frac{i(i+1)}{2}\,. \tag{1.18}$$

Then we might guess that we have a polynomial

$$b_n \;=\; \alpha + \beta n + \gamma n^2 + \kappa n^3\,. \tag{1.19}$$

Note that to have $b_0 = 0$ we must have $\alpha = 0$. So really,

$$b_n \;=\; \beta n + \gamma n^2 + \kappa n^3\,. \tag{1.20}$$

We must have $\beta + \gamma + \kappa = b_1 = a_1 = 1(1+1)/2 = 1$. We must also have $2\beta + 4\gamma + 8\kappa = b_2 = b_1 + a_2 = 1 + a_2 = 1 + 3 = 4$. Since $2\beta + 2\gamma + 2\kappa = 2$, that means

$$2\gamma + 6\kappa \;=\; 2\,, \quad \text{or} \quad \gamma + 3\kappa \;=\; 1\,, \tag{1.21}$$

which we obtain by subtracting (and then dividing). We must also have $3\beta+9\gamma+27\kappa = b_2+a_3 = 4+6 = 10$. Subtracting $3\beta + 3\gamma + 3\kappa$, we have

$$6\gamma + 24\kappa \;=\; 10 - 3 \;=\; 7\,. \tag{1.22}$$

Then subtracting $6\gamma + 18\kappa$, we obtain

$$6\kappa \;=\; 7 - 6 \;=\; 1\,. \tag{1.23}$$

So $\kappa = 1/6$. Then that means, since $\gamma = 1 - 3\kappa$, that

$$\gamma \;=\; 1 - \frac{1}{2} \;=\; \frac{1}{2} \;=\; \frac{3}{6}\,. \tag{1.24}$$

Then, since $\beta = 1 - \gamma - \kappa$ we get

$$\beta \;=\; 1 - \frac{1}{2} - \frac{1}{6} \;=\; \frac{6-3-1}{6} \;=\; \frac{2}{6} \;=\; \frac{1}{3}\,. \tag{1.25}$$

So we have the guess that

$$b_n \;=\; \sum_{i=1}^{n} \frac{i(i+1)}{2} \quad \text{is equal to} \quad \frac{2n + 3n^2 + n^3}{6}\,. \tag{1.26}$$

Note that this is the same as $n(n+1)(n+2)/6$.

Now let us prove that our guess is correct by using induction. For the initial step, we note that $b_0 = 0$ and that is equal to $(0)(0+1)(0+3)/6$ because that equals 0. Then for the induction step, if we assume that $b_n = (2n + 3n^2 + n^3)/6$, then we have for $m = n+1$,

$$b_m \;=\; b_{n+1} \;=\; b_n + a_{n+1} \;=\; b_n + \frac{(n+1)(n+2)}{2}\,, \tag{1.27}$$

which since we assumed $b_n = (2n + 3n^2 + n^3)/6$, we can rewrite as

$$\begin{aligned} b_{n+1} \;&=\; \frac{2n + 3n^2 + n^3}{6} + \frac{(n+1)(n+2)}{2} \\ &=\; \frac{n(n+1)(n+2)}{6} + \frac{3(n+1)(n+2)}{6} \\ &=\; \frac{(n+1)(n+2)(n+3)}{6}\,. \end{aligned} \tag{1.28}$$

But that is equal to $m(m+1)(m+2)/6$. So, we have proved the induction step,

Now, by proof-by-induction, we have proved our formula:

$$\sum_{i=1}^{n} \frac{i(i+1)}{2} \;=\; \frac{n(n+1)(n+2)}{6}\,. \tag{1.29}$$

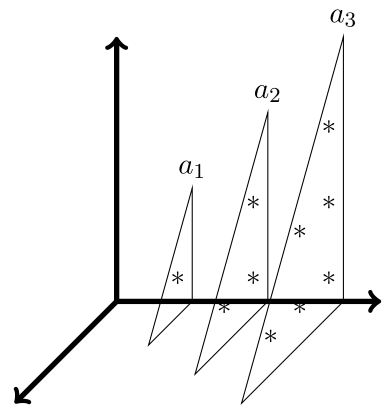


Figure 1.3: In the left-most plot we include the points in a_n's in two dimensions, and we go from term to term of sequence of a_n's in a third dimension. This is a pictorial perspective on the formula from this subsection. Note $a_1 = 1$, $a_2 = 3$, $a_3 = 6$. So $b_3 = 10$, as in the formula because $(3)(4)(5)/6 = 10$. This is a discrete version of a simplex.

There are many other examples that can be proved by similar techniques. One may note that the most important part of $(i^2 + i)/2$ for large i is just $i^2/2$. Similarly, the most important part of $n(n + 1)(n + 2)/6$, which is $(2n + 3n^2 + n^3)/6$ is just $n^3/6$ for large n. That suggests that we might have the approximation

$$\sum_{i=1}^{n} \frac{i^2}{2} \approx \frac{n^3}{6}\,. \tag{1.30}$$

More precisely we may guess that

$$\lim_{n\to\infty} \frac{1}{n^3} \sum_{i=1}^{n} i^2 = \frac{2}{6} = \frac{1}{3}\,. \tag{1.31}$$

In a sense, we all know this formula, because using Riemann sums this gives the formula $\int_0^1 t^2\, dt = (1^3 - 0^3)/3 = 1/3$, where for a fixed n, we have $t_n = i/n$ and $\Delta t = 1/n$. So $\lim_{n\to\infty} \sum_{i=1}^{n} t_i^2 \Delta t = \int_0^1 t^2\, dt = 1/3$. We can understand all of this better, later, after we learn a few more techniques in combinatorics. We will see that we are looking at one portion of a hypercube, where for d dimensions the number of symmetric regions, whose disjoint union comprises the total hypercube, is equal to the number of permutations of d objects. See Figure 1.3 for a picture corresponding to $\sum_{i=1}^{n} \frac{i(i+1)}{2} = \frac{n(n+1)(n+2)}{6}$. The general formula we will learn is called the Chu-Vandermonde identity. We will learn it later, in the next chapter.

1.A Optional appendix: Cardinality and rigorous proofs

The purpose of this class is not to introduce students to the challenge of writing proofs. There are other classes for that, such as Advanced Calculus. More importantly, we do not need to do that. Combinatorics and Graph Theory are important topics, at the same level of applicability as a Differential Equations class or a Probability Theory class. The first courses for those topics are taught without undue much emphasis on rigorous proofs. One may try the same for combinatorics and graph theory.

However, it is also true that some advanced mathematics students may appreciate rigorous proofs. When a student trains at a particular task full-time for years, they begin to narrow their focus to that type of task. So, especially for graduate students in mathematics, they will often not trust the presentation in a mathematics class unless most of the results can be presented in their language: the language of rigorous proofs. Therefore, we will present this using proofs at roughly the level of Advanced Calculus. An even better reference would be Set Theory. A good textbook for an introduction to set theory is by Herbert Enderton [3].

Notation: The symbol $\forall$ means "for all" or "for every." The symbol $\exists$ means "there exists." For two statements (or propositions) P and Q, the notation $P \Rightarrow Q$ means that "proposition P implies proposition Q."

We note that at a high level in computer science, some portions of set theory are actually useful there too. But that is primarily the finitary portion of set theory, since we know that computers cannot perform infinite tasks. Another name for an important type of set theory is called "intuitionist," or "constructive." This does get back to considerations such as those of Edward Nelson and other induction skeptics.

1.A.1 Bijective and surjective mappings

Suppose that we have two sets A and B. Suppose that we have a function,

$$f\,:\,A \to B\,. \tag{1.32}$$

Definition 1.A.1 *1. The function f is called* **surjective** *if and only if*

$$\forall y \in B\,, \quad \text{we have } \exists x \in A\,, \quad \text{such that } f(x) = y\,. \tag{1.33}$$

Another way to say the same thing is to say that f is **onto**.

2. The function f is called **injective** *if and only if*

$$\forall x \in A\,,\ \forall y \in A\,, \quad \text{we have the implication: } \Big(\big(f(x) = f(y)\big) \Rightarrow \big(x = y\big)\Big)\,. \tag{1.34}$$

Another way to say the same thing is to say that f is **one-to-one**.

3. The function f is called **bijective** *if and only if it has both properties: it is injective and it is surjective.*

Whether or not f is bijective we define the inverse function mapping $f^{-1} : \mathcal{P}(B) \to \mathcal{P}(A)$ where the power set of A is the collection of all subsets of A

$$\mathcal{P}(A) \;=\; \{X\,:\,X \subseteq A\}\,, \tag{1.35}$$

and the power set of B is the collection of all subsets of B

$$\mathcal{P}(B) \;=\; \{Y\,:\,Y \subseteq B\}\,. \tag{1.36}$$

The formula for the inverse function as a set function is

$$\forall Y \in \mathcal{P}(B)\,, \quad \text{we have} \quad f^{-1}(Y) \;=\; \{x \in A \,:\, f(x) \in Y\}\,. \tag{1.37}$$

If f is bijective, then for each $y \in B$ there is one and only one element $x \in A$ such that $f(x) = y$. In other words, we have that $f^{-1}(\{y\})$ is the set $\{x\}$. So, if f is bijective then we may define a new function $g : B \to A$ such that

$$\forall y \in B\,, \quad \text{we have } \; g(y) = x \text{ for that unique } x \text{ satisfying } f^{-1}(\{y\}) = \{x\}. \tag{1.38}$$

Definition 1.A.2 *If $f : A \to B$ is a bijection, then we denote the inverse function $f^{-1} : B \to A$, defined such that*

$$\forall x \in A\,,\ \forall y \in B\,,\ \Big(\big(f(x) = y\big) \Leftrightarrow \big(f^{-1}(y) = x\big)\Big)\,. \tag{1.39}$$

This is a slight abuse of notation, because f^{-1} also denotes the set-function inverse. But this usually does not cause much confusion.

Note that, if we take the composition of two functions both of which are bijections, then the resulting function is also a bijection. Also, for a bijective function, if we take the inverse function as defined just above, then the inverse function is also a bijection. We say that sets A and B are bijectively equivalent if there is a bijection from A onto B.

For us let $\mathbb{N} = \{1, 2, 3, \dots\}$. A main proof technique is proof-by-induction which we demonstrate in the next theorem.

Theorem 1.A.3 *Suppose that m and n are two elements of $\mathbb{N}$ such that $m < n$, and suppose that we have a function $f : \{1, \dots, m\} \to \{1, \dots, n\}$. Then f is not a surjection.*

As we have stated in the last section, proof-by-induction works as follows. Suppose that we have a set of the form $A = \{a, a+1, a+2, \dots\}$, and suppose that for each element of the set $x \in A$ we want to prove a particular statement $P(x)$. Then if we prove the first statement $P(a)$, for the case of the element a, and if we also prove the related result,

$$\forall x \in A\,, \quad \text{we have} \quad \big(P(x) \Rightarrow P(x+1)\big)\,, \tag{1.40}$$

then that proves the desideratum

$$\forall x \in A\,, \quad \text{we have} \quad P(x)\,. \tag{1.41}$$

Proving $P(a)$ is called the "initial step." Proving (1.40) is called the "induction step."

Before proving Theorem 1.A.3, let us introduce a notation that will get used and re-used. Given two elements $m, n \in \mathbb{N}$ if we have $r < n$ then we define a function

$$\gamma_{r,n} : \{1, \dots, n\} \setminus \{r\} \to \{1, \dots, n-1\}\,, \tag{1.42}$$

such that

$$\forall k \in \{1, \dots, n\} \setminus \{r\}, \quad \text{we have} \quad \gamma_{r,n}(k) \;=\; \begin{cases} k & \text{if } k < r, \\ k-1 & \text{if } k > r. \end{cases} \tag{1.43}$$

By treating the two cases, it is easy to see that $\gamma_{r,n}$ is bijective. We also define $\gamma_{n,n}$ as the identity mapping from $\{1, \dots, n-1\}$ back to itself. See Figure 1.4.

Notation: If we have a function $f : A \to B$, and if C is a subset of the domain $C \subset A$, then we define $f \restriction C$ to be the restriction of f to C, written $\big(f \restriction C\big) : C \to B$.

In case we need it, we also define another function

$$\widetilde{\gamma}_{m,n} : \{1, \dots, n\} \to \{1, \dots, n-1\}, \tag{1.44}$$

by the formula

$$\forall k \in \{1, \dots, n\}, \quad \text{we have} \quad \widetilde{\gamma}_{m,n}(k) \;=\; \begin{cases} \gamma_{m,n}(k) & \text{if } k \in \{1, \dots, n\} \setminus \{m\}, \\ 1 & \text{if } k = m. \end{cases} \tag{1.45}$$

Proof of Theorem 1.A.3:
The proof is by induction on $m \in \mathbb{N} = \{1, 2, \dots\}$.

INITIAL STEP: Suppose $m = 1$ and also assume $n > 1$. Note that since $n > 1$, this means $n \neq 1$. Now let $r = f(1)$. Note that $\{1, \dots, m\}$ is the singleton set $\{1\}$ since $m = 1$. So the entire range of f in this case is $\{f(1)\}$. In other words, the entire range of f is just $\{r\}$, as a subset of $\{1, \dots, n\}$. Then there are potentially three possibilities:

- $r = 1$,
- $1 < r < n$,
- $r = n$.

In the first case n is not in the range of f. In the second case neither 1 nor n is in the range of f. In the third case 1 is not in the range of f. In every case, $\operatorname{ran}(f)$ is not all of $\{1, \dots, n\}$. It either misses 1 or it misses n, or it misses both 1 and n. So f is not surjective. The initial step has been proved.

INDUCTION STEP: Now suppose that for some $k \in \mathbb{N}$, the conclusion of the theorem has been proved for $m = k$. That is called the "induction hypothesis." For the induction step, we instead take $m = k + 1$. Suppose that we have a mapping

$$f \;:\; \{1, \dots, k+1\} \to \{1, \dots, n\}, \tag{1.46}$$

for some number $n \in \mathbb{N}$ satisfying $n > k + 1$. By the induction hypothesis, we know that for any $n' > k$, and for any function

$$g \;:\; \{1, \dots, k\} \to \{1, \dots, n'\}, \tag{1.47}$$

that function g is not surjective onto $\{1, \dots, n'\}$. We are going to use our given function f to construct the new function g. Namely, first we take $n' = n - 1$. Note that we assumed $n > k+1$, so we conclude that $n - 1 > k$. So $n' > k$, as needed for applying the induction hypothesis to g. Next we need to actually construct the function g.

Let us take $r = f(k+1)$. Then we use the mapping $\widetilde{\gamma}_{r,n}$ defined before. Our function g is

$$g = \widetilde{\gamma}_{r,n} \circ (f \restriction \{1,\dots,k\}), \tag{1.48}$$

where $f \restriction \{1,\dots,k\}$ is the restriction of f to $\{1,\dots,k\}$ (and where $\circ$ denotes composition so that $g(i) = \widetilde{\gamma}_{r,n}(f(i))$ for those numbers $i \in \{1,\dots,k\}$). As mentioned before, g satisfies the hypotheses of the theorem, but with $m = k$. Therefore, by our induction hypothesis, we may assume that the conclusion of the theorem applies to g. So g is not a surjection.

Therefore, there is some $s \in \{1,\dots,n-1\}$ such that s is not in the range of g. Let us consider several cases.

- Suppose $s < r$. Then that means that $\widetilde{\gamma}_{n,r}^{-1}(\{s\})$ is a set which is either $\{s\}$ (if $s > 1$) or else $\{1, r\}$ if $s = 1$. In either case we know that $\{s\}$ is at least a subset of $\widetilde{\gamma}_{n,r}^{-1}(\{s\})$. We have assumed that $g^{-1}(\{s\}) = \emptyset$. But recall the "shoes and socks" principle: You put your socks on first and then your shoes second, but you take your shoes off first, then your socks off second. This is supposed to remind you that for any set A we have $g^{-1}(A) = f^{-1}(\widetilde{\gamma}_{n,r}^{-1}(A)) \cap \{1,\dots,k\}$. When you take the inverse of a composition you reverse the order of the functions in the composition defining the inverse: $(F \circ G)^{-1} = G^{-1} \circ F^{-1}$. (We have written $(f \restriction \{1,\dots,k\})^{-1}$ as $f^{-1}(\cdot) \cap \{1,\dots,k\}$, which is valid.) Therefore, if $g^{-1}(\{s\}) = \emptyset$, then that means there is no $i \in \{1,\dots,n\}$ such that $f(i) = s$. In this case f is not surjective, because s is not in the range of f. (Note that $f(k+1)$ is also not equal to s since $f(k+1) = r$ and $s < r$.)

- Now suppose that $s \geq r$, then that means $\widetilde{\gamma}_{n,r}^{-1}(\{s\}) = \{s+1\}$. So that means that there is no $i \in \{1,\dots,n\}$ such that $f(i) = s+1$. Note that since $s \geq r$ that means $s+1 > r$. So also $f(k+1) = r \neq s+1$. So f is not surjective because $s+1$ is not in the range of f.

Thus in both cases we see that either s or $s+1$ is not in the range of f. So f is not surjective.

So we have proved the induction step. If we assume that the theorem is true with $m = k$, then we can conclude it is also true for $m = k+1$. We also checked the initial step: the theorem is true for $m = 1$. Therefore, by proof-by-induction, that means the theorem is true for every choice of $m \in \mathbb{N}$. □

Note that results such as Theorem 1.A.3 are the type of theorems proved in Advanced Calculus, MA 440/441 or MA 540/541. But those classes are not necessarily prerequisites for MA 693, Combinatorics and Graph Theory. The only pre-requisites are Calculus I and II. ***The present section is optional. It is mainly included in case some mathematics graduate students want to see it.***

Definition 1.A.4 *If a set A has the property that there is a function*

$$f : \{1,\dots,n\} \to A, \tag{1.49}$$

for some $n \in \mathbb{N}$, and the function f is a bijection, then $|A| = n$. The **cardinality** *of A is n.*

Note that the cardinality formula is well-defined (there cannot be two or more different values for the same thing) because if $m \neq n$ then it is impossible for A to be bijectively equivalent to $\{1, \ldots, n\}$ and $\{1, \ldots, m\}$. That would violate Theorem 1.A.3 by transitivity and symmetry of bijective equivalence. That would imply $\{1, \ldots, m\}$ is bijectively equivalent to $\{1, \ldots, n\}$ and vice-versa, and since $m \neq n$ we must either have $m < n$ or $n < m$.

Theorem 1.A.5 *Suppose we have two sets A and B which are bijectively equivalent. If $|A| = n$ for some $n \in \mathbb{N}$, then we also have $|B| = n$.*

Proof: Suppose $f : A \to B$ is a bijection. By the definition above, since $|A| = n$ we know that there is another bijection $g : \{1, \ldots, n\} \to A$. Then we may take the composition $f \circ g : \{1, \ldots, n\} \to B$, which is also a bijection. Using Definition 1.A.4, this proves $|B| = n$. □

In Combinatorics in general one tries to prove identities such as

$$a_n = b_n \tag{1.50}$$

for two sequences $(a_n)_{n \in \mathbb{N}}$, $(b_n)_{n \in \mathbb{N}}$ which have been defined in some way that it is apparent that $a_n, b_n \in \mathbb{N}$ for all $n \in \mathbb{N}$. In Enumerative Combinatorics one tries to prove identities such as this for cardinalities of sets. If we are given sets A_n and B_n such that $|A_n| = a_n$ and $|B_n| = b_n$, then one approach is to seek a direct bijection from A_n onto B_n and then use Theorem 1.A.5.

But there are other techniques as well. One other technique is induction. Sometimes, induction can be used to construct direct bijections, too. Yet another technique involves generating functions. But again, one might check a generating function identity by using induction. But there are other times where generating functions yield a truly alternative approach to induction.

1.B Optional appendix: Proof of the pigeon-hole principle

Again, this entire section is purely optional. But it may help those who are trying to understand the previous proof to see pictures of the mappings defined.

Now we are going to prove one version of the pigeon hole principle in great detail. Let us first re-state Theorem 1.A.3 in a slightly different way:

Lemma 1.B.1 *For each m and n in $\mathbb{N}$ if $m < n$ then there does not exist any mapping*

$$f : \{1, \ldots, m\} \to \{1, \ldots, n\},$$

which is a surjection.

This is equivalent to Theorem 1.A.3. But let us review the argument.

Proof: It can be proved by induction on m. If $m = 1$ then $\mathrm{ran}(f) = \{f(1)\}$ consists of just 1 number. If $n \geq 2$ that means:

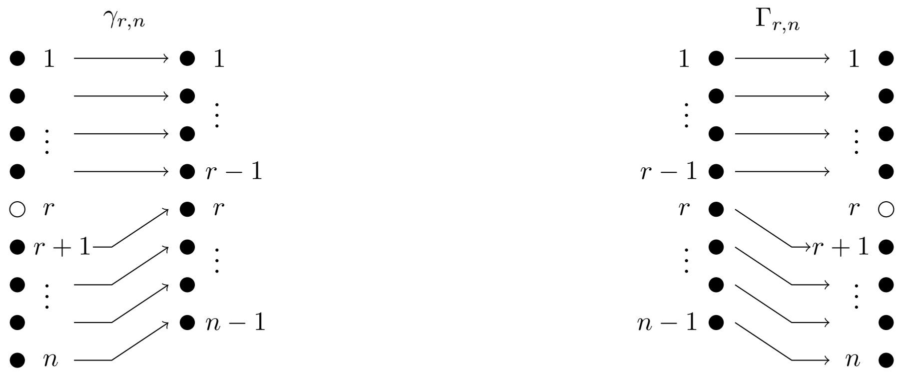


Figure 1.4: The mappings $\gamma_{r,n}$ and $\Gamma_{r,n}$ from the proof.

- if $f(1) = 1$ then $2 \notin \operatorname{ran}(f)$,
- if $f(1) = 2$ then $1 \notin \operatorname{ran}(f)$,
- if $f(1) \notin \{1, 2\}$ then neither 1 nor 2 is in $\operatorname{ran}(f)$.

In every case f is not a surjection since $\operatorname{ran}(f) \neq \{1, \ldots, n\}$. So the initial step of the proof-by-induction is proved, $m = 1$.

Now let us consider the induction step.

Suppose that k is some number in $\mathbb{N}$, and suppose that Theorem 1.3 has already been proved for $m = k$. That is the induction hypothesis. Now let us suppose that $m = k + 1$ and $n > m$. We will try to prove the result also for $m = k + 1$, using the induction hypothesis. That will prove the *induction step.*

Suppose we have a mapping

$$f \,:\, \{1, \ldots, k + 1\} \to \{1, \ldots, n\}\,.$$

Let $r \in \{1, \ldots, n\}$ be equal to $f(k + 1)$. Now we make a technical assumption.

Technical assumption: $\forall i \in \{1, \ldots, k\}$ we have $f(i) \neq r$.

We will explain how to remove the technical assumption at the end of the proof. Let us delay this step.

Let us use the notion of the restriction of a function to a subset of the domain. The restriction of f to $\{1, \ldots, k\}$ is defined simply as

$$\forall i \in \{1, \ldots, k\}\,, \quad \text{we have } \big(f \restriction \{1, \ldots, k\}\big)(i) \;=\; f(i)\,.$$

By the technical assumption we know that r is not in the range of the restriction

$$r \notin \operatorname{ran}\big(f \restriction \{1, \ldots, k\}\big)\,.$$

So we may view the restriction as a mapping into $\{1, \ldots, n\} \setminus \{r\}$,

$$\big(f \restriction \{1, \ldots, k\}\big) : \{1, \ldots, k\} \to \big(\{1, \ldots, n\} \setminus \{r\}\big)\,.$$

Now, there is a bijection

$$\gamma_{r,n} : \big(\{1, \ldots, n\} \setminus \{r\}\big) \to \{1, \ldots, n - 1\}\,.$$

See Figure 1.4. So defining the composition

$$g = \gamma_{r,n} \circ (f \restriction \{1, \ldots, k\}),$$

we have a mapping

$$g : \{1, \ldots, k\} \to \{1, \ldots, n-1\}.$$

But since $n > k + 1$ we know $n - 1 > k$.
So by the induction hypothesis, we know that g is not surjective.
But that implies f is not surjective.

- If f were surjective onto $\{1, \ldots, n\}$, then $\big(f \restriction \{1, \ldots, k\}\big)$ would be surjective onto $\{1, \ldots, n\} \setminus \{r\}$.
- But since $\gamma_{r,n} : \{1, \ldots, n\} \setminus \{r\}$ is surjective, that would imply the composition of the two functions would be surjective, since the composition of two surjective mappings is surjective.
- Since we know that g is not surjective, by proof-by-contradiction we conclude that f is not surjective.

Technical assumption bypass: Define a replacement function

$$\mathfrak{f} : \{1, \ldots, k+1\} \to \{1, \ldots, n\}.$$

If $r > 1$, let $s = 1$. If $r = 1$ then let $s = 2$.
For each $i \in f^{-1}(\{r\}) \setminus \{k+1\}$ let $\mathfrak{f}(i) = s$, otherwise $\mathfrak{f}(i) = f(i)$.
Since $\mathfrak{f}$ satisfies the *technical assumption*, it is not surjective.
Since $\operatorname{ran}(\mathfrak{f}) = \operatorname{ran}(f) \cup \{s\}$, f is not surjective. □

For a picture of the bijections used, see Figure 1.4. For $n > 2$ and $r < n$ we have

$$\gamma_{r,n} : \{1, \ldots, n\} \setminus \{r\} \to \{1, \ldots, n-1\},$$

$$\gamma_{r,n}(k) = \begin{cases} k & \text{if } k < r, \\ k-1 & \text{if } k > r. \end{cases}$$

Also define another mapping,

$$\Gamma_{r,n} : \{1, \ldots, n-1\} \to \{1, \ldots, n\} \setminus \{r\}$$

$$\Gamma_{r,n}(k) = \begin{cases} k & \text{if } k < r, \\ k+1 & \text{if } k \geq r. \end{cases}$$

If $r = n$, define $\gamma_{n,n} : \{1, \ldots, n-1\} \to \{1, \ldots, n-1\}$ and $\Gamma_{n,n} : \{1, \ldots, n-1\} \to \{1, \ldots, n-1\}$ to both be equal to the identity map.

These mappings are inverses. See this as follows. For $k < r$ we have $\gamma_{r,n}(k) = k$ and then $\Gamma_{r,n}(\gamma_{r,n}(k)) = k$. But for $k > r$ we have $\gamma_{r,n}(k) = k - 1$ and $k - 1 \geq r$. So $\Gamma_{r,n}(\gamma_{r,n}(k))$ equals $\Gamma_{r,n}(k-1)$ which is $k - 1 + 1 = k$. Therefore, for all $k \in \{1, \ldots, n\} \setminus \{r\}$, we have $\big(\Gamma_{r,n} \circ \gamma_{r,n}\big)(k) = k$.

Similarly, for all $k < r$ we have $\Gamma_{r,n}(k) = k$ and $\gamma_{r,n}(\Gamma_{r,n}(k)) = k$.

For $k \in \{r, \ldots, n-1\}$, we have $\Gamma_{r,n}(k) = k+1$ and $k+1 > r$. So $\gamma_{r,n}(\Gamma_{r,n}(k)) = \gamma_{r,n}(k+1) = k+1-1 = k$. Therefore, for all $k \in \{1, \ldots, n-1\}$ we have $\big(\gamma_{r,n} \circ \Gamma_{r,n}\big)(k) = k$.

In general, inverse pair mappings are bijections.

We proved that $\gamma_{r,n}(\Gamma_{r,n}(k)) = k$ and $\Gamma_{r,n}(\gamma_{r,n}(k)) = k$. Hence, $\Gamma_{r,n} = \gamma_{r,n}^{-1}$. For completeness we prove that invertible functions are bijections.

Suppose we have functions $\gamma : A \to B$ and $\Gamma : B \to A$ such that:

$$\forall x \in A\,,\ \Gamma(\gamma(x)) = x \quad \text{and} \quad \forall y \in B\,,\ \gamma(\Gamma(y)) = y\,.$$

Then we can prove that γ is bijective as follows.

- For $x_1, x_2 \in A$ if $\gamma(x_1) = \gamma(x_2)$ then $x_1 = \Gamma(\gamma(x_1)) = \Gamma(\gamma(x_2)) = x_2$.
- So γ is injective.
- For any $y \in B$, let $x = \Gamma(y)$.
- Then $\gamma(x) = \gamma(\Gamma(y)) = y$.
- So γ is surjective.

It is a special type of exercise to check a proof such as the previous one, step by step, starting from basic foundations. We do this type of exercise for some theorems at the beginning of the notes. But we will not do so later.

Another such theorem is the pigeon-hole principle. For any set A and any $n \in \mathbb{N}$, we say that

$$|A| = n\,,$$

if and only if there is a bijection $\varphi : \{1, \ldots, n\} \to A$.

The pigeon hole principle says the following.

Theorem 1.B.2 *Given a number $n \in \mathbb{N}$ and two sets A and B, assume that*

$$|A| = |B| = n\,.$$

Then for any function $f : A \to B$, we have the equivalence: f is injective if and only if f is surjective.

This is called the pigeon-hole principle, because if some pigeon hole has 2 or more "pigeons" in it, then that means that some box must have 0 "pigeons" in it, and vice-versa. Pictorially, that means: f is not 1-to-1 if and only if it is not onto. See Figure 1.5 for one example.

What follows was originally presented in slides. But here we include it in this section. ***Again, we mention that this whole section is optional. That is why we will not include slides for this topic in the present version of the course.***

We start by establishing the following, which is one half of a specific instance of the pigeon-hole principle.

Lemma 1.B.3 *For a number $n \in \mathbb{N}$ with $n > 1$ and a function*

$$f : \{1, \ldots, n\} \to \{1, \ldots, n\}\,,$$

suppose that f is not injective. More precisely assume that $f(n) = f(1)$. Then f is not surjective.

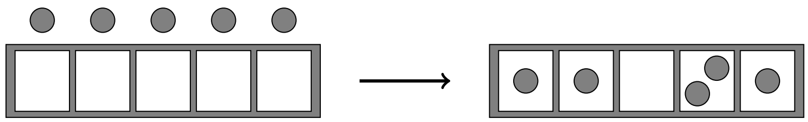


Figure 1.5: This is an illustration of one version of the pigeon-hole principle. If we have n balls to place in n boxes, then there is at least 1 empty box if and only if there is at least 1 box with more than one ball.

Proof: In order to prove this, let r denote the number $f(1)$. Note that we have assumed $f(n) = f(1)$. So we have $f(n) = f(1) = r$.

Technical assumption: $f^{-1}(\{r\}) = \{1, n\}$.

Because of the technical assumption, if we restrict,

$$\big(f \restriction \{2, \dots, n-1\}\big) : \{2, \dots, n-1\} \to \{1, \dots, n\},$$

we know that r is not in its range.

So we may apply the mapping

$$\gamma_{r,n} : \{1, \dots, n\} \setminus \{r\} \to \{1, \dots, n-1\}.$$

Actually, we define the function,

$$g : \{1, \dots, n-2\} \to \{1, \dots, n-1\},$$

by the formula

$$g(k) = \gamma_{r,n}(f(k+1)),$$

which is well-defined because for k in $\{1, \dots, n-2\}$ we have that $k+1$ is in $\{2, \dots, n-1\}$.

By the main theorem, we know that g is not surjective. So there is some $k \in \{1, \dots, n-1\}$ which is not in the range of g. Define $j = \gamma_{r,n}^{-1}(k)$, which is in $\operatorname{ran}\left(\gamma_{r,n}^{-1}\right)$ which is $\{1, \dots, n\} \setminus \{r\}$.

If $f(i) = j$ for some $i \in \{2, \dots, n-1\}$, then we have

$$g(i-1) = \gamma_{r,n}(f(i)) = \gamma_{r,n}(j) = \gamma_{r,n}\left(\gamma_{r,n}^{-1}(k)\right) = k.$$

But k is not in the range of g, so there is no i in $\{2, \dots, n-1\}$ such that $f(i) = j$.

Also $j \neq r$ so we do not have that $f(i) = j$ for $i = 1$ nor for $i = n$. Hence $j \notin \operatorname{ran}(f)$: f is not surjective.

Technical assumption bypass:

If $r = 1$, let $s = 2$. If $r > 1$, let $s = 1$.

Define $\mathfrak{f} : \{1, \dots, n\} \to \{1, \dots, n\}$ such that for $i \in f^{-1}(\{r\}) \setminus \{1, n\}$ we take $\mathfrak{f}(i) = s$, and otherwise $\mathfrak{f}(i) = f(i)$.

Since $\mathfrak{f}$ satisfies the technical assumption, $\mathfrak{f}$ is not surjective.

Since $\operatorname{ran}(\mathfrak{f}) = \operatorname{ran}(f) \cup \{s\}$, f is not surjective. □

Now let us prove a specific instance of the second half of the pigeon-hole principle.

Lemma 1.B.4 *For a number $n \in \mathbb{N}$ with $n > 1$ and a function*

$$f : \{1, \dots, n\} \to \{1, \dots, n\},$$

suppose that f is not surjective. More precisely assume that n is not in $\mathrm{ran}(f)$*. Then f is not injective.*

Proof: In order to reach a contradiction, suppose that f is injective. Let $A = \mathrm{ran}(f)$. So $A \subseteq \{1, \dots, n-1\}$ since we assumed $n \notin \mathrm{ran}(f)$. For each $j \in \mathrm{ran}(f)$, there is at least one i in $f^{-1}(\{j\})$. Since f is injective, there is at most one i in $f^{-1}(\{j\})$. Define a function $g : A \to \{1, \dots, n\}$, such that for each $j \in A$ we let $i \in \{1, \dots, n\}$ be that number such that $f(i) = j$: $g(j) = i$.

For every $i \in \{1, \dots, n\}$ we know that $i \in f^{-1}(\{j\})$ for $j = f(i)$. So $g(j)$ equals i. So every $i \in \{1, \dots, n\}$ is in $\mathrm{ran}(g)$. So g is surjective. But $A \subseteq \{1, \dots, n-1\}$, so this will be seen to violate the main theorem.

Technically, define $\mathfrak{g} : \{1, \dots, n-1\} \to \{1, \dots, n\}$ such that for $i \in \{1, \dots, n-1\} \setminus A$ we have $\mathfrak{g}(i) = 1$. Otherwise $\mathfrak{g}(i) = g(i)$.

Then $\mathrm{ran}(\mathfrak{g}) \supseteq \mathrm{ran}(g)$. But $\mathrm{ran}(g) = \{1, \dots, n\}$. So $\mathfrak{g}$ is surjective, violating the main theorem. This is a contradiction. So, our assumption must be false: f is not injective. □

Now let us generalize the second half of the pigeon-hole principle.

Corollary 1.B.5 *For a number $n \in \mathbb{N}$ with $n > 1$ and a function*

$$f : \{1, \dots, n\} \to \{1, \dots, n\},$$

suppose that f is not surjective. Then f is not injective.

Proof: By hypothesis, there is a number $r \in \{1, \dots, n\}$ not in $\mathrm{ran}(f)$. Then we can apply the bijection $\gamma_{r,n} : \{1, \dots, n\} \setminus \{r\} \to \{1, \dots, n-1\}$. Define $g : \{1, \dots, n\} \to \{1, \dots, n\}$ by taking $g = \gamma_{r,n} \circ f$. Now n is the number not in $\mathrm{ran}(g)$. So, by Lemma 1.B.4, g is not injective. So one of f and $\gamma_{r,n}$ is not injective, but $\gamma_{r,n}$ is. □

Now let us state the second half of the pigeon-hole principle for two general sets with the same cardinality.

Proposition 1.B.6 *Suppose that $|A| = |B| = n$ for some $n \in \mathbb{N}$ with $n > 1$ and*

$$f : A \to B,$$

is a function which is not surjective. Then f is not injective.

Proof: Since $|A| = n$, there is a bijection

$$\varphi : \{1, \dots, n\} \to A.$$

Since $|B| = n$, there is a bijection

$$\psi : \{1, \dots, n\} \to B.$$

Let us define

$$g : \{1, \dots, n\} \to \{1, \dots, n\},$$

by $g = \psi^{-1} \circ f \circ \varphi$.

Because φ and ψ are bijections, we know that g is not surjective because f is not surjective. To see this, for example, note that

$$f \;=\; \psi \circ g \circ \varphi^{-1} \,,$$

and ψ and φ^{-1} are also bijections.

- If g were a surjection, then $\psi \circ g \circ \varphi^{-1}$ would be a composition of surjective functions.
- So $\psi \circ g \circ \varphi^{-1} = f$ would be a surjection.
- But f is not a surjection.

So by Corollary 1.B.5 that means g is not an injection. But then by a similar argument that means that f is not an injection (otherwise $\psi^{-1} \circ f \circ \varphi$ would be a composition of injective functions, hence injective: but g is not injective). □

Now let us extend the first half result above.

Corollary 1.B.7 *For a number $n \in \mathbb{N}$ with $n > 1$ and a function*

$$f : \{1, \ldots, n\} \to \{1, \ldots, n\} \,,$$

suppose that f is not injective. Then f is not surjective.

Proof: Now we do not assume that $f(1) = f(n)$. Instead we assume

$$f(s) \;=\; f(t) \,,$$

for some numbers s and t with

$$1 \;\le\; s \;<\; t \;\le\; n \,.$$

So we need a bijection between $\{2, \ldots, n-1\}$ and $\{1, \ldots, n\} \setminus \{s, t\}$.

We first make a bijection,

$$\Gamma^{(2)}_{s,t,n} \;:\; \{1, \ldots, n-2\} \to \{1, \ldots, n\} \setminus \{s, t\} \,.$$

Just take the composition: $\Gamma^{(2)}_{r,s,n} = \Gamma_{s,n} \circ \Gamma_{r,n-1}$. See Figure 1.6.

Next, take

$$\gamma^{(2)}_{s,t,n} \;:\; \{1, \ldots, n\} \setminus \{s, t\} \to \{1, \ldots, n-2\} \,.$$

by the formula

$$\gamma^{(2)}_{s,t,n} \;=\; \left(\Gamma^{(2)}_{s,t,n}\right)^{-1} \,.$$

Let r equal the value of $f(s)$ and $f(t)$.

Define the mapping

$$g \;:\; \{1, \ldots, n\} \to \{1, \ldots, n\}$$

by

$$g(k) \;=\; \begin{cases} f\left(\Gamma^{(2)}_{s,t,n}(k-1)\right) & \text{for } k \in \{2, \ldots, n-1\}, \\ r & \text{for } k \in \{1, n\}. \end{cases}$$

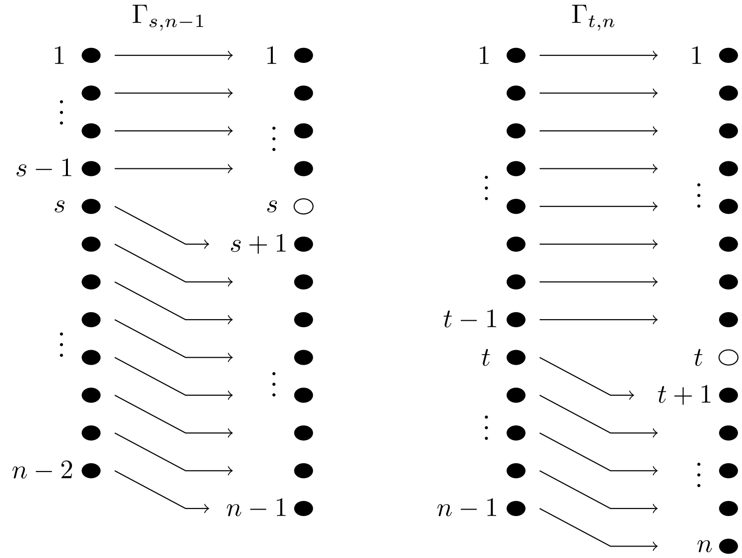


Figure 1.6: This shows how to compose two instances of $\Gamma_{r,m}$ for choices of r and m, to get $\Gamma^{(2)}_{s,t,n}$. Note that, as is usual, when giving pictures for the composition, we draw from left-to-right. But when writing compositions, following the standard mathematical notation that $\big(f \circ g\big)(x) = f(g(x))$, we write compositions from right-to-left.

So g satisfies the hypotheses of Lemma 1.B.3, so it is not surjective.

But then we also have the formula

$$f(k) \;=\; \begin{cases} g\left(\gamma^{(2)}_{s,t,n}(k)+1\right) & \text{for } k \in \{1,\ldots,n\} \setminus \{s,t\}, \\ r & \text{for } k \in \{s,t\}. \end{cases}$$

Note that the mapping

$$k \mapsto \gamma^{(2)}_{s,t,n}(k) + 1 \quad \text{for} \quad k \in \{1,\ldots,n\} \setminus \{s,t\}$$

defines a bijection from $\{1,\ldots,n\} \setminus \{s,t\}$ onto $\{2,\ldots,n-1\}$ whose inverse is the mapping

$$k \mapsto \Gamma^{(2)}_{s,t,n}(k-1)\,.$$

So $\operatorname{ran}(f) = \operatorname{ran}(g)$.

So, since g is not surjective, $\operatorname{ran}(g)$ is not $\{1,\ldots,n\}$. So $\operatorname{ran}(f)$ is not $\{1,\ldots,n\}$, so f is not surjective. $\square$

Now we prove the second half of the pigeon-hole principle for general sets that have the same cardinality.

Proposition 1.B.8 *Suppose that $|A| = |B| = n$ for some $n \in \mathbb{N}$ with $n > 1$ and*

$$f \,:\, A \to B\,,$$

is a function which is not injective.

Proof: Since f is not injective, there are some numbers x_1 and x_2 in A, with $x_1 \neq x_2$ and $f(x_1) = f(x_2)$. Since $|A| = n$, there is a bijection

$$\varphi : \{1, \ldots, n\} \to A\,.$$

Since $|B| = n$, there is a bijection

$$\psi : \{1, \ldots, n\} \to B\,.$$

Let s and t denote the two points $\varphi^{-1}(x_1)$ and $\varphi^{-1}(x_2)$ ordered so

$$s < t\,.$$

Let

$$g : \{1, \ldots, n\} \to \{1, \ldots, n\}\,,$$

be defined by $g = \psi^{-1} \circ f \circ \varphi$.

Then g satisfies the hypotheses of Corollary 1.B.7. So g is not surjective.

But φ and ψ are bijections. So if g is not surjective and

$$g = \psi^{-1} \circ f \circ \varphi\,,$$

then that means f is not surjective. □

By combining Propositions 1.B.8 and 1.B.6, we obtain Theorem 1.B.2.

Chapter 2

Examples from Combinatorics

1 A first combinatorial identity

Theorem 1.1 (Fundamental Principle of Combinatorics)
Suppose that m and n are numbers in $\mathbb{N}$ and suppose that A and B are two sets such that

$$|A| = m \quad \text{and} \quad |B| = n\,. \tag{2.1}$$

Define the Cartesian product to be the set of ordered pairs

$$A \times B = \{(x, y) : x \in A \quad \text{and} \quad y \in B\}\,. \tag{2.2}$$

Then $|A \times B| = mn$.

Proof: Consider the $n \times m$ matrix M defined as

$$M = \begin{bmatrix} 1 & 2 & 3 & \dots & m \\ m+1 & m+2 & m+3 & \cdots & 2m \\ 2m+1 & 2m+2 & 2m+3 & \cdots & 3m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (n-1)m+2 & (n-1)m+2 & (n-1)m+3 & \cdots & nm \end{bmatrix}. \tag{2.3}$$

It is easy to see that $h : \{1, \dots, m\} \to \{1, \dots, n\} \to \{1, \dots, mn\}$, defined by $h\big((i, j)\big) = a_{j,i}$, is a bijection. Then, given two bijections $f : \{1, \dots, m\} \to A$ and $g : \{1, \dots, n\} \to B$, we can define a new bijection from $\{1, \dots, mn\}$ to $A \times B$, where for $k \in \{1, \dots, mn\}$ we let (i, j) be the unique preimage under h and then we map that to $(f(i), g(j))$. $\square$

Theorem 1.2 (A first combinatorics result) *Consider for each $n \in \mathbb{N}$ the set A_n, a subset of the Cartesian product $\{1,\dots,n\} \times \{1,\dots,n\}$ defined by*

$$A_n = \big\{(i,j) \,:\, i,j \in \{1,\dots,n\} \;\; \text{and} \;\; j \le i\big\}\,. \tag{2.4}$$

Then $|A_n| = n(n+1)/2$. In other words, since we may write

$$A_n = \bigcup_{i=1}^{n} \big\{(i,j) \,:\, j \in \{1,\dots,i\}\big\}\,, \tag{2.5}$$

we may conclude that

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}\,. \tag{2.6}$$

Before proving the theorem, let us note another important fact about cardinality. Suppose that we have two sets E and F. Then

$$\text{if } |E| = m \text{ and } |F| = n \text{ for } m,n \in \mathbb{N}\,, \text{ and if } E \cap F = \emptyset\,, \text{ then } |E \cup F| = m+n\,. \tag{2.7}$$

Indeed, if $f : \{1,\dots,m\} \to E$ and $g : \{1,\dots,n\} \to F$ are two bijections, then we can define a new bijection $h : \{1,\dots,m+n\} \to E \cup F$ by using the formula

$$h(k) = \begin{cases} f(k) & \text{if } 1 \le k \le m, \text{ and} \\ g(k-n) & \text{if } m+1 \le k \le m+n. \end{cases} \tag{2.8}$$

This mapping is always onto, whether or not E and F are disjoint. But if $E \cap F = \emptyset$ (meaning the two sets are disjoint) then h is also 1-to-1.

Lemma 1.3 (Preliminary version of P.I.E.) *For any two finite sets E and F, whether or not they are disjoint, the cardinality of the union is*

$$|E \cup F| = |E| + |F| - |E \cap F|\,. \tag{2.9}$$

This is a preliminary version of the principle of inclusion exclusion. The full principle of inclusion exclusion extends this to more than just two sets.

Proof of Lemma 1.3: Define

$$G = E \cap F \text{, and then } E' = E \setminus G \text{ and } F' = F \setminus G. \tag{2.10}$$

Recall that the definition of the set-minus is $E \setminus G = \{x \in E : x \notin G\}$ and $F \setminus G = \{x \in F : x \notin G\}$. Then E' and F' and G are pairwise disjoint. So, using equation (2.7) two times,

$$|E' \cup F' \cup G| \,=\, |E'| + |F'| + |G|\,. \tag{2.11}$$

But also, it is easy to see that $|E' \cup G| = |E'| + |G|$ and $|F' \cup G| = |F'| + |G|$. Since these are also equal to $|E|$ and $|F|$, respectively, we conclude that

$$|E'| = |E| - |G| \text{ and } |F'| = |F| - |G|. \tag{2.12}$$

So finally, since $E' \cup F' \cup G = E \cup F$, we get the cardinality

$$|E \cup F| \,=\, (|E| - |G|) + (|F| - |G|) + |G| \,=\, |E| + |F| - |G|\,. \tag{2.13}$$

But that is $|E| + |F| - |E \cap F|$. So we have proved (2.9). $\square$

The proof of Lemma 1.3 is an example of "double-counting." In double-counting, we use Theorem 1.A.5 to count the cardinality of a set in two different ways. Then, because both ways count the same thing, the two numbers must be equal. The website

`https://en.wikipedia.org/wiki/Double_counting_(proof_technique)`

gives a good description. But we will also use it to prove Theorem 1.2. For finite sets $E_1, \dots, E_r$,

$$\begin{aligned} |E_1 \cup \dots \cup E_r| \,=\, &\sum_{i=1}^{r} |E_i| - \sum_{1 \le i_1 < i_2 \le r} |E_{i_1} \cap E_{i_2}| + \dots \\ &+ (-1)^k \sum_{1 \le i_1 < \dots < i_k \le r} |E_{i_1} \cap \dots \cap E_{i_k}| + \dots + (-1)^r |E_1 \cap \dots \cap E_r|\,. \end{aligned} \tag{2.14}$$

That is the principle of inclusion-exclusion. Lemma 1.3 is a special case, for $r = 2$.

Proof of Theorem 1.2 (double-counting method): Let us also define a set B_n, also a subset of the Cartesian product $\{1, \dots, n\} \times \{1, \dots, n\}$, by the formula

$$B_n \,=\, \big\{(i,j) \,:\, i,j \in \{1, \dots, n\} \;\text{ and }\; i \le j\big\}\,. \tag{2.15}$$

Let us define the interchange mapping on the Cartesian product

$$\varphi \,:\, \big(\{1, \dots, n\} \times \{1, \dots, n\}\big) \to \big(\{1, \dots, n\} \times \{1, \dots, n\}\big) \tag{2.16}$$

defined such that

$$\forall (i,j) \in \{1, \dots, n\} \times \{1, \dots, n\}\,, \quad \text{we have } \varphi\big((i,j)\big) \,=\, (j,i)\,. \tag{2.17}$$

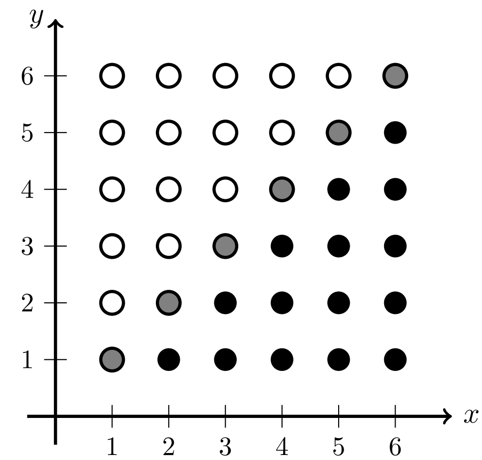


Figure 2.1: We have given a pictorial representation of the three sets A'_n, B'_n and D_n for $n = 6$. Points in A'_n are represented by black-filled dots. Points in D_n are represented by gray filled dots. Points in B'_n are represented by holes.

This is an example of an *involution.* Namely, φ is its own inverse, since $\varphi\big(\varphi\big((i,j)\big)\big)$ equals (i,j), again. So, obviously, φ is a bijection (since it has an inverse, itself).

But also $\varphi \restriction A_n$ maps A_n onto B_n. So $|A_n| = |B_n|$ by Theorem 1.A.5. Therefore, by (2.9), we have

$$|A_n \cup B_n| \;=\; |A_n| + |B_n| - |A_n \cap B_n| \;=\; 2|A_n| - |A_n \cap B_n|\,. \tag{2.18}$$

But since either $i \leq j$ or $j \leq i$, or both, we have $A_n \cup B_n$ is the entire Cartesian product $\{1,\ldots,n\} \times \{1,\ldots,n\}$. So by Theorem 1.1

$$n^2 \;=\; |\{1,\ldots,n\} \times \{1,\ldots,n\}| \;=\; |A_n \cup B_n| \;=\; 2|A_n| - |A_n \cap B_n|\,. \tag{2.19}$$

So $2|A_n| = n^2 - |A_n \cap B_n|$. But (i,j) is in $A_n \cap B_n$ if and only if we have both: $i \leq j$ and $j \leq i$. In other words,

$$A_n \cap B_n \;=\; D_n\,, \quad \text{where} \quad D_n \;=\; \big\{(i,i) \;:\; i \in \{1,\ldots,n\}\big\}\,. \tag{2.20}$$

But it is easy to see from Definition 1.A.4 that the cardinality of D_n is n, by just mapping $i \in \{1,\ldots,n\}$ onto its duplication-double $(i,i) \in D_n$, which defines a bijection. So $|A_n| = \frac{1}{2}(n^2 - |D_n|)$ which is $n(n-1)/2$. $\square$

We may also define $A'_n = A_n \setminus D_n$ and $B'_n = B_n \setminus D_n$ so that the Cartesian product may be written

$$\{1,\ldots,n\} \times \{1,\ldots,n\} \;=\; A'_n \cup B'_n \cup D_n\,, \tag{2.21}$$

which is a *disjoint union*, since A'_n, B'_n and D_n are pairwise disjoint. An example for $n = 6$ is shown in Figure 2.1. We could also find an alternative proof.

Proof of equation (2.6) by induction:
Initial step:
If $n = 1$ then $\sum_{i=1}^{n} i$ is $\sum_{i=1}^{1} i$ which is just 1. It is a sum with only a single summand, and that summand is i which is the index, which is 1. But, if $n = 1$, then $n(n+1)/2$ is given by $1(2)/2$, which is also 1. So the identity is true for $n = 1$.

Induction step:
Now suppose that $k \in \mathbb{N}$, and also assume that equation (2.6) has been proved for $n = k$. Take $n = k + 1$, instead. Then we have

$$\sum_{i=1}^{n} i \,=\, \sum_{i=1}^{k+1} i \,=\, \left(\sum_{i=1}^{k} i\right) + (k+1)\,, \tag{2.22}$$

by the inductive definition of the summation (Sigma notation). But we already know that the first sum $\sum_{i=1}^{k} i$ equals $k(k+1)/2$ by our assumption that (2.6) is true for $n = k$. Therefore, we may rewrite

$$\left(\sum_{i=1}^{k} i\right) + (k+1) \,=\, \frac{k(k+1)}{2} + (k+1) \,=\, \frac{k(k+1)}{2} + \frac{2(k+1)}{2} \,=\, \frac{(k+2)(k+1)}{2}\,. \tag{2.23}$$

But since $n = k+1$ that is equal to $(n+1)n/2$ which is the same as $n(n+1)/2$. So (2.6) is also true for $n = k+1$. So the induction step has been proved. So the result holds by induction. $\square$

2 A first result about S_n, the set of permutations

Let us mention the *recursion principle*. It is a basic theorem of set theory, but it relies on other axioms to prove, such as the axiom of induction. See, for example, Enderton [3]. Suppose we are given a set $\mathcal{X}$, and an element $\mathfrak{x} \in \mathcal{X}$. Suppose we are also given a function $f : \mathcal{X} \to \mathcal{X}$ which does not necessarily need to be one-to-one nor onto. Then the recursion principle says that there is a mapping

$$\varphi \,:\, \mathbb{N} \to \mathcal{X}\,, \tag{2.24}$$

satisfying the conditions: $\varphi(1) = \mathfrak{x}$, and for each $n \in \mathbb{N}$ we have $\varphi(n+1) = f(\varphi(n))$. Moreover, the mapping φ is uniquely determined by these conditions.

As an example, consider the factorial function. As usual, we define the factorial such that $0! = 1$ and such that, for each $n \in \mathbb{N}$ we have

$$n! \,=\, n \cdot (n-1)!\,. \tag{2.25}$$

In other words, we could define $\mathcal{X} = \mathbb{R} \times \mathbb{N}$ and then define $f : \mathcal{X} \to \mathcal{X}$ by the formula $f\big((x,n)\big) = (n \cdot x, n+1)$. We choose the element $\mathfrak{x} \in \mathcal{X}$ to be $(1,1)$. Then we would be able to obtain the factorial function from the recursion formula, thus:

$$\varphi(1) \,=\, (1,1) \,=\, (0!,1)\,, \quad \text{and} \quad \varphi(n+1) \,=\, f\big(\varphi(n)\big) \,=\, (n!, n+1)\,. \tag{2.26}$$

This works since, if $\varphi(n) = ((n-1)!, n)$, then $\varphi(n+1)$ equals $f\Big(\big((n-1)!, n\big)\Big) = \big(n\cdot(n-1)!, n+1\big)$. In other words, what we are saying is that the factorial function can be defined this way, using the recursion principle to give us the definition.That would be pedagogical in the context of set-theory.

We usually will not go to the trouble of checking how to formally apply the recursion principle, even when that is what we should do to be complete. We just say that a sequence is defined inductively or recursively and give enough information that the reader or student could imagine the rest of the details in their head. The set of permutations is also defined by reference to recursion or induction.

Given a set A we denote the Cartesian product $A \times A$ as just A^2. It is the set of all ordered pairs (x, y) with $x, y \in A$. Also, for $n \in \mathbb{N}$, if we have already defined A^n, then we will let A^{n+1} be defined to be the Cartesian product $A \times A^n$. But we will eliminate unnecessary extra appearances of parentheses. So A^n is just the set of all ordered n-tuples $(x_1, \dots, x_n)$ such that $x_1, \dots, x_n \in A$. (As an example, from our definition, technically we would have defined A^3 to be the set of all $\big(x_1, (x_2, x_3)\big)$ for $x_1, x_2, x_3 \in A$. But by general consensus we all write it as (x_1, x_2, x_3) instead.)

Lemma 2.1 *Define S_n to be the set of all permutations:*

$$S_n = \big\{(\pi_1, \dots, \pi_n) \in \{1, \dots, n\}^n : \{\pi_1, \dots, \pi_n\} = \{1, \dots, n\}\big\}. \tag{2.27}$$

Then $|S_n| = n!$. We denote elements of S_n as $\pi = (\pi_1, \dots, \pi_n)$, for example.

Before proving this lemma let us pause to state one version of the **pigeon-hole principle.** We are not going to prove it, here, although it could be proved by induction, using methods similar to the proof of Theorem 1.A.3. (And it is usually proved in MA 440/441 and MA 540/541.) To any student who needs to know the proof, please look it up on Wikipedia

`https://en.wikipedia.org/wiki/Pigeonhole_principle`

under the sub-heading, "alternative formulations." Here the version of the pigeon-hole principle that we like is this: if we have a function $f : A \to B$ and if $|A| = |B| = n$ for some $n \in \mathbb{N}$, then

$$\Big(f \text{ is injective}\Big) \Leftrightarrow \Big(f \text{ is surjective}\Big). \tag{2.28}$$

It may seem a bit inconsistent for us to skip the proof of this result, even though we did prove Theorem 1.A.3, which is even more basic. But we cannot prove everything. (The purpose of this course is not just to review basic set theory results.)

Proof of Lemma 2.1: Suppose that $n > 1$. Let $\pi_n = r$. Note that, defining a mapping

$$f : \{1, \ldots, n\} \to \{1, \ldots, n\}\,, \tag{2.29}$$

by the formula

$$\forall k \in \{1, \ldots, n\}\,, \quad \text{we have} \quad f(k) \;=\; \pi_k\,, \tag{2.30}$$

the function f is a surjection because $\mathrm{ran}(f) = \{1, \ldots, n\}$ according to the definition of S_n in equation (2.27). Therefore f is an injection because, according to the pigeon-hole principle (2.28), since it is a surjection it must also be an injection. So for every $i < n$ we have $f(i) \neq r$.

Considering $f \restriction \{1, \ldots, n-1\}$, the restriction of f to $\{1, \ldots, n-1\}$, we have

$$\mathrm{ran}(f \restriction \{1, \ldots, n-1\}) \;=\; \{1, \ldots, n\} \setminus \{r\}\,. \tag{2.31}$$

Therefore, defining a new mapping,

$$g : \{1, \ldots, n-1\} \to \{1, \ldots, n-1\}\,, \quad \text{such that} \quad g \;=\; \gamma_{n,r}^{-1} \circ (f \restriction \{1, \ldots, n-1\})\,, \tag{2.32}$$

this mapping is a surjection, hence also an injection. Again we have used the pigeon hole principle (2.28). So g is a bijection. So we can define a new element $\rho \in S_{n-1}$ by the formula

$$\rho \;=\; (\rho_1, \ldots, \rho_{n-1})\,, \quad \text{such that} \quad \forall k \in \{1, \ldots, n-1\}\,, \quad \text{we have} \quad \rho_k \;=\; g(k)\,. \tag{2.33}$$

Let us denote the mapping taking $\pi \in S_n$ to the pair (r, ρ) in $\{1, \ldots, n\} \times S_{n-1}$ by

$$\mathcal{T}_n : S_n \to \{1, \ldots, n\} \times S_{n-1}\,. \tag{2.34}$$

Given any pair

$$(r, \rho) \in \{1, \ldots, n\} \times S_{n-1}\,, \tag{2.35}$$

we may first define a function

$$h : \{1, \ldots, n{-}1\} \to \{1, \ldots, n{-}1\}\,, \quad \text{such that} \quad \forall k \in \{1, \ldots, n{-}1\}\,, \quad \text{we have} \quad h(k) \;=\; \rho_k\,. \tag{2.36}$$

Then we may define an element $\sigma \in S_n$ according to the formula

$$\forall k \in \{1, \ldots, n\}\,, \quad \text{we have} \quad \sigma_k \;=\; \begin{cases} \gamma_{n,r}(h(k)) & \text{if } k \in \{1, \ldots, n-1\}, \\ r & \text{if } k = n. \end{cases} \tag{2.37}$$

Let us denote this mapping, taking the pair $(r, \rho) \in \{1, \ldots, n\} \times S_{n-1}$ to $\sigma \in S_n$, by

$$\widetilde{\mathcal{T}}_n : \{1, \ldots, n\} \times S_{n-1} \to S_n\,. \tag{2.38}$$

It is an exercise to see that $\mathcal{T}_n$ and $\widetilde{\mathcal{T}}_n$ are inverses, and we will skip this exercise. (The student may do it.) Therefore S_n and $\{1, \ldots, n\} \times S_{n-1}$ are bijectively equivalent.

Then we can prove the lemma inductively. It is easy to see that $S_n = \{(1)\}$ has cardinality 1. Then by our result, if we know $|S_{n-1}| = (n{-}1)!$, then by Theorems 1.A.5 and 1.1, $|S_n| = n|S_{n-1}|$, since S_n is bijectively equivalent to $\{1, \ldots, n\} \times S_{n-1}$. By the induction hypothesis that means $|S_n| = n\,(n-1)!$. Finally by the inductive definition of the factorial, this equals $n!$. $\square$

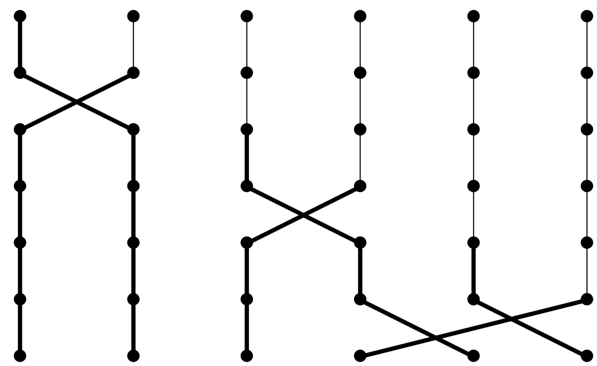


Figure 2.2: An example of the Fisher-Yates-Knuth algorithm. Take an example of $r = (1, 1, 3, 3, 5, 4)$ in $\{1\} \times \{1, 2\} \times \cdots \times \{1, \ldots, 6\}$. Then we get $\pi = (2, 1, 5, 3, 6, 4) \in S_6$. In this example, one can check the formula $\mathrm{Inv}(\pi) = \sum_{k=1}^{n} (k - r_k)$, because $0+1+0+1+0+2 = 4$ and that is the number of crossings in the picture due to the inversion-pairs $\{\{1, 2\}, \{3, 4\}, \{3, 6\}, \{5, 6\}\}$.

Using the proof, if we have

$$r = (r_1, r_2 \ldots, r_n) \in \{1\} \times \{1, 2\} \times \cdots \times \{1, \ldots, n\}, \tag{2.39}$$

then we obtain $\pi \in S_n$ by: first taking $\pi^{(1)} \in S_1$ to be (r_1) (which is necessarily (1)); then taking $\pi^{(2)} \in S_2$ to be $\widetilde{\mathcal{T}}_2(r_2, \pi^{(1)})$; and proceeding inductively until we get $\pi^{(n)} \in S_n$ to be $\widetilde{\mathcal{T}}_n(r_n, \pi^{(n-1)})$. We take $\pi = \pi^{(n)}$. This is called the Fisher-Yates(-Knuth) bijection on Wikipedia

`https://en.wikipedia.org/wiki/Fisher-Yates_shuffle`

An example is shown in Figure 2.2. This will be useful to us in a later, lecture. It has a good property that it demonstrates a construction directly related to the inversion-number of a permutation π:

$$\mathrm{Inv}(\pi) = |\{(i, j) : 1 \leq i < j \leq n \text{ and } \pi_i > \pi_j\}|. \tag{2.40}$$

3 Some first facts about binomial coefficients

For the moment, let us define binomial coefficients as follows. Firstly, we define

$$\forall n \in \{0, 1, 2, \ldots\},\ \forall k \in \{0, \ldots, n\}, \quad \text{we have } \binom{n}{k} = \frac{n!}{k!\,(n-k)!}. \tag{2.41}$$

Given $n \in \{0, 1, 2, \ldots\}$, let us extend the definition of $\binom{n}{k}$ to all $k \in \mathbb{Z}$ by setting $\binom{n}{k} = 0$ if $k < 0$ or $k > n$. Then the most fundamental fact for the binomial coefficients is Pascal's relation, as follows.

Theorem 3.1 *For any $n \in \mathbb{N}$ and for any $k \in \mathbb{Z}$, we have*

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}. \tag{2.42}$$

We have not yet given the alternative definition of the binomial coefficients (which we will do in the next section). So we will not state the elementary bijective proof. Instead we do a calculation.

Proof: If $k < 0$ then obviously (2.42) becomes the identity $0 = 0 + 0$. If $k > n$ then it is also that identity because then $k - 1 > n - 1$. Otherwise if $k = 0$ then the identity becomes $1 = 1 + 0$ because $\binom{n}{0} = \binom{n-1}{0} = 1$ and $\binom{n-1}{-1} = 0$. If $k = n$ then (2.42) is still true using the equation

$$\binom{n}{k} = \binom{n}{n-k}. \tag{2.43}$$

So $\binom{n}{n} = \binom{n-1}{n-1} = 1$ but $\binom{n-1}{n} = 0$.

Finally, if $0 < k < n$ then

$$\binom{n-1}{k} + \binom{n-1}{k-1} = \frac{(n-1)!}{k!\,(n-k-1)!} + \frac{(n-1)!}{(k-1)!\,(n-k)!} = \frac{(n-1)!}{k!\,(n-k)!} \cdot \left(\frac{n-k}{1} + \frac{k}{1}\right), \tag{2.44}$$

which is $n!/(k!(n-k)!)$ also known as $\binom{n}{k}$. □

It is typical to draw several lines of the binomial coefficients in a stylized triangular array (centered, rather than left-justified as it should properly be). This is called Pascal's triangle

$$\begin{array}{ccccccccccccccccc}
&&&&&&&&1&&&&&&&& \\
&&&&&&&1&&1&&&&&&& \\
&&&&&&1&&2&&1&&&&&& \\
&&&&&1&&3&&3&&1&&&&& \\
&&&&1&&4&&6&&4&&1&&&& \\
&&&1&&5&&10&&10&&5&&1&&& \\
&&1&&6&&15&&20&&15&&6&&1&& \\
&1&&7&&21&&35&&35&&21&&7&&1& \\
1&&8&&28&&56&&70&&56&&28&&8&&1
\end{array}$$

It is easy to check Pascal's relation for the examples written by looking at the triangle. Or, defining the triangle by Pascal's relation, one may calculate the entries and then check that they are equal to the binomial coefficients.

From Pascal's relation, we may conclude the binomial formula.

Theorem 3.2 *Suppose that* $n \in \{0, 1, 2, \dots\}$. *Then for each* $z \in \mathbb{C}$ *we have*

$$\sum_{k=0}^{n} \binom{n}{k} z^k = (1+z)^n \,. \tag{2.45}$$

Proof: For $n = 0$ we have

$$\sum_{k=0}^{n} \binom{n}{k} z^k = \sum_{k=0}^{0} \binom{n}{k} z^k = \binom{0}{0} z^0 = 1 \cdot 1 = 1 \,. \tag{2.46}$$

Since $(1+z)^0$ is also 1, equation (2.45) is true for $n = 0$.

Now suppose that for some $m \in \{0, 1, 2, \dots\}$ we have proved (2.45) for $n = m$. Now let us take $n = m + 1$. Then

$$\sum_{k=0}^{n} \binom{n}{k} z^k = \sum_{k=0}^{m+1} \binom{m+1}{k} z^k \,. \tag{2.47}$$

Note that we can extend the sum to all of $\mathbb{Z}$ if we want, and by the Pascal relation

$$\sum_{k=-\infty}^{\infty} \binom{m+1}{k} z^k = \sum_{k=-\infty}^{\infty} \left(\binom{m}{k} + \binom{m}{k-1} \right) z^k \,. \tag{2.48}$$

There are only finitely many non-zero summands because both $\binom{m}{k}$ and $\binom{m}{k-1}$ are 0 if $k < 0$ or $k > m + 1$. So distributivity may be applied

$$\sum_{k=-\infty}^{\infty} \left(\binom{m}{k} + \binom{m}{k-1} \right) z^k = \sum_{k=-\infty}^{\infty} \binom{m}{k} z^k + \sum_{k=-\infty}^{\infty} \binom{m}{k-1} z^k \,. \tag{2.49}$$

But in the second sum, we may replace the dummy variable k by j and then we may define $k = j - 1$ so that $j = k + 1$. That results in

$$\sum_{k=-\infty}^{\infty} \binom{m}{k} z^k + \sum_{k=-\infty}^{\infty} \binom{m}{k-1} z^k = \sum_{k=-\infty}^{\infty} \binom{m}{k} z^k + \sum_{k=-\infty}^{\infty} \binom{m}{k} z^{k+1} \,. \tag{2.50}$$

So by elementary algebra steps such as distributivity again, we have

$$\sum_{k=-\infty}^{\infty} \binom{m}{k} z^k + z \sum_{k=-\infty}^{\infty} \binom{m}{k} z^k = (1+z) \sum_{k=-\infty}^{\infty} \binom{m}{k} z^k \,. \tag{2.51}$$

And by our induction hypothesis, the last summation is $\sum_{k=0}^{n} \binom{m}{k} z^k$ (because all other choices of $k \in \mathbb{Z}$ give a binomial coefficient equal to 0) and this equals $(1+z)^m$. So, in the end, we obtained $(1+z) \cdot (1+z)^m$ which is $(1+z)^{m+1}$ (recursively). Since $n = m+1$, that equals $(1+z)^n$. So (2.45) is also proved for $n = m + 1$. Therefore, the result follows from proof-by-induction. □

Note that if $x \in \mathbb{C}$ and $y \in \mathbb{C} \setminus \{0\}$, then taking $z = x/y$ and multiplying by y^n, we have

$$\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = (x+y)^n . \tag{2.52}$$

By the symmetry $\binom{n}{k} = \binom{n}{n-k}$, if we assume $x \neq 0$ and y is any number in $\mathbb{C}$, the identity above is still true. Finally if both $x = 0$ and $y = 0$ the identity is evidently true because $0 = 0$. As a special case, if $x = y = 1$ then we obtain the identity

$$\forall n \in \{0, 1, 2, \dots\}, \quad \text{we have} \quad \sum_{k=0}^{n} \binom{n}{k} = 2^n . \tag{2.53}$$

Th identity (2.52) is often what we think of as the binomial formula, but (2.45) has advantages, too.

4 Newton's binomial formula and Catalan's numbers

Let us digress from the main line of our first lesson. It lets us see that there is some connection to concrete analysis. Of course, combinatorics is an important tool in many parts of analysis. But, in some sense, some concrete analysis results in turn lead to combinatorics. (In mathematical physics, these results would often be called "exactly solvable models," and increasingly in probability, too.)

Definition 4.1 *For any $k \in \{0, 1, 2, \dots\}$ and any $\nu \in \mathbb{C}$ define*

$$(\nu)_k = \prod_{j=1}^{k} (\nu + 1 - j) = \nu(\nu - 1) \cdot (\nu - k + 1) , \tag{2.54}$$

with the understanding that an empty product is always interpreted as the multiplicative identity 1. So $(\nu)_0 = 1$.

These are always called falling factorials. We will also call them Pochhammer symbols (which they are not always called). See, for instance, Wikipedia

`https://en.wikipedia.org/wiki/Falling_and_rising_factorials`

Note that if $n \in \{0, 1, 2, \dots\}$ and $k \in \{0, \dots, n\}$, then

$$(n)_k = n(n-1)\cdots(n-k+1) = \frac{n(n-1)\cdot(n-k+1)\cdot(n-k)(n-k-1)\cdots(1)}{(n-k)(n-k-1)\cdots 1} = \frac{n!}{(n-k)!} . \tag{2.55}$$

Therefore, we could rewrite $\binom{n}{k} = (n)_k/k!$.

Theorem 4.2 (Newton's binomial formula) *For any $\nu \in \mathbb{C}$ and $z \in \mathbb{C}$ satisfying $|z| < 1$,*

$$(1+z)^\nu \;=\; e^{\nu \log(1+z)} \;=\; \sum_{k=0}^{\infty} \frac{(\nu)_k}{k!}\, z^k\,. \tag{2.56}$$

Proof: Apply the Taylor theorem at the origin and use the fact that $\frac{d^k}{dz^k}(1+z)^\nu \;=\; (\nu)_k(1+z)^{\nu-k}$, which may be proved by induction. $\square$

Of course, if $\nu = n \in \{0,1,2,\dots\}$, then $(n)_{n+1} = 0$ because $n+1-(n+1) = 0$. So $(n)_k = 0$ for all $k \in \{n+1, n+2, \dots\}$ (recursively). So Theorem 4.2 reduces to Theorem 3.2 for $\nu = n \in \{0,1,2,\dots\}$.

A particular interesting application of this theorem occurs for $\nu = 1/2$. For $|z| < 1$, we have

$$\sqrt{1+z} \;=\; (1+z)^{1/2} \;=\; \sum_{k=0}^{\infty} \frac{(1/2)_k}{k!}\, z^k\,. \tag{2.57}$$

But then note that, for $k > 0$, we have

$$(1/2)_k \;=\; \frac{1}{2}\cdot\frac{(-1)}{2}\cdot\frac{(-3)}{2}\cdots\frac{(-2k+3)}{2} \;=\; \frac{(-1)^{k-1}}{2^k}(2k-3)(2k-5)\cdots(5)(3)(1)\,. \tag{2.58}$$

This may be rewritten as

$$(1/2)_k \;=\; \frac{(-1)^{k-1}}{2^k}\cdot\frac{(2k-2)(2k-3)(2k-4)(2k-5)\cdots(6)(5)(4)(3)(2)(1)}{(2k-2)(2k-4)\cdots(6)(4)(2)}\,, \tag{2.59}$$

which may be written as

$$(1/2)_k \;=\; \frac{(-1)^{k-1}}{2^{2k-1}}\cdots\frac{(2k-2)!}{(k-1)!}\,, \tag{2.60}$$

because $(2)(4)(6)\cdot(2k-4)(2k-2) = 2^{k-1}\,(k-1)!$. So Newton's binomial formula gives this one particular implication that

$$\sqrt{1+z} \;=\; 1 + \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{2^{2k-1}}\cdot\frac{(2k-2)!}{(k-1)!k!}\, z^k\,. \tag{2.61}$$

This can also be written as

$$\frac{1-\sqrt{1-4z}}{4z} \;=\; \frac{1}{2}\sum_{k=0}^{\infty}\frac{1}{k+1}\cdot\binom{2k}{k}\, z^k\,, \tag{2.62}$$

by rewriting k as j and then letting the newly available variable k stand for $j-1$, which means $j = k+1$. The numbers

$$C_k \;=\; \frac{1}{k+1}\binom{2k}{k}\,, \tag{2.63}$$

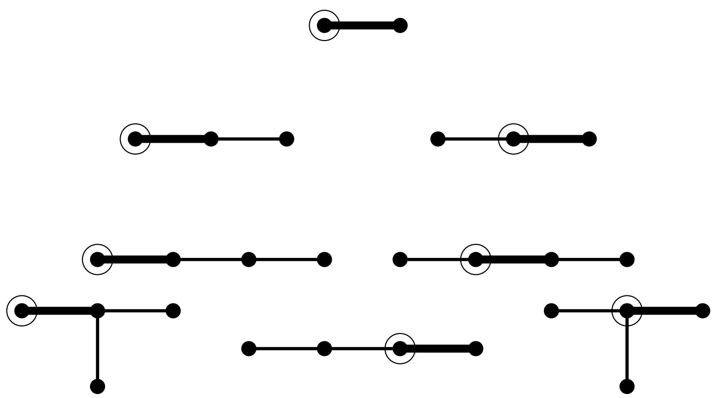


Figure 2.3: Let Θ_n be the set of possible planar maps of trees with n vertices, with one vertex labeled as the root and with one edge incident to the root labeled as the first edge. In the first line we see the 1 planar map tree of Θ_1. In the second line we see the 2 planar map trees of Θ_2. In the last two lines we see the 5 planar map trees of Θ_3. The first three Catalan numbers (after the 0th Catalan number) are 1, 2 and 5.

whose definitions allow us to rewrite

$$\frac{1-\sqrt{1-4z}}{4z} = \frac{1}{2}\sum_{k=0}^{\infty} C_k\, z^k\,, \tag{2.64}$$

are called Catalan's numbers. See

https://en.wikipedia.org/wiki/Catalan_number

We will discuss them later in the context of walks on $\mathbb{N}$ beginning and ending at the left endpoint. They also arise in another context that we will consider: the number of planar maps of trees with a distinguished vertex as the root, and with one of the edges connected to the root distinguised (as the first edge traversed in a ribbon-graph style walk around the tree). See Figure 2.3 for some examples. Then we can rewrite (2.64) as the generating function. Then we will use generating functions to check a quadratic recurrence relation for Catalan's numbers, which will allow us to prove that they count the planar map trees, as shown in Figure 2.3 for just the first 3 numbers. Also see Appendix 2.A for additional related figures. We will save the topic of the quadratic recurrence relation for a later lecture or a later homework exercise.

In a different context, we will enumerate all labeled trees, also called Cayley's theorem, as a special case of the more general Matrix-Tree theorem of Kirchoff. But technically, Cayley's theorem is unrelated to the Catalan numbers.

5 Counting shuffle permutations

Given $n \in \{1, 2, 3, \dots\}$ and $k \in \{0, \dots, n\}$, define $\mathrm{Sh}(k, n-k)$ as a subset of S_n equal to all the permutations $\pi = (\pi_1, \dots, \pi_n)$ such that

$$\pi_1 < \cdots < \pi_k \quad \text{and} \quad \pi_{k+1} < \cdots < \pi_n\,. \tag{2.65}$$

We call these the shuffle permutations of shape $(k, n-k)$.

Another couple of sets which are bijectively equivalent to $\mathrm{Sh}(k, n-k)$ are these:

$$\mathcal{P}_k(\{1,\dots,n\}) \,=\, \{A \subset \{1,\dots,n\} \,:\, |A| = k\}\,, \tag{2.66}$$

which are the cardinality-k subsets of $\{1,\dots,n\}$ (or cardinality-k elements of the power set $\mathcal{P}(\{1,\dots,n\})$); and the binary strings of length n with k ones,

$$\mathcal{B}_n(k) \,=\, \big\{(\sigma_1,\dots,\sigma_n) \in \{0,1\}^n \,:\, \sigma_1 + \dots + \sigma_n = k\big\}\,. \tag{2.67}$$

We write elements of $\mathcal{B}_n(k)$ as $\sigma = (\sigma_1,\dots,\sigma_n)$.

Given a binary string $\sigma \in \mathcal{B}_n(k)$, we can define a set

$$A \,=\, \big\{k \in \{1,\dots,n\} \,:\, \sigma_k = 1\big\}\,. \tag{2.68}$$

Because $\sigma_1 + \dots + \sigma_n = k$ we know that the cardinality of this set is $|A| = k$. So not only is A a subset of $\{1,\dots,n\}$, but it is also in $\mathcal{P}_n(\{1,\dots,n\})$.

Similarly, given a set $A \in \mathcal{P}_k(\{1,\dots,n\})$, we can define a permutation $\pi \in S_n$ by first taking $\pi_1,\dots,\pi_k$ to be the numbers such that: $\pi_1 < \dots < \pi_k$; and

$$A \,=\, \{\pi_1,\dots,\pi_k\}\,. \tag{2.69}$$

We then let $\pi_{k+1},\dots,\pi_n$ be the numbers such that: $\pi_{k+1} < \dots < \pi_n$; and

$$\{1,\dots,n\} \setminus A \,=\, \{\pi_{k+1},\dots,\pi_n\}\,. \tag{2.70}$$

Therefore, not only is $\pi \in S_n$, but it is in $\mathrm{Sh}(k, n-k)$.

Finally, in this triple of mappings, given $\pi \in \mathrm{Sh}(k, n-k)$, let $\sigma \in \{0,1\}^n$ be the binary string such that $\sigma_i = \sum_{j=1}^k \mathbf{1}_{\{i\}}(\pi_j)$, where

$$\mathbf{1}_{\{i\}}(x) \,=\, \begin{cases} 1 & \text{if } x \in \{i\}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.71}$$

Then since we are summing the indicator functions over k terms which will each give 1 one, that means that σ has k ones. So σ is in $\mathcal{B}_n(k)$.

If one composes these three transformations, one just obtains the identity. That would also be true if we cyclically permuted the order. Hence, the three sets are bijectively equivalent.

Theorem 5.1 *For each $n \in \{0,1,2,\dots\}$, we have*

$$|\,\mathrm{Sh}(k, n-k)| \,=\, \binom{n}{k}\,. \tag{2.72}$$

Let us give two proofs of this theorem to demonstrate two methods.

Proof of Theorem 5.1, first method: Let us make a polynomial defined as

$$p_n(z) \;=\; \sum_{k=0}^{n} |\operatorname{Sh}(k, n-k)| z^k \,. \tag{2.73}$$

This is called the generating function of the finite sequence $(|\operatorname{Sh}(k, n-k)|)_{k=0}^{n}$. Because of the bijective equivalences, we may rewrite it as

$$p_n(z) \;=\; \sum_{k=0}^{n} |\mathcal{B}_n(k)| z^k \,. \tag{2.74}$$

But note that we can rewrite this (by summing z^k to itself a total of $|\mathcal{B}_n(k)$ times) as

$$p_n(z) \;=\; \sum_{k=0}^{n} \left(\sum_{(\sigma_1,\ldots,\sigma_n)\in\mathcal{B}_n(k)} z^k \right). \tag{2.75}$$

Then, for each $(\sigma_1, \ldots, \sigma_n) \in \mathcal{B}_n(k)$, we have $k = \sigma_1 + \cdots + \sigma_n$. Therefore, we may rewrite it as

$$p_n(z) \;=\; \sum_{k=0}^{n} \left(\sum_{(\sigma_1,\ldots,\sigma_n)\in\mathcal{B}_n(k)} z^{\sigma_1+\cdots+\sigma_n} \right). \tag{2.76}$$

Finally, note that $\bigcup_{k=0}^{n} \mathcal{B}_n(k)$ is merely $\{0,1\}^n$, and it is a disjoint union. Therefore, the sum over these disjoint sets equals the total sum over the disjoint union:

$$p_n(z) \;=\; \sum_{(\sigma_1,\ldots,\sigma_n)\in\{0,1\}^n} z^{\sigma_1+\cdots+\sigma_n} \,. \tag{2.77}$$

Then we may rewrite the sum over the Cartesian product set $\{0,1\}^n$ as a product over sums

$$p_n(z) \;=\; \sum_{\sigma_1=0}^{1} z^{\sigma_1} \sum_{\sigma_2=0}^{1} z^{\sigma_2} \cdots \sum_{\sigma_n=0}^{1} z^{\sigma_n} \,, \tag{2.78}$$

because $z^{\sigma_1+\cdots+\sigma_n} = z^{\sigma_1} z^{\sigma_2} \cdots z^{\sigma_n}$. Each of the individual sums is equal to $z^0 + z^1$ or $1 + z$. So

$$p_n(z) \;=\; (1+z)^n \,, \tag{2.79}$$

But we know $(1+z)^n$ is the generating function for the binomial coefficients by Theorem 3.2. So

$$|\operatorname{Sh}(k, n-k)| \;=\; \frac{1}{k!} \frac{d^k}{dz^k} (1+z)^n \Big|_{z=0} \;=\; \binom{n}{k} . \tag{2.80}$$

□

More generally, if we ever want to check equivalences based on generating functions in case the generating functions are polynomials, we may want to recall the Fundamental theorem of algebra or maybe the identity theorem from Complex Analysis. We will not probe more deeply into generating functions at the moment. They will be the subject of a second set of notes. They are the subject of entire monographs! See, for example, Wilf [9].

Now we consider a second proof. This proof will be based on the "double-counting" method. This is a variant of direct bijective equivalence. If we want to prove that $|A_n| = a_n$ and we already know that $|B_n| = a_n$, then it suffices to prove that A_n is bijectively equivalent to B_n.

But sometimes instead we want to prove that $|A_n| = a_n$, and what we can prove is that for some other sets B_n and C_n, we have $A_n \times B_n$ is bijectively equivalent to C_n. If we know that $|C_n|/|B_n| = a_n$, then we also accomplish the desired goal. This is called double-counting because we count the number of terms in the two sides of the equation

$$A_n \times B_n \,=\, C_n \tag{2.81}$$

in two different ways. But due to the equation, both counting methods must yield the same number. (We gave a reference to the Wikipedia page earlier in these notes.)

Proof of Theorem 5.1, second method: We are going to elide over some details. Suppose that we have a permutation $\pi \in \mathrm{Sh}(k, n-k)$. Also, suppose that we have additionally two permutations $\rho \in S_k$ and $\sigma \in S_{n-k}$. Create a first function $f : \{1, \dots, n\} \to \{1, \dots, n\}$ by the formula

$$\forall j \in \{1, \dots, n\}\,, \quad \text{we have} \quad f(j) \,=\, \pi_j\,. \tag{2.82}$$

Of course this means $f(1) < \cdots < f(k)$ and $f(k+1) < \cdots < f(n)$, since π in in $\mathrm{Sh}(k, n-k)$. Then create two other functions, $g : \{1, \dots, n\} \to \{1, \dots, n\}$, defined so that

$$\forall j \in \{1, \dots, n\}\,, \quad \text{we have} \quad g(j) \,=\, \begin{cases} \rho_j & \text{if } j \leq k, \\ j & \text{if } j > k, \end{cases} \tag{2.83}$$

and $h : \{1, \dots, n\} \to \{1, \dots, n\}$, defined so that

$$\forall j \in \{1, \dots, n\}\,, \quad \text{we have} \quad h(j) \,=\, \begin{cases} j & \text{if } j \leq k, \\ k + \sigma_{j-k} & \text{if } k+1 \leq j \leq n. \end{cases} \tag{2.84}$$

It is easy to see that $g \circ h = h \circ g$. We obtain our final permutation $\tau \in S_n$ such that, for $\tau = (\tau_1, \dots, \tau_n)$,

$$\forall j \in \{1, \dots, n\}\,, \quad \text{we have } \tau_j \,=\, \big(f \circ g \circ h\big)(j)\,. \tag{2.85}$$

Then it is an exercise to see that this mapping from $\mathrm{Sh}(k, n-k) \times S_k \times S_{n-k}$ to S_n is a bijection. We skip this exercise. Therefore, by the double-counting principle and Theorem 1.1 and Lemma 2.1

$$|\,\mathrm{Sh}(k, n-k)| \cdot (k!) \cdot ((n-k)!) \,=\, n!\,. \tag{2.86}$$

Hence, $|\,\mathrm{Sh}(k, n-k)| = n!/(k! \cdot (n-k)!)$. □

We have skipped the details of the bijection. This can be done. In some sense, one good approach is to think along the lines of the Fisher-Yates-Knuth algorithm again. Indeed, that way one can even show that $\mathrm{Inv}(\tau) = \mathrm{Inv}(\pi) + \mathrm{Inv}(\rho) + \mathrm{Inv}(\sigma)$, which we will see again (in lecture or in homework exercises) in connection with the q-binomial formula. An example is shown in Figure 2.4.

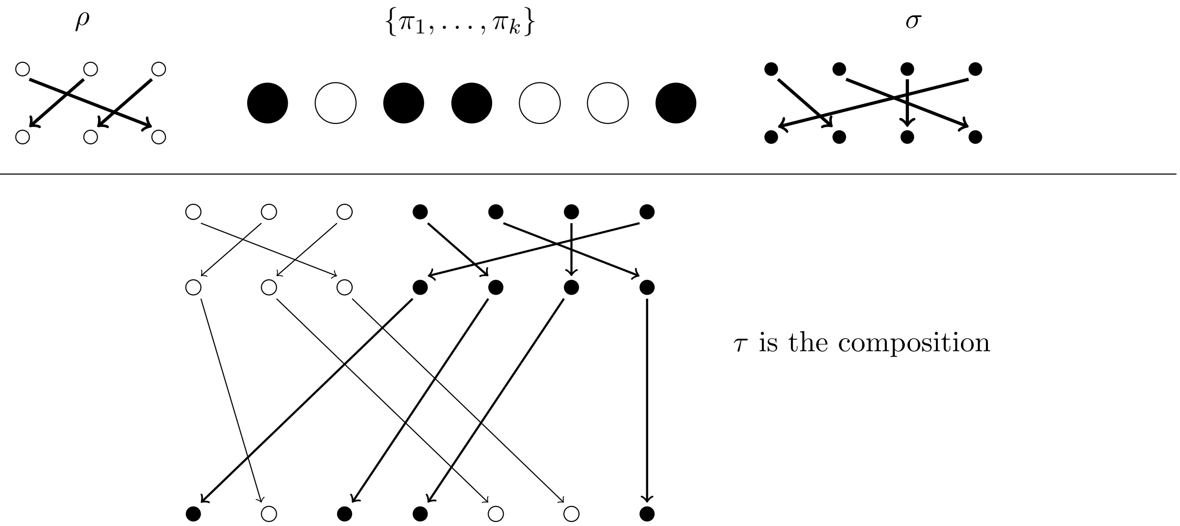


Figure 2.4: This figure demonstrates an example of the mapping taking $\mathrm{Sh}(k, n-k) \times S_k \times S_{n-k}$ bijectively onto S_n. In the top picture the black dots represent the locations of $\{\pi_1, \ldots, \pi_k\}$ and the white dots represent the locations of $\{\pi_{k+1}, \ldots, \pi_n\}$. The final permutation is $\tau = (6, 2, 5, 3, 7, 4, 1)$. Notice that in this example $\mathrm{Inv}(\tau) = \mathrm{Inv}(\pi) + \mathrm{Inv}(\rho) + \mathrm{Inv}(\sigma)$ because $7 + 2 + 4 = 13$ and the inversion pairs of τ are $\{\{1,2\},\{1,3\},\{1,4\},\{1,6\},\{1,7\},\{2,7\},\{3,4\},\{3,6\},\{3,7\},\{4,7\},\{5,6\},\{5,7\},\{6,7\}\}$.

6 The "hockey stick" identity

A question arises about the *best* way to generalize the first inductive identity (2.6). Note that the identity in question is used in lower-level mathematics, for example in proving that

$$\int_0^1 x\,dx = \frac{1}{2}\,, \tag{2.87}$$

because

$$\int_0^1 x\,dx = \lim_{n\to\infty} \sum_{i=1}^n x_i\,\Delta x = \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^n i = \lim_{n\to\infty} \frac{n+1}{2n}\,, \tag{2.88}$$

where $x_i = i/n$ and $\Delta x = 1/n$. Therefore, in a first course on integrals sometimes one encounters several other identities of the same type as (2.6), but with higher powers, such as

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}\,. \tag{2.89}$$

But such identites become hard to remember for much higher powers. A simpler identity to remember is the following.

Theorem 6.1 *For each $k \in \{0, 1, 2, \dots\}$ and each $m \in \{k, k+1, k+2, \dots\}$, we have*

$$\sum_{n=k}^{m} \binom{n}{k} = \binom{m+1}{k+1}. \tag{2.90}$$

Proof: This can be proved using induction and Pascal's relation, Theorem 3.1, by inducting on m. We leave it as a homework exercise to do that. $\square$

The reason that this is called the hockey-stick identity is that in Pascal's triangle it somewhat resembles a hockey-stick. This is especially true if one draws Pascal's triangle in a left-justified format

$$\begin{array}{ccccccc}
1 \\
1 & 1 \\
1 & 2 & \boxed{1} \\
1 & 3 & \boxed{3} & 1 \\
1 & 4 & \boxed{6} & 4 & 1 \\
1 & 5 & \boxed{10} & 10 & 5 & 1 \\
1 & 6 & 15 & \boxed{20} & 15 & 6 & 1
\end{array}$$

In the case drawn $k = 2$ and $m = 5$. We boxed the number $\binom{n}{k}$ for $k \le n \le m$, and we also boxed the sum $\binom{m+1}{k+1}$. The pattern of the boxes is said to resemble a hockey-stick. (You can also look this up on Wikipedia for another more colorful representation.)

Note that the hockey-stick identity is sufficient to deduce that $\int_0^1 x^n\,dx = 1/(n+1)$, as is needed in a first course in integration. That is because

$$\lim_{m\to\infty} m^{-k-1} \sum_{n=k}^{m} \binom{n}{k} = \frac{1}{k!} \lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} \frac{\prod_{j=1}^{k}(n+1-j)}{m^k} = \frac{1}{k!} \lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} \left(\frac{n}{m}\right)^k , \tag{2.91}$$

since $m^{-k} \prod_{j=1}^{k}(n+1-j) = \prod_{j=1}^{k} \left(\frac{n}{m} - \frac{j-1}{m}\right)$ can be expanded into a finite sum of terms such as $(n/m)^k$ as well as lower order terms, expressed using a sum over $r \in \{1, \dots, k\}$, where the rth summand is $(n/m)^{k-r}$ times

$$(-1)^r \sum_{1 \le j_1 < \cdots < j_r \le k-1} j_1 \cdots j_r \cdot m^{-r} , \tag{2.92}$$

where the displayed sum may be seen to converge to 0 in the $m \to \infty$ limit. But writing $x_m = n/m$ and $\Delta x = 1/m$ the right-most side of (2.91) does equal

$$\frac{1}{k!} \lim_{m\to\infty} \sum_{n=k}^{m} x_m^k \, \Delta x , \tag{2.93}$$

which can be seen to equal $(k!)^{-1}\int_0^1 x^k\,dx$ (because only the first $k-1$ terms of the Riemann sum are missing). But, by the hockey-stick identity, the left-most side of (2.91) is equal to

$$\lim_{m\to\infty}\frac{1}{m^{k+1}}\binom{m+1}{k+1} = \frac{1}{(k+1)!}\lim_{m\to\infty}\prod_{j=1}^{k+1}\left(1-\frac{j-2}{m}\right), \tag{2.94}$$

which equals $1/(k+1)!$ by a reasoning similar to that given before. (One could use induction to clean-up some of the unwieldy parts of the argument.) So this proves

$$\frac{1}{k!}\int_0^1 x^k\,dx = \frac{1}{(k+1)!}, \tag{2.95}$$

hence proving $\int_0^1 x^k\,dx = 1/(k+1)$ for $k\in\{0,1,2,\dots\}$, prior to using the fundamental theorem of calculus. This would satisfy the ambitious calculus instructor's goal.

7 Two special cases of the Chu-Vandermode identity

We wish to note two more combinatorial identities, which will be proved by direct bijections or double-counting, rather than by induction. Both of these are special cases of the Chu-Vandermonde identity. The first identity is sometimes called the man-woman committee identity

Theorem 7.1 *Suppose that we have numbers $m,n,k\in\{0,1,2,\dots\}$. Then*

$$\sum_{j=0}^{k}\binom{m}{j}\binom{n}{k-j} = \binom{m+n}{k} \tag{2.96}$$

Instead of proving this directly, imagine that there is a group of m men and n women as the permanent faculty in a college department. It is decided to make a committee with exactly k people, of which there are $\binom{m+n}{k}$ possible choices. But after the fact, the administrator states that they actually do care about how many men and women are in the committee. If we want j men and $k-j$ women, then the number of possibilities are $\binom{m}{j}\binom{n}{k-j}$. But at the administrator's whim, they revert their decision, so that once again they do not care about the number of men versus women on the committee, after all. So if you sum over the j, you must get the identity written above.

We will not systematize this argument into a more formal proof.

Another picture will help to understand, for the following related identity, which directly generalizes the hockey-stick identity, Theorem 6.1.

Theorem 7.2 *For $L\in\{0,1,2,\dots\}$, $r\in\mathbb{N}$ and $j\in\{1,\dots,r\}$, we have*

$$\sum_{x=j}^{L-r+j}\binom{x-1}{j-1}\binom{L-x}{r-j} = \binom{L}{r}. \tag{2.97}$$

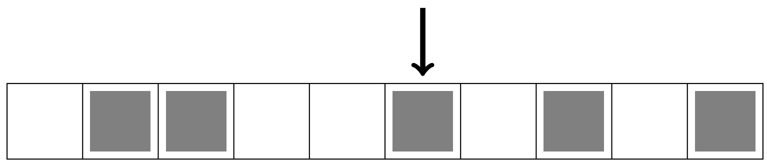


Figure 2.5: Suppose that, like with the man-woman-committee problem, we condition on something. That means we first count the number of ways to obtain a given choice with that value fixed, and then we drop the requirement of fixing that number. So we sum over all possibilities. Here we condition on where the j^{th} point will be. If it is as x then there are $\binom{x-1}{j-1}$ choices for the points before and $\binom{L-x}{r-j}$ choices for the points afterwards. Here L is the total length of the pill box, and r is the number of squares to be filled in. Of course if we do not condition, then the total number of choices are $\binom{L}{r}$. In this example $L=10$, $r=5$ and $j=3$ with $x=6$ in this case (and only 1 of the choices for the 2 positions preceding and only 1 of the choices for the two positions after the j^{th} particle shown).

Note that if we view $\binom{-1}{k}$ as $\delta_{k,-1}$ then we can also allow $k=0$ in which case we could also allow r to equal 0. (That identity would make sense if we used Gamma functions in place of factorials. A common theme in combinatorics these days is to explore generalizations that may be true even if they were not originally defined that way.)

The way to get the hockey-stick identity is to take $n=x-1$, $k=j-1$ and $r=j$. Then note that means $r=k+1$. So making all necessary substitutions gives

$$\sum_{n=k}^{L-1} \binom{n}{k}\binom{L-n-1}{0} = \binom{L}{k+1}. \tag{2.98}$$

Using the fact that $\binom{L-n-1}{0}=1$ as long as $L \geq n+1$, and taking $L=m+1$, we obtain (2.90).

In Figure 2.5 we show a graphical way of checking the theorem. Of course we arbitrarily chose to make a linearly ordered set of boxes. But at least that does allow one special extra result, which we mention now, but which we will not prove, now. Namely, there is a q-Chu-Vandermonde identity which relies on the 1-dimensional linear ordering of the boxes. But the q-Chu-Vandermonde identity will be the subject for a later lecture or else a later homework exercise.

In a certain way, the latter identity is related to the negative hypergeometric distribution in probability theory. More importantly, there is a way, taking certain asymptotic limits, to see these identities as related to the Beta integral identities in cases of integer arguments. In this way, the negative hypergeometric distribution is related to the Beta distribution which one first encounters in a statistics course when learning the formula of Bayes. We will save this discussion for the first homework assignment.

2.A The number of "positive" walks on $\mathbb{Z}$

Here we want to give two pictures to better describe two alternative sets enumerated by the Catalan numbers. These are shown in Figure 2.6 and 3.1. A good reference for all of this is the

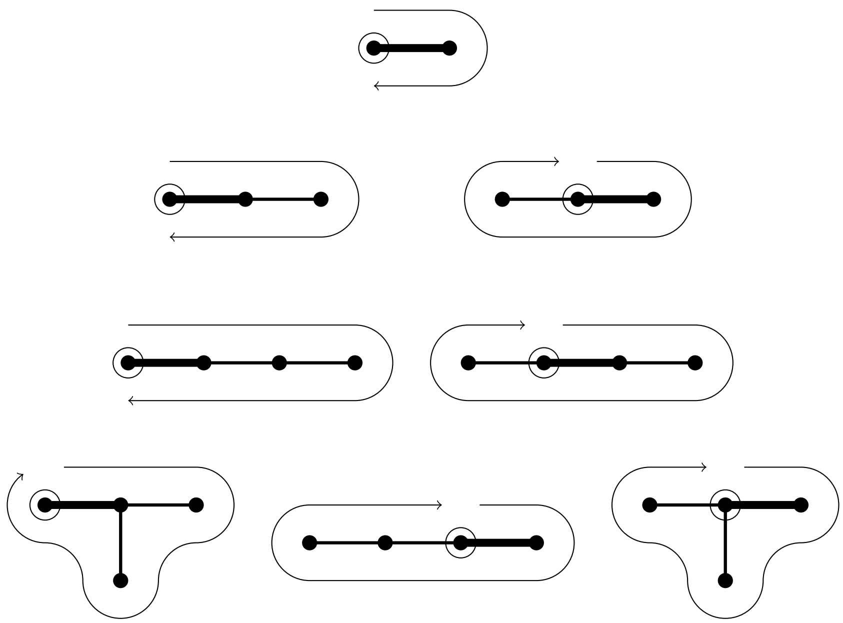


Figure 2.6: Here we plot the planar map trees of Θ_n again, for $n = 1$ (first line), $n = 2$ (second line) and $n = 3$ (third line). But this time we draw the ribbon-graph style path encircling the tree. More specifically we draw a loop around the map of the tree going counterclockwise. We start at the root, and the first edge we pass is the distinguised edge. For each edge, view its positive direction being that going out from the root towards the leaves of the tree. Each time we pass an edge for the first time, we pass on the left. When we pass the edge the second (and final) time, we pass it on the right.

probability textbook by William Feller [4]. But we will save that discussion for a later lecture or a later homework exercise.

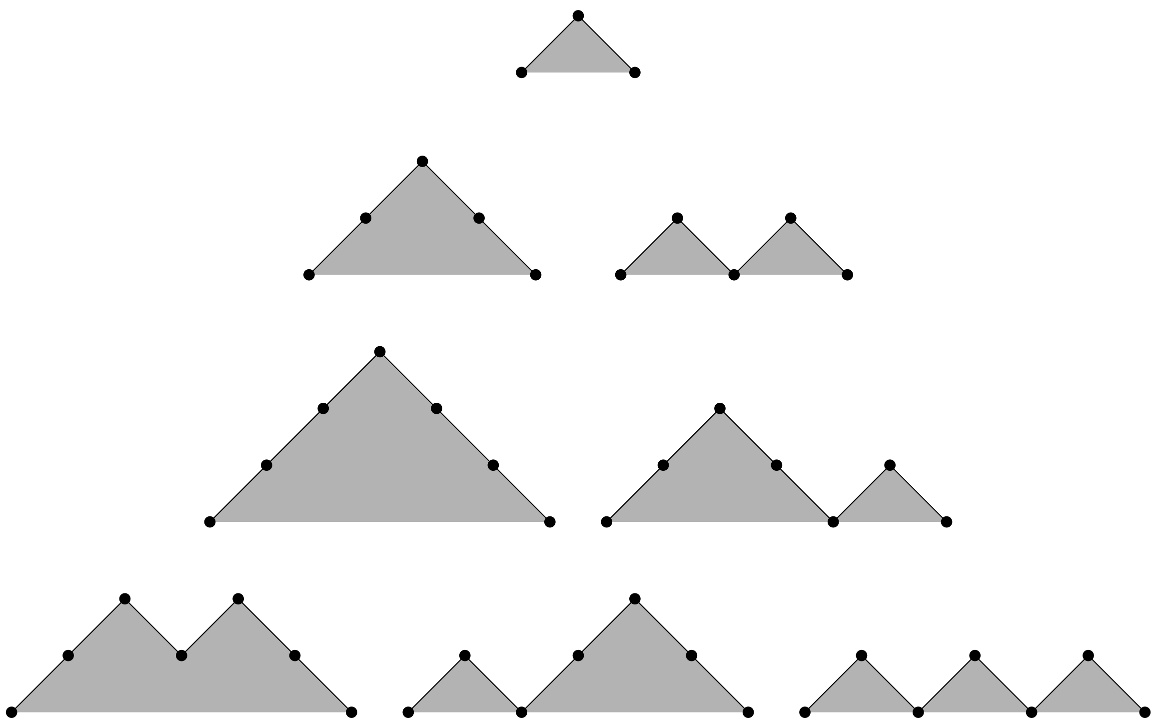


Figure 2.7: Referring to the planar map trees, we use the following rule to draw walks on $\mathbb{N}$. The first time you pass a give edge on the outer loop around a planar map tree, you step up in the walk on $\mathbb{N}$. The second time you pass the edge you step down. These map to walks in $\mathbb{N}$ of length $2n$ beginning and ending at the left endpoint of $\mathbb{N}$. These are called "positive walks" in some contexts.

Chapter 3

Generating Functions

1 Generating function using power series

Given two sets, X and Y, let us denote the set of all functions $f : X \to Y$ as X^Y.

Theorem 1.1 *Suppose that X and Y are finite, non-empty sets. Then*

$$\left|X^Y\right| = (|X|)^{|Y|} . \tag{3.1}$$

Proof: Recall from "Some Notes on Basic Combinatorics," Theorem 1.5, that if two sets A and B are bijective and if $|A| = n$, then also $|B| = n$. Let us call this the "basic fact of combinatorics." But also, the "fundamental principle of combinatorics" says that for two finite sets A and B, the cardinality of the Cartesian product $A \times B$ is equal to $|A|$ times $|B|$: Theorem 2.1 in the same notes.

We will first prove the formula in the special case that $Y = \{1, \dots, n\}$ for some $n \in \mathbb{N}$. We use proof-by-induction. Firstly, assume $Y = \{1\}$. Then $f : Y \to X$ is determined by one number $f(1) \in X$. The mapping taking f to $f(1)$ is a bijection from $X^{\{1\}}$ onto X. Its inverse mapping takes any number $x \in X$ and defines a function $f : \{1\} \to X$ by letting $f(1)$ be equal to x. Therefore, by the "basic fact of combinatorics," $|X^{\{1\}}| = |X|$ which is $|X|^1$.

Now assume the theorem was proved for $Y = \{1, \dots, n\}$ for the special case of $n = k$ for some fixed $k \in \mathbb{N}$. Let us consider the new problem of taking $Y = \{1, \dots, n\}$ for $n = k + 1$. Then we can map a function $f : Y \to X$ to a pair (g, x) where g is in $X^{\{1,\dots,k\}}$ and x is in X. Namely, let $g = f \restriction \{1, \dots, k\}$ and let $x = f(k + 1)$. This is a bijection onto $X^{\{1,\dots,k\}} \times X$, with inverse mapping taking a general (g, x) onto the function $f : \{1, \dots, k + 1\} \to X$ satisfying that $f(i) = g(i)$ for $i \leq k$ and $f(k + 1) = x$. So by the The Fundamental Principle of Combinatorics, we have that $|X^{\{1,\dots,k+1\}}| = |X^{\{1,\dots,k\}}| \cdot |X|$ which is $|X|^k \cdot |X|$ by our induction hypotheses. But that is $|X|^{k+1}$, proving the induction step.

If Y were some other finite, non-empty set, then there would be a bijection $\psi : \{1, \dots, n\} \to Y$ for that $n \in \mathbb{N}$ equal to the cardinality of Y. Then we could define a bijection from X^Y to $X^{\{1,\dots,n\}}$ by mapping any $f : Y \to X$ to the function $f \circ \psi$. This is a bijection whose inverse mapping takes a function $g : \{1, \dots, n\} \to X$ to the function $g \circ \psi^{-1}$. So by the "basic fact of combinatorics," we know that $|X^Y| = |X^{\{1,\dots,n\}}|$. In the previous two paragraphs, we learned that this is $|X|^n$ which is $|X|^{|Y|}$ since $n = |Y|$. $\square$

That theorem justifies the notation. But we often use the notation even when the underlying sets are infinite.

Definition 1.2 *Suppose that we have two sequences,* $\boldsymbol{a} \in \mathbb{C}^{\{0,1,\dots\}}$ *and* $\boldsymbol{\gamma} \in (\mathbb{C} \setminus \{0\})^{\{0,1,\dots\}}$ *written as*

$$\boldsymbol{a} = (a_0, a_1, a_2, \dots) \quad \text{and} \quad \boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2, \dots), \tag{3.2}$$

where we assume that $\gamma_0 \neq 0$, $\gamma_1 \neq 0$, *etc., because* $\boldsymbol{\gamma}$ *is in* $(\mathbb{C} \setminus \{0\})^{\{0,1,\dots\}}$. *(We have written them as ordered* ∞*-tuples, as is conventional, instead of as functions.) Then, if we have the condition that*

$$\limsup_{n\to\infty} |a_n \gamma_n|^{1/n} < \infty, \tag{3.3}$$

we may construct the power series

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) = \sum_{n=0}^{\infty} a_n \gamma_n z^n, \tag{3.4}$$

where the radius of convergence satisfies

$$R \geq \left(\limsup_{n\to\infty} |a_n \gamma_n|^{1/n} \right)^{-1}. \tag{3.5}$$

The reason for allowing a choice of $\boldsymbol{\gamma}$ is to allow to define two types of generating functions. We want the radius of convergence to be positive. But also, certain types of generating functions allow for different tools for proving combinatorial identites. We choose to avoid discussion of the analytical details of the proof of the convergence criterion. But if any student wants to review it, for example from MA 440/441or MA 540/541, the topic to consult includes Weierstrass's M-test.

Definition 1.3 *Suppose that we have a sequence,* $\boldsymbol{a} = (a_0, a_1, \dots)$ *in* $\mathbb{C}^{\{0,1,\dots\}}$. *If we take* $\boldsymbol{\gamma}$ *in* $(\mathbb{C} \setminus \{0\})^{\{0,1,\dots\}}$ *by writing*

$$\gamma_0 = \gamma_1 = \dots = 1, \tag{3.6}$$

then the resulting generating function is called the ordinary power series generating function

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) = \sum_{n=0}^{\infty} a_n z^n, \tag{3.7}$$

sometimes abbreviated opsgf in Wilf [9]. If, instead, we take $\boldsymbol{\gamma} \in (\mathbb{C} \setminus \{0\})^{\{0,1,\dots\}}$ *as* $\gamma_n = 1/n!$, *then the resulting generating function is called the exponential generating function*

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) = \sum_{n=0}^{\infty} \frac{a_n z^n}{n!}. \tag{3.8}$$

Recall that a central type of problem in combinatorics is to prove that two sequences of numbers are equal. For example, we may have sets $A_0, A_1, A_2, \ldots$ and $B_0, B_1, B_2, \ldots$ and we want to prove that

$$\forall n \in \{0, 1, \ldots\}, \quad \text{we have} \quad |A_n| = |B_n|. \tag{3.9}$$

Then we could define $\boldsymbol{a}$ and $\boldsymbol{b}$ such that for each $n \in \{0, 1, \ldots\}$, we have $a_n = |A_n|$ and $b_n = |B_n|$. Then we could check if the opsgf or exponential generating functions are equal.

Theorem 1.4 *Suppose that we have two sequences* $\boldsymbol{a} = (a_0\,,\ a_1\,,\ a_2\,,\ \ldots)$ *and* $\boldsymbol{b} = (b_0\,,\ b_1\,,\ b_2\,,\ \ldots)$ *in* $\mathbb{C}^{\{0,1,\ldots\}}$*. Also suppose that we have a sequence* $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots)$ *in* $(\mathbb{C} \setminus \{0\})^{\{0,1,\ldots\}}$ *such that*

$$\limsup_{n\to\infty} |\gamma_n a_n|^{1/n} < \infty \quad \textit{and} \quad \limsup_{n\to\infty} |\gamma_n b_n|^{1/n} < \infty. \tag{3.10}$$

If there exists some $\epsilon > 0$ *such that*

$$forall z \in \mathbb{C} \quad \textit{such that } |z| < \epsilon, \quad \textit{we have} \quad f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) = f_{\boldsymbol{\gamma},\boldsymbol{b}}(z), \tag{3.11}$$

then $\boldsymbol{a} = \boldsymbol{b}$.

Proof: By the convergence criteria in (3.10) we know that $f_{\boldsymbol{\gamma},\boldsymbol{a}}(z)$ and $f_{\boldsymbol{\gamma},\boldsymbol{a}}(z)$ both have positive radii of convergence. They are both infinitely differentiable (with respect to z) inside their radius of convergence. Moreover,

$$\forall k \in \{0, 1, \ldots\}, \quad \text{we have} \quad a_k = \frac{1}{k!\,\gamma_k} \cdot \frac{d^k}{dz^k} f_{\boldsymbol{\gamma},\boldsymbol{a}}(z)\bigg|_{z=0}, \tag{3.12}$$

and

$$\forall k \in \{0, 1, \ldots\}, \quad \text{we have} \quad b_k = \frac{1}{k!\,\gamma_k} \cdot \frac{d^k}{dz^k} f_{\boldsymbol{\gamma},\boldsymbol{b}}(z)\bigg|_{z=0}. \tag{3.13}$$

If the functions are equal for all $|z| < \epsilon$, then all their derivatives at 0 are equal. So the two sequences $\boldsymbol{a}$ and $\boldsymbol{b}$ are equal. □

1.1 An example: the factorial sequence and Gamma integrals

As an example, let us prove the following identity.

Theorem 1.5 *For each* $k \in \{0, 1, \ldots\}$*, we have*

$$k! = \int_0^\infty t^k e^{-t}\, dt. \tag{3.14}$$

We will prove this by the generating function method.

Let $a_k = k!$ and $b_k = \int_0^\infty t^k e^{-t}\,dt$, and let us take $\gamma_k = 1/k!$. So $a_k\gamma_k = 1$ for every $k \in \{0,1,\dots\}$ which means that the limsup for the $\boldsymbol{a}$ sequence in equation (3.10) is 1. Let us quickly note the well-known geometric sum identity

$$\forall n \in \{0,1,\dots\}\,,\quad \text{we have}\ \sum_{k=0}^{n} x^k \;=\; \frac{1-x^{n+1}}{1-x}\,, \tag{3.15}$$

as polynomial, rational functions. (So the identity is true for all $x \neq 1$, while the limit of the right-hand-side is equal to the limit of the left-hand-side at $x = 1$.) That can be proved by induction since it is evidently true for $n = 0$ since we get 1 on both sides, while for $n \in \mathbb{N}$ we have

$$\sum_{k=0}^{n} x^k \;=\; 1 + x\sum_{k=0}^{n-1} x^k\,. \tag{3.16}$$

If we made the induction hypothesis, this would equal

$$1 + x\cdot\frac{(1-x^n)}{1-x} \;=\; 1 + \frac{x-x^{n+1}}{1-x} \;=\; \frac{1-x}{1-x} + \frac{x-x^{n+1}}{1-x} \;=\; \frac{1-x^{n+1}}{1-x}\,. \tag{3.17}$$

Similarly, for $z \in \mathbb{C}$ such that $|z| < 1$, we may take limits to conclude the geometric series identity

$$\sum_{k=0}^{\infty} z^k \;=\; \frac{1}{1-z}\,. \tag{3.18}$$

So, since $a_k\gamma_k = 1$ for all k, we have, for all $z \in \mathbb{C}$ such that $|z| < 1$,

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \;=\; \frac{1}{1-z}\,. \tag{3.19}$$

Proof of Theorem 1.5: In order to calculate the exponential generating function for the $\boldsymbol{b}$ sequence, note that for $x \geq 0$ we may rearrange the order of the sum and integral

$$\sum_{k=0}^{\infty}\frac{x^k}{k!}\int_0^\infty t^k e^{-t}\,dt \;=\; \int_0^\infty \sum_{k=0}^{\infty}\frac{(xt)^k}{k!}\,e^{-t}\,dt\,. \tag{3.20}$$

The latter is equal to $\int_0^\infty e^{xt}e^{-t}\,dt$. So it is integrable as long as $x < 1$, i.e., finite.

Then that implies that the radius of converges for $f_{\boldsymbol{\gamma},\boldsymbol{b}}$ is also $R = 1$. Moreover, since $e^{zt}e^{-t} = e^{-(1-z)t}$, we have, for all $z \in \mathbb{C}$ satisfying $|z| < 1$,

$$\int_0^\infty e^{-(1-z)t}\,dt \;=\; \frac{1}{1-z}\,, \tag{3.21}$$

this proves that $f_{\boldsymbol{\gamma},\boldsymbol{b}}$ is equal to $f_{\boldsymbol{\gamma},\boldsymbol{a}}$. So, that implies (3.14) by Theorem 2.1. $\square$

The Gamma function $\Gamma : (0,\infty) \to \mathbb{R}$ is defined as the integral

$$\Gamma(r) \;=\; \int_0^\infty t^{r-1}e^{-t}\,dt\,. \tag{3.22}$$

Note that by scaling $\int_0^\infty t^{r-1}e^{-xt}\,dt = \Gamma(r)/x^r$ for each $x > 0$. That is a motivation for why the power of t is $r-1$ instead of r.

So equation (3.14) may be rewritten as saying that

$$\forall k \in \{0,1,\dots\}\,,\quad \text{we have}\ \ k! \;=\; \Gamma(k+1)\,. \tag{3.23}$$

1.2 Another example: moment generating functions

Although this is not strictly within the topic of combinatorics, it does intersect with combinatorics: moment generating functions for probability distribution functions. Suppose that we have a function $\rho : \mathbb{R} \to \mathbb{R}$ satisfying

- for all $x \in \mathbb{R}$ actually $\rho(x) \geq 0$,
- and $\int_{-\infty}^{\infty} \rho(x)\, dx = 1$.

Then ρ is called a probability distribution function. Properties that follow from this set-up form the basis of probability theory. Many examples in probability theory are related to specific combinatorial identities.

For a probability distribution function, we define the kth moments to be

$$\forall k \in \{0, 1, \dots\}\,, \quad \text{we have} \;\; m_k \;=\; \int_{-\infty}^{\infty} x^k\, \rho(x)\, dx\,, \tag{3.24}$$

assuming that integrand is integrable. These are collected in an exponential generating function

$$\varphi(z) \;=\; \sum_{k=0}^{\infty} \frac{m_k}{k!}\, z_k\,. \tag{3.25}$$

But by the Fubini-Tonelli theorem, we may rewrite this as

$$\varphi(z) \;=\; \int_{-\infty}^{\infty} e^{zx}\, \rho(x)\, dx\,, \tag{3.26}$$

assuming that we are within the radius of convergence of the exponential generating function.

For example, there is a way to see that $\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}$. Therefore,

$$\rho(x) \;=\; \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\,, \tag{3.27}$$

does define a probability distribution function called the standard, normal distribution because of its central position in probability theory. Then the moment generating function is

$$\varphi(t) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2}\, dx\,. \tag{3.28}$$

Suppose that t is a real number, not a complex number. By combining the exponents and changing variables to $y = x + t$ it is possible to see that

$$\varphi(t) \;=\; e^{t^2/2}\,. \tag{3.29}$$

This may be Taylor expanded as $\sum_{k=0}^{\infty} t^{2k}/(2^k\, k!)$. So this implies that for the standard, normal probability distribution function the moments are

$$\forall k \in \{0, 1, \dots\}\,, \quad \text{we have} \;\; m_k \;=\; \begin{cases} k!/(2^{k/2}\,(k/2)!) & \text{if } k \in \{0, 2, 4, 6, \dots\}, \text{ and} \\ 0 & \text{if } k \in \{1, 3, 5, \dots\}. \end{cases} \tag{3.30}$$

Note that

$$\forall k \in \{0, 2, 4, 6, \dots\}\,, \quad \text{we have} \quad \frac{k!}{2^{k/2}\,(k/2)!} \;=\; (k-1)(k-3)(k-5)\cdots(5)(3)(1)\,, \tag{3.31}$$

and this quantity is often called $(k-1)!!$. See for instance the Wikipedia page on the standard, normal distribution under "moments" or

`https://en.wikipedia.org/wiki/Double_factorial`

Note that we may make a change of variables by defining $x^2 = y$ in the formula

$$\forall k \in \{0, 1, 2, \dots\}\,, \quad \text{we have} \quad m_{2k} \;=\; \frac{2}{\sqrt{2\pi}} \int_0^\infty x^{2k} e^{-x^2/2}\, dx\,. \tag{3.32}$$

Then we obtain

$$\forall k \in \{0, 1, 2, \dots\}\,, \quad \text{we have} \quad \frac{(2k)!}{2^k\, k!} \;=\; \frac{1}{\sqrt{2\pi}} \int_0^\infty y^k e^{-y/2}\, y^{-1/2}\, dy\,. \tag{3.33}$$

But from our previous discussion this is $2^{k+(1/2)}\Gamma\left(k+\frac{1}{2}\right)$. Therefore, we derived this:

Theorem 1.6 *For each $k \in \{0, 1, \dots\}$, we have*

$$\Gamma\left(k + \frac{1}{2}\right) \;=\; \sqrt{\pi}\, \frac{(2k)!}{2^{2k}\, k!}\,. \tag{3.34}$$

Of course, historically, the most difficult step was determining that $\Gamma(1/2) = \sqrt{\pi}$ which can be seen as part of Stirling's formula (the part that Stirling did which de Moivre could not do). In other words, when we quoted the fact that $\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}$, that implicitly used $\Gamma(1/2) = \sqrt{\pi}$ since that is an equivalent calculation (which must be done).

2 Formal power series and products of generating functions

Let us start with the following theorem for opsgf's.

Theorem 2.1 *Suppose that we have two sequences* $\boldsymbol{a} = (a_0\,,\ a_1\,,\ a_2\,,\ \dots)$ *and* $\boldsymbol{b} = (b_0\,,\ b_1\,,\ b_2\,,\ \dots)$ *in* $\mathbb{C}^{\{0,1,\dots\}}$*, such that*

$$\limsup_{n\to\infty} |a_n|^{1/n} \,<\, \infty \quad \text{and} \quad \limsup_{n\to\infty} |b_n|^{1/n} \,<\, \infty\,. \tag{3.35}$$

Suppose that $\boldsymbol{\gamma} \in (\mathbb{C}\setminus\{0\})^{\{0,1,\dots\}}$ *is* $\gamma_0 = \gamma_1 = \dots = 1$*. Then, defining* $\boldsymbol{c} = (c_0\,,\ c_1\,,\ c_2\,,\ \dots)$ *by*

$$\forall n \in \{0,1,\dots\}\,, \quad \text{we have} \quad c_n \,=\, \sum_{k=0}^{n} a_k b_{n-k}\,, \tag{3.36}$$

we have that

$$\limsup_{n\to\infty} |c_n|^{1/n} \,<\, \infty\,, \tag{3.37}$$

and, inside the common radius of convergence of the three opsgf's,

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) f_{\boldsymbol{\gamma},\boldsymbol{b}}(z) \,=\, f_{\boldsymbol{\gamma},\boldsymbol{c}}(z)\,. \tag{3.38}$$

Proof: Because of (3.35), we know that there is some finite $M > 0$ such that

$$\forall n \in \{0,1,\dots\}\,, \quad \text{we have} \quad |a_n| \,\le\, M^{n+1} \quad \text{and} \quad |b_n| \,\le\, M^{n+1}\,. \tag{3.39}$$

Therefore, using the triangle inequality and (3.36)

$$\forall n \in \{0,1,\dots\}\,, \quad \text{we have} \quad |c_n| \,\le\, (n+1)M^{n+1}\,. \tag{3.40}$$

But by Bernoulli's inequality, we know $(1+t)^n \ge 1+nt$ for all $t > -1$. Taking $t = 1$, we see that

$$\forall n \in \{0,1,\dots\}\,, \quad \text{we have} \quad |c_n| \,\le\, 2^n M^{n+1}\,. \tag{3.41}$$

Therefore, $\limsup_{n\to\infty} |c_n|^{1/n} \le 2M$.

Inside the common radius of convergence, we may rearrange the power series as we wish. (Recall that if a series were not absolutely convergent, but only conditionally convergent, then we could not rearrange the sums without potentially changing the limit.) But

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) f_{\boldsymbol{\gamma},\boldsymbol{b}}(z) \,=\, \sum_{k=0}^{\infty} a_k z^k \sum_{\ell=0}^{\infty} b_\ell z^\ell \,=\, \sum_{k,\ell=0}^{\infty} a_k b_\ell z^{k+\ell}\,. \tag{3.42}$$

Now rearrange the final double-sum series by changing variables from $(k,\ell) \in \{0,1,\dots\} \times \{0,1,\dots\}$ to $(m,k) \in \bigcup_{n=0}^{\infty} \{n\} \times \{0,\dots,n\}$ where $m = \ell + k$. □

In some classes one would call the sequence $\boldsymbol{c}$ the convolution of $\boldsymbol{a}$ and $\boldsymbol{b}$. In essence that is a main idea behind the definition of the ring of formal power series.

The requirement to have a finite radius of convergence for the generating function is overly restrictive if one is interested in combinatorics at its most basic levels. Consider for example the identity $|S_n| = n!$ for the cardinality of the set of all permutations of n objects. If we made the generating function for $\boldsymbol{a}$ with $a_k = k!$ or for $\boldsymbol{b}$ with $b_k = |S_k|!$ (and $b_0 = 1$) then unless we set $\gamma_k = 1/k!$ we do not necessarily get convergence. After all

$$\sum_{k=0}^{\infty} k! z^k \tag{3.43}$$

does not converge for any $z \neq 0$. Note that $k! \geq (k/2)^{k/2}$ by its definition and $(z^2 k/2)^{k/2}$ will diverge as $k \to \infty$ for every z other than 0. But the generating function for this or another rapidly growing combinatorial sequence could still be a useful tool. The formal power series is a series written as in equation (3.43) for example without any assumption of convergence. Rather, we just assume that we do know what all the coefficients are.

If we study the power series without assuming any type of convergence then they are called "formal power series." One way to think about them is to imagine this as a topic within algebra, where one often does not worry about convergence issues. There are facts about generating functions and more general power series that can be understood at this level. A good reference is Wilf's textbook [9]. That reference is also freely available online on the author's website:

`https://www2.math.upenn.edu/~wilf/gfology2.pdf`

Definition 2.2 *The* ring of formal power series *is the ring whose set consist of all opsgf's viewed as formal power series: for any $\boldsymbol{a} \in \mathbb{C}^{\{0,1,\dots\}}$, let $\boldsymbol{\gamma} \in (\mathbb{C} \setminus \{0\})^{\{0,1,\dots\}}$ be $\gamma_0 = \gamma_1 = \cdots = 1$, and take*

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) = \sum_{n=0}^{\infty} a_n z^n , \tag{3.44}$$

viewed as a formal power series. We drop the assumption of a positive radius of convergence. Given any second sequence $\boldsymbol{b} \in \mathbb{C}^{\{0,1,\dots\}}$, we define *the product of the two opsgf's,*

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) f_{\boldsymbol{\gamma},\boldsymbol{b}}(z) = f_{\boldsymbol{\gamma},\boldsymbol{c}}(z) , \tag{3.45}$$

where the two opsgf's are viewed as formal power series. The definition *of $\boldsymbol{c} \in \mathbb{C}^{\{0,1,\dots\}}$, is given by*

$$\forall n \in \{0, 1, \dots\} , \quad \text{we have} \quad c_n = \sum_{k=0}^{n} a_k b_{n-k} . \tag{3.46}$$

In the ring of formal power series one may ask the usual questions, such as

"for which formal power series may one find a multiplicative inverse?"

One may also ask whether for a pair of given formal power series if it makes sense to take their composition, inside the ring of formal power series. These questions and others are considered in Wilf [9]. We just refer to that free online textbook for the interested student.

2.1 Example: the number of elements in discrete simplices

For $n \in \{0, 1, \dots\}$ and for $d \in \mathbb{N}$, let us define

$$\Sigma(n, d-1) \;=\; \big\{(k_1, \dots, k_d) \in \{0, 1, \dots\}^d \;:\; k_1 + \cdots + k_d = n\big\}\,, \tag{3.47}$$

which is an analogue of a $(d-1)$-dimensional simplex, but discrete. The step-size may be thought of as $1/n$ so that in the n-to-∞ limit one would recover the $(d-1)$-dimensional simplex in $\mathbb{R}^d$

$$\Sigma^{(d-1)} \;=\; \big\{(x_1, \dots, x_d) \in \mathbb{R}^d \;:\; 0 \le x_1, \dots, x_d \le 1 \;\text{ and }\; x_1 + \cdots + x_d = 1\big\}\,. \tag{3.48}$$

Simplices such as this are important. For example, they form the set of all d-point probability mass functions in probability theory. They also form good topological manifolds homeomorphic to closed disks of all dimensions, which will be useful when we consider Brouwer's fixed point theorem in the graph theory portion of class. Let us note that

$$\forall n \in \{0, 1, \dots\}\,,\; \forall d \in \mathbb{N}\,, \quad \text{we have } \; |\Sigma(n, d-1)| \;=\; \binom{n+d-1}{d-1}\,. \tag{3.49}$$

The way to see this is just to note that

$$\sum_{k_1=0}^{\infty} \cdots \sum_{k_d=0}^{\infty} z^{k_1+\cdots+k_d} \;=\; \sum_{n=0}^{\infty} |\Sigma(n, d-1)| z^n \;=\; \left(\sum_{k=0}^{\infty} z^k\right)^d \;=\; \frac{1}{(1-z)^d}\,, \tag{3.50}$$

and then to use Newton's version of the binomial theorem. Recall that $(-1)^n\binom{-d}{n}$ is equal to $\binom{n+d-1}{d-1}$. Note that this would give yet another version of the type of identity which would be useful in a Calculus class for proving that (in the $n \to \infty$ limit) we get $\int_0^1 x^{d-1}\, dx \;=\; 1/d$ although the argument is a bit longer.

3 The Catalan numbers

One way that generating functions are useful is to help guess identities. For many combinatorial identities, if the identity is true then you can try to prove the identity using proof-by-induction, although the proof may be technical and difficult if induction is not the simplest approach. But it is often difficult to guess what form an identity may take to begin with.

Given two numbers $i, j \in \mathbb{N} = \{1, 2, \dots\}$ let us define $W_{i,j}^{(n)}(\mathbb{N})$ to be the set of all vectors $(x_0, \dots, x_n) \in \mathbb{N}^{n+1}$ satisfying

- $x_0 = i$,
- $\forall t \in \{1, \dots, n\}$ we have $x_t \in \{x_{t-1}+1, x_{t-1}-1\} \cap \mathbb{N}$,
- $x_n = j$.

These are the set of walks in $\mathbb{N}$ with n steps such that the walk starts at i and ends at j. For $n \in \mathbb{N}$ let

$$a_n \;=\; \left|W_{1,1}^{(n)}(\mathbb{N})\right|\,, \tag{3.51}$$

which is the number of all n-step walks on $\mathbb{N}$ that begin and end at 1. By parity, we know that

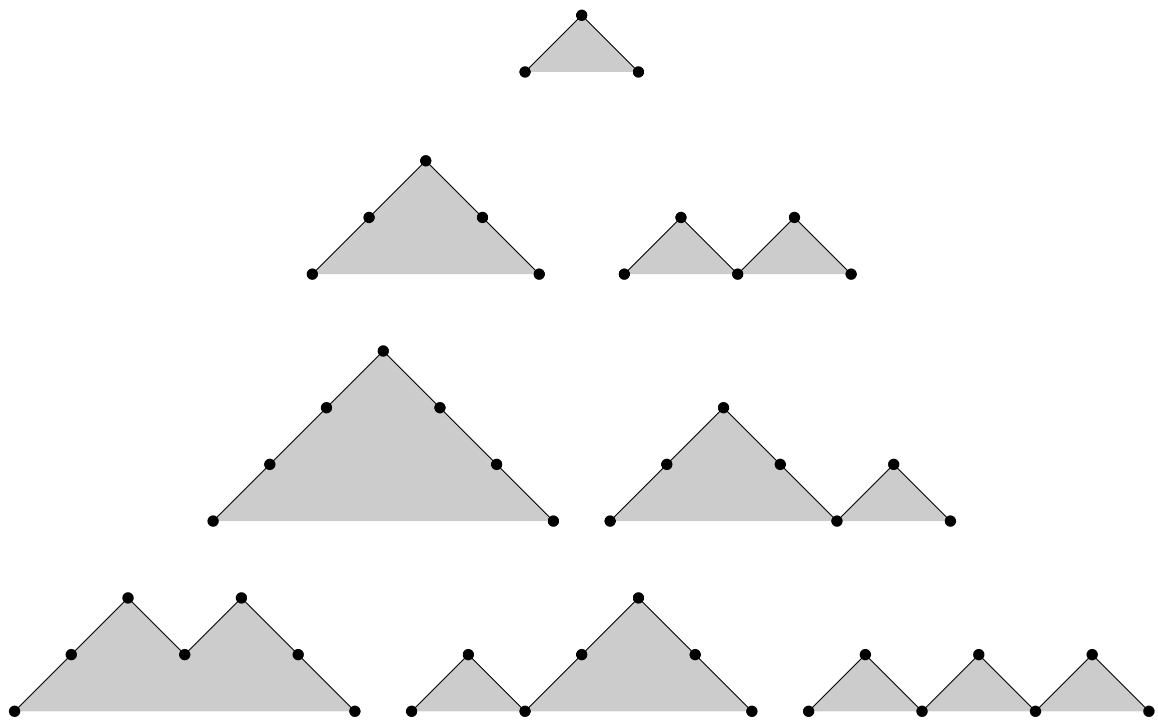


Figure 3.1: These are the walks in $\mathbb{N}$ of length $2n$ beginning and ending at the left endpoint of $\mathbb{N}$, namely 1. In the first row we have the single walk in $W_{1,1}^{(2)}(\mathbb{N})$. In the second row we have the two walks in $W_{1,1}^{(4)}(\mathbb{N})$. In the last two rows we have the 5 distinct walks in $W_{1,1}^{(6)}(\mathbb{N})$.

$a_n = 0$ unless $n \in \{0, 2, 4, \dots\}$. By convention, let $a_0 = 1$. In Figure 3.1, we have depicted the walks in $W_{1,1}^{(n)}(\mathbb{N})$ for the first few cases $n = 2$, $n = 4$ and $n = 6$.

Theorem 3.1 *With the definitions made above, $|W_{1,1}^{(n)}(\mathbb{N})|$ is Catalan's sequence for even n's:*

$$\forall n \in \{0, 1, \dots\}, \quad \textit{we have} \quad \left|W_{1,1}^{(n)}(\mathbb{N})\right| \;=\; \begin{cases} \frac{2}{n+2}\binom{n}{n/2} & \textit{if } n \in \{0, 2, 4, \dots\}, \textit{ and} \\ 0 & \textit{if } n \in \{1, 3, 5, \dots\}. \end{cases} \tag{3.52}$$

Proof: The way we will prove this is to first let $\boldsymbol{\gamma}$ be such that $\gamma_0 = \gamma_1 = \cdots = 1$ and take the (ordinary power series) generating function

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \;=\; 1 + \sum_{k=1}^{\infty} \left|W_{1,1}^{(2k)}(\mathbb{N})\right| z^{2k} \,. \tag{3.53}$$

Now we will use some combinatorial reasoning. Let us define

$$b_n \;=\; \left|\left\{(x_0, \dots, x_n) \in W_{1,1}^{(n)}(\mathbb{N}) \,:\, \forall t \in \{1, \dots, n-1\}\,,\; x_t > 1\right\}\right| \,. \tag{3.54}$$

So that is the set of all walks that first return to 1 at time n, itself. So $b_0 = a_0 = 1$, practically by convention. But for $n \in \{2, 4, 6, \dots\}$, we have that the first step of a walk enumerated by b_n is up and the last step is down. But $(x_1, \dots, x_{n-1})$ are all in $\{2, 3, \dots\}$. So, subtracting 1 from each point, we have $(x_1 - 1, \dots, x_{n-1} - 1)$ is still a walk in $\{1, 2, \dots\}$, and it begins and ends at 1. So

$$b_0 \;=\; 1 \quad \text{and} \quad \forall k \in \mathbb{N}\,, \quad \text{we have} \quad b_{2k} \;=\; a_{2k-2} \,. \tag{3.55}$$

Now consider a walk in $W_{1,1}^{(n)}(\mathbb{N})$, for $n > 0$. Let k be the smallest index in $\{1, \dots, n\}$ such that $x_k = 1$. Then $(x_0, \dots, x_k)$ is one of the walks enumerated by b_k. If $k < n$ then $(x_k, \dots, x_n)$ is a walk in $W_{1,1}^{(n-k)}(\mathbb{N})$. Even if $k = n$, since we set $a_0 = 1$ we may think of the single number (x_n) (which is (1)) as an element of $W_{1,1}^{(0)}(\mathbb{N})$. Because of this, we have

$$\forall k \in \mathbb{N}\,, \quad \text{we have} \quad a_{2k} \;=\; \sum_{j=1}^{k} b_{2j} a_{2k-2j} \,. \tag{3.56}$$

This is essentially the formula for the convolution. But now, using equation (3.55), we may rewrite this

$$\forall k \in \mathbb{N}\,, \quad \text{we have} \quad a_{2k} \;=\; \sum_{j=1}^{k} a_{2j-2} a_{2k-2j} \,. \tag{3.57}$$

Now let us incorporate this into the generating function

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \;=\; 1 + \sum_{k=1}^{\infty} a_{2k}\, z^{2k} \;=\; 1 + \sum_{k=1}^{\infty} \sum_{j=1}^{k} a_{2j-2} a_{2k-2j}\, z^{2k} \,. \tag{3.58}$$

So, we may rewrite and rearrange the sum as

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \,=\, 1+\sum_{k=1}^{\infty}\sum_{j=1}^{k} a_{2j-2}z^{2j-2}a_{2k-2j}\,z^{2k-2j}z^2 \,=\, 1+z^2\sum_{i=0}^{\infty}\sum_{\ell=0}^{\infty} a_{2i}z^{2i}a_{2\ell}\,z^{2\ell}\,, \tag{3.59}$$

where $i=j-1$ and $\ell=k-j$. But then we can split the two series on the right-most side,

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \,=\, 1+z^2\left(f_{\boldsymbol{\gamma},\boldsymbol{a}}(z)\right)^2\,. \tag{3.60}$$

This describes a quadratic equation for the generating function, whose algebraic solution is one of the two possibilities $\left(1\pm\sqrt{1-4z^2}\right)/(2z^2)$. But since $f_{\boldsymbol{\gamma},\boldsymbol{a}}(z)$ must converge to 1 at $z=0$, we conclude that we must take the $-$ in front of the square-root, not $+$. Then by L'Hospital's rule, or Taylor expansion, we see that the limit at 0 is correct. We are in the realm of using analysis, now, since we took limits. But it is easy to see that $|W_{1,1}^{(n)}(\mathbb{N})|\leq 2^n$, because there are n steps which are each either stepping up by $+1$ or stepping back by -1. So $R=1/2$ at least. According to the formula we have derived above R is equal to $1/2$:

$$\forall z\in\mathbb{C}\,,\quad \text{if } |z|<1/2, \text{ then }\ f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \,=\, \frac{1-\sqrt{1-4z^2}}{2z^2}\,. \tag{3.61}$$

Then we can calculate all the Taylor coefficients by using Newton's version of the binomial formula as in Section 5 of the pdf file "Some Notes on Basic Combinatorics." So that shows the formula. □

The underlying idea used in this proof appears in another famous problem: derivation of the Wigner semi-circle law for the spectral distribution function for the large-n limit of eigenvalues in the bulk of random matrix theories such as the GOE and GUE. That problem is actually equivalent to this one, but various references would need to be made to show that. We will consider the equivalence more in homework assignments or else as a potential final group project.

Note that there is another derivation of Theorem 3.1, which is arguably simpler. Let $\mathcal{W}_1^{(n)}(\mathbb{N})$ note all the walks of length 1 in $\mathbb{N}$ starting at 1 and ending anywhere in $\mathbb{N}$. Let $\mathcal{W}_{0,0}^{(n)}(\mathbb{Z})$ denote all the walks in $\mathbb{Z}$ starting and ending at 0. It is easy to see that

$$\left|\mathcal{W}_{0,0}^{(n)}(\mathbb{Z})\right| \,=\, \begin{cases}\binom{n}{n/2} & \text{if } n\in\{0,2,4,\dots\}, \text{ and}\\ 0 & \text{if } n\in\{1,3,5,\dots\}.\end{cases} \tag{3.62}$$

Moreover, by a geometric argument of Ed Nelson reported in Feller [4] one can see that

$$\left|\mathcal{W}_1^{(2n)}(\mathbb{N})\right| \,=\, \left|\mathcal{W}_{0,0}^{(2n)}(\mathbb{Z})\right|\,. \tag{3.63}$$

A great place to see the figure is in Kim Border's class lecture notes for his class at Caltech

`http://www.math.caltech.edu/~2016-17/2term/ma003/Notes/Lecture16.pdf`

Especially see figures 16.6 to 16.11 to see a mapping forward, as well as the reverse mapping backwards, to see that equation (3.63) is true.

But then it is easy to see that

$$\left|\mathcal{W}_1^{(2n+2)}(\mathbb{N})\right| = 4\left(\left|\mathcal{W}_1^{(2n)}(\mathbb{N})\right| - \left|\mathcal{W}_{1,1}^{(2n)}(\mathbb{N})\right|\right) + 2\left|\mathcal{W}_{1,1}^{(2n)}(\mathbb{N})\right| . \tag{3.64}$$

In other words,

$$\frac{1}{2^{2n+2}}\left|\mathcal{W}_1^{(2n+2)}(\mathbb{N})\right| = \frac{1}{2^{2n}}\left|\mathcal{W}_1^{(2n)}(\mathbb{N})\right| - \frac{1}{2^{2n+1}}\left|\mathcal{W}_{1,1}^{(2n)}(\mathbb{N})\right| . \tag{3.65}$$

But that means

$$\frac{1}{2^{2n+1}}\left|\mathcal{W}_{1,1}^{(2n)}(\mathbb{N})\right| = \frac{1}{2^{2n}}\left|\mathcal{W}_1^{(2n)}(\mathbb{N})\right| - \frac{1}{2^{2n+2}}\left|\mathcal{W}_1^{(2n+2)}(\mathbb{N})\right| , \tag{3.66}$$

which is

$$\frac{1}{2^{2n}}\binom{2n}{n} - \frac{1}{2^{2n+2}}\binom{2n+2}{n+1} . \tag{3.67}$$

But by writing the binomial coefficients in terms of the factorials and using recursion, this equals

$$\frac{1}{2^{2n}}\binom{2n}{n}\left(1 - \frac{2n+1}{2n+2}\right) . \tag{3.68}$$

But then this does imply that

$$\left|\mathcal{W}_{1,1}^{(2n)}(\mathbb{N})\right| = \frac{1}{n+1}\binom{2n}{n} . \tag{3.69}$$

It seems to be the case that the original formula may have been derived by the ballot problem, which is a problem related to all of this. But what is true, and what shows up in this important example, is that it is often tricky to first identify a combinatorial identity. Once the identity is known by one technique, there may be other proofs which independently re-derive the result.

We will state another type of example by Ira Gessel, where he derived a formula by one technique. Then, once the formula was known, other derivations could be obtained by other techniques, for an odd-time generalization of an identity of Chung and Feller for the number of walks with a prescribed number of positive midpoints.

4 The snake oil method

A good reference for the snake oil method is Wilf's book, Section 4.3. It is a method for calculating and proving identities involving combinatorial sums. For example, let the Fibonacci numbers be defined as the sequence F_0, F_1, F_2, etc. Define it inductively as

$$F_0 = 1 , \; F_1 = 1 , \quad \text{and for each } n \in \mathbb{N}, \text{ let } \; F_{n+1} = F_n + F_{n-1} . \tag{3.70}$$

With this, we can prove the following formula for a particular sum of binomial coefficients.

Theorem 4.1 *For each $n \in \{0, 1, \ldots\}$, we have*

$$F_n = \sum_{k=0}^{n}\binom{n-k}{k} , \tag{3.71}$$

where, as usual, the summand $\binom{n-k}{k}$ is equal to 0 if $k > n - k$ (i.e., if $2k > n$).

As an example, for $n = 4$ the non-zero terms in the summation are

$$\binom{4}{0} + \binom{3}{1} + \binom{2}{2} \;=\; 1 + 3 + 1 \;=\; 5\,. \tag{3.72}$$

And it is true that F_4 equals 5: $F_2 = 1 + 1 = 2$ and $F_3 = 2 + 1 = 3$ so F_4 is $2 + 3$. Before describing the snake-oil method, let us use simple recursion to prove the following.

Lemma 4.2 *Let $\boldsymbol{a}$ be such that $a_k = F_k$ for every $k \in \{0, 1, \dots\}$ and let $\boldsymbol{\gamma}$ be $\gamma_0 = \gamma_1 = \cdots = 1$. Then the ordinary power series generating function is*

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \;=\; \frac{1}{1 - z - z^2} \tag{3.73}$$

for $z < (\sqrt{5} - 1)/2$.

Proof of Lemma 4.2: Note that we can prove $F_n \le 2^n$ by induction, since $F_0 = 1 = 2^0$ and $F_1 = 1 < 2^1$ and for the induction step $F_{n+1} = F_n + F_{n-1} \le 2^n + 2^{n-1} \le 2^{n+1}$. So the radius of convergence is at least $R \ge 1/2$. Then, using the induction formula

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \;=\; 1 + z + \sum_{n=2}^{\infty} F_n z^n \;=\; 1 + z + \sum_{n=2}^{\infty} (F_{n-1} + F_{n-2}) z^n\,. \tag{3.74}$$

Rearranging the sums (inside the radius of convergence) gives

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \;=\; 1 + z + \sum_{m=1}^{\infty} F_m z^{m+1} + \sum_{\ell=0}^{\infty} F_\ell z^{\ell+1}\,, \tag{3.75}$$

writing $m = n - 1$ and $\ell = n - 2$. So this proves

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \;=\; 1 + z + z(f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) - 1) + z^2 f_{\boldsymbol{\gamma},\boldsymbol{a}}(z)\,. \tag{3.76}$$

Solving gives equation (3.73) within the radius of convergence. Since the power series has all nonnegative coefficients, the radius of convergence will be the smallest R such that the power series diverges for some number greater than R. This is related to a well-known result, Pringsheim's theorem. So R is the smallest root of $1 - z - z^2$ which is $(\sqrt{5} - 1)/2$ (which of course is greater than $1/2$). $\square$

Here is the snake-oil method, writing in our own words the algorithm stated in Wilf.

1. Consider the identity to be $a_n = b_n$ where a_n is a sum of combinatorial terms, for some range of values of n such as $n \in \{0, 1, \dots\}$.
2. Make the ordinary power series generating function $f_{\boldsymbol{\gamma},\boldsymbol{a}}(z)$, taking $\gamma_0 = \gamma_1 = \cdots = 1$.
3. Calculate the ordinary power series generating function $f_{\boldsymbol{\gamma},\boldsymbol{b}}(z)$, taking $\gamma_0 = \gamma_1 = \cdots = 1$. This requires being able to calculate the generating function for that sequence, for example using the geometric series identity, or Newton's version of the binomial formula.

4. For $f_{\boldsymbol{\gamma},\boldsymbol{a}}(z)$, use the sum formula for a_n in order to write the series as a double-sum.
5. Change the order of summation for the two sums, hopefully being able to directly evaluate the inner-summation, and then to identify the final summation as $f_{\boldsymbol{\gamma},\boldsymbol{b}}(z)$.

Proof of Theorem 4.1: Let $b_n = F_n$ and $\gamma_0 = \gamma_1 = \dots$ so that

$$f_{\boldsymbol{\gamma},\boldsymbol{b}}(z) \;=\; \frac{1}{1-z-z^2}\,, \tag{3.77}$$

for $|z| < (\sqrt{5}-1)/2$ by Lemma 4.2. Let

$$a_n \;=\; \sum_{k=0}^{n} \binom{n-k}{k}\,. \tag{3.78}$$

So the ordinary power series generating function is

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \;=\; \sum_{n=0}^{\infty} z^n \sum_{k=0}^{n} \binom{n-k}{k}\,. \tag{3.79}$$

In order to interchange the order of summation, let $m = n-k$ so that $n = m+k$. If k satisfies $k \leq n$, then n satisfies $n \geq k$. So m satisfies $m \geq 0$. Thus we get

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \;=\; \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} z^{m+k}\binom{m}{k}\,. \tag{3.80}$$

Actually now it helps to change the order of summation again:

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \;=\; \sum_{m=0}^{\infty} z^m \sum_{k=0}^{\infty} z^k \binom{m}{k}\,. \tag{3.81}$$

Then the inner summation is $(1+z)^m$. Therefore, we obtain

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \;=\; \sum_{m=0}^{\infty} z^m (1+z)^m\,. \tag{3.82}$$

Changing the variable to $w = z(1+z)$ we obtain the geometric series with w as a variable, so that does give equation (3.77). $\square$

Note that in this proof we did switch the order of summation back. But we could have avoided that by using Newton's version of the binomial formula to deduce that

$$\sum_{m=k}^{\infty} z^m \binom{m}{k} \;=\; \frac{z^k}{(1-z)^{k+1}}\,, \tag{3.83}$$

for example because $\binom{k+\ell}{k} = (-1)^\ell \binom{-k-1}{\ell}$. So, if we start at equation (3.80), then we get

$$f_{\boldsymbol{\gamma},\boldsymbol{a}}(z) \;=\; \frac{1}{1-z}\sum_{k=0}^{\infty} z^k \frac{z^k}{(1-z)^k} \;=\; \frac{1}{1-z}\sum_{k=0}^{\infty}\left(\frac{z^2}{1-z}\right) \;=\; \frac{1}{1-z}\cdot\frac{1}{1-\left(\frac{z^2}{1-z}\right)}\,. \tag{3.84}$$

That does also give equation (3.77), after algebraic simplification.

5 A recent result by Gessel and independently Grünbaum

Consider walks $\mathcal{W}_0^n(\mathbb{Z})$. Given an element $(x_0, \dots, x_n)$, let the midpoints be

$$\left(\frac{x_0 + x_1}{2}, \frac{x_1 + x_2}{2}, \dots, \frac{x_{n-1} + x_n}{2}\right). \tag{3.85}$$

In [2], Chung and Feller considered the combinatorial problem of counting the number of walks in $\mathcal{W}_0^n(\mathbb{Z})$ with a prescribed number of positive midpoints, but only for the case that n was even.

Note that a walk in $\mathcal{W}_0^n(\mathbb{Z})$ has some points at 0. So the points of the walks itself $\{x_0, \dots, x_n\}$ cannot be perfectly partitioned into positive points and negative points (since some are 0). But what is true is that if t is an even element of $\mathbb{N}$ then x_n is an even element of $\mathbb{Z}$, and vice-versa. Therefore, for each t, we do know that $x_t + x_{t-1}$ is odd. So it is not zero. Therefore, the midpoints can be perfectly partitioned into positive midpoints and negative midpoints. The number $(x_t + x_{t-1})/2$ is the midpoint of x_{t-1} and x_t and sits directly between the two in a graph of the walk using linear interpolation to graph edges between points.

Chung and Feller were motivated by trying to generalize a continuous theorem of Paul Lèvy for Brownian motion [5], known as the arcsine law (actually several results). But to get their combinatorial result they changed the perspective from positive points to positive midpoints. In many subjects, once a result is known in the continuous setting, researchers often find that discovering combinatorial analogues poses a greater challenge (although potentially bypassing some difficulties of continuous analysis).

The proof of this occupies a chapter in each of the editions of Feller, with improvements in the pedagogy of the presentation seeming to motivate Feller to write new editions. An excellent reference is Kim Borders's notes. The formula is

$$\left|\left\{(x_0, \dots, x_{2k}) \in \mathcal{W}_0^{2k}(\mathbb{Z}) \,:\, |\{t \in \{1, \dots, 2k\} : (x_t + x_{t-1}) > 0\}| = 2j\right\}\right| = \binom{2j}{j}\binom{2k-2j}{k-j}. \tag{3.86}$$

A simple proof is provided in McKean's textbook page 141 [7].

What is interesting is that no generalization to odd values of n was accomplished until the 21st century, despite the fact that the Chung, Feller theorem is a fundamental result for random walks. Ira Gessel gave a formula using generating functions in

`https://people.brandeis.edu/~gessel/homepage/slides/chung-feller-slides.pdf`

His formula is that

$$\begin{aligned}&\left|\left\{(x_0, \dots, x_{2k+1}) \in \mathcal{W}_0^{2k+1}(\mathbb{Z}) \,:\, |\{t \in \{1, \dots, 2k+1\} : (x_t + x_{t-1}) > 0\}| = 2j+1\right\}\right| \\ &\qquad = \binom{2j}{j}\binom{2k-2j}{k-j} \cdot \frac{2j+1}{k+1},\end{aligned} \tag{3.87}$$

and the number of walks whose positive midpoints number $2k + 1 - (2j + 1)$ is the same as the number of walks whose positive midpoints number $2j + 1$, by symmetry.

Once this formula is known, other proofs, not using generating functions, may be used. Independently, Alberto Grünbaum re-derived the formula by using Feynman-Kac, a tool best understood in the context of probability theory (or quantum theory) [6]. It may also be proved directly using induction. (That may appear as a potential final group project.) For combinatorial identities such as this, one must first understand the formula, by some method. Induction does

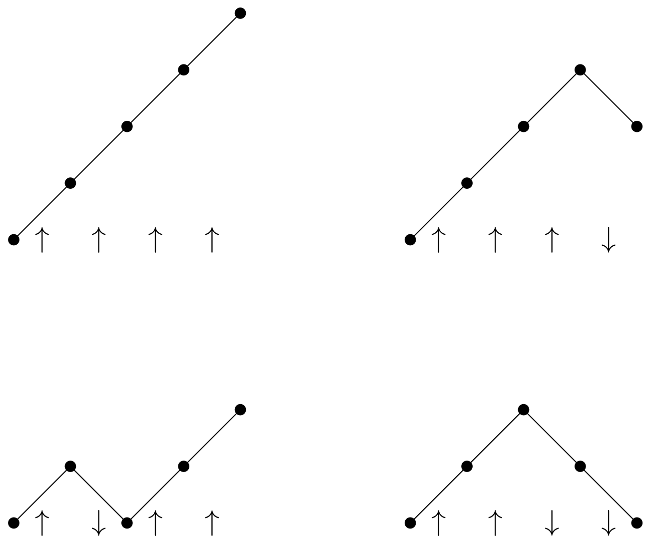


Figure 3.2: All walks of length $2k = 4$ with $2j = 4$ midpoints positive. The equation of Chung and Feller (3.86) gives the number of these as $\binom{4}{2}\binom{0}{0}$ which is 6. We also indicate whether each step is a step up or a step down using $\uparrow$ and $\downarrow$ which is a shorter notation.

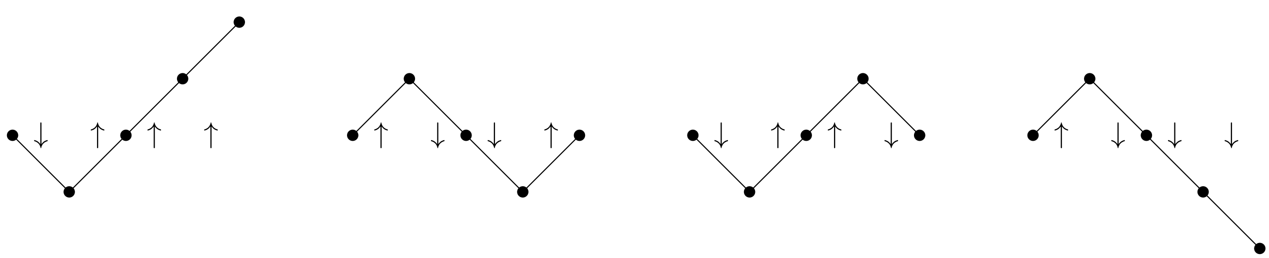


Figure 3.3: All walks of length $2k = 4$ with $2j = 2$ midpoints positive. The equation of Chung and Feller (3.86) gives the number of these as $\binom{2}{1}\binom{2}{1}$ which is 4. We also indicate whether each step is a step up or a step down using $\uparrow$ and $\downarrow$ which is a shorter notation.

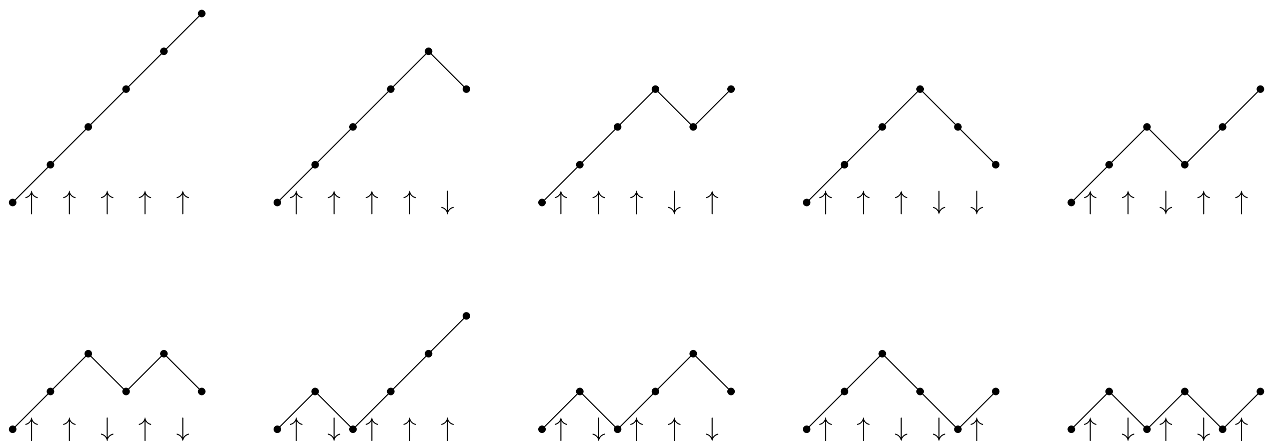


Figure 3.4: All walks of length $2k+1=5$ with $2j+1=5$ midpoints positive. The equation of Gessel and of Grünbaum (3.87) gives the number of these as $\binom{4}{2}\binom{0}{0}\cdot\frac{5}{3}$ which is 10.

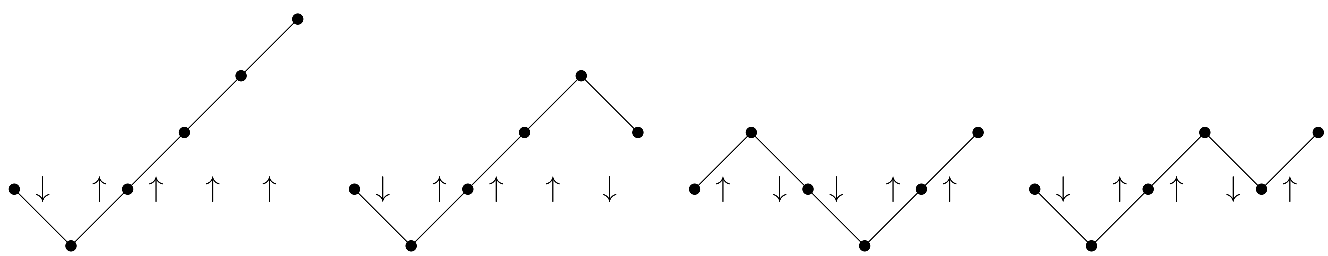


Figure 3.5: All walks of length $2k+1=5$ with $2j+1=3$ midpoints positive. The equation of Gessel and of Grünbaum (3.87) gives the number of these as $\binom{2}{1}\binom{2}{1}\cdot\frac{3}{3}$ which is 4.

not usually help to guess a formula. So induction is not a good method to derive the simple form of the right-hand-side in (3.87). Presumably that is why it took so many years for Chung and Feller's theorem for even times to be generalized to odd times.

6 Generating functions and connections beyond combinatorics

An important topic in combinatorics is the counting of integer partitions. Let $\mathcal{P}_n$ denote the set

$$\mathcal{P}_n \,=\, \left\{\boldsymbol{\lambda} = (\lambda_1, \lambda_2 \dots) \in \{0,1,\dots\}^{\mathbb{N}} \,:\, \lambda_1 \geq \lambda_2 \geq \lambda_2 \geq \dots \ \text{ and } \lambda_1 + \lambda_2 + \dots = n\right\}. \tag{3.88}$$

Because of the conditions any element of $\mathcal{P}_n$ will be a sequence $\boldsymbol{\lambda} = (\lambda_1, \lambda_2 \dots)$ such that for some ℓ we have $\lambda_{\ell+1} = \lambda_{\ell+2} = \dots = 0$. We let

$$\ell(\boldsymbol{\lambda}) \,=\, \min(\{k \in \{0,1,\dots\} \,:\, \lambda_{k+1} = 0\})\,. \tag{3.89}$$

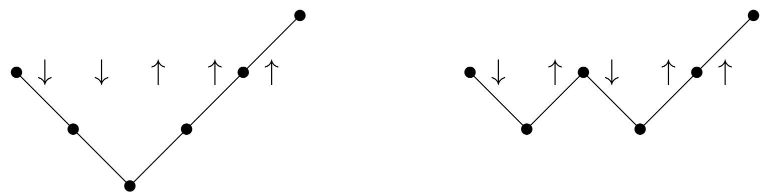


Figure 3.6: All walks of length $2k+1=5$ with $2j+1=1$ midpoints positive. The equation of Gessel and of Grünbaum (3.87) gives the number of these as $\binom{0}{0}\binom{4}{2}\cdot\frac{1}{3}$ which is 2.

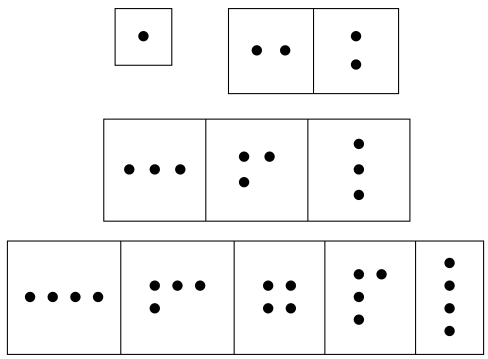


Figure 3.7: These are the Young diagrams for the partitions in $\mathrm{Par}(n)$ for $n=1,2,3,4$. With $p_n=|\,\mathrm{Par}(n)|$, we have $p_1=1$, $p_2=2$, $p_3=3$ and $p_4=5$. But there is no known simple formula for p_n in general.

Then for $\ell=\ell(\lambda)$ we have

$$\lambda_1,\dots,\lambda_\ell \in \mathbb{N}=\{1,2,\dots\}\,, \qquad \lambda_1\geq\lambda_2\geq\cdots\geq\lambda_\ell \quad \text{and} \quad \lambda_1+\cdots+\lambda_\ell=n\,. \tag{3.90}$$

Given a partition $\boldsymbol{\lambda}$, the Young diagram is a pictorial representation. Draw a dot at each index (i,j) such that $j\leq\ell$ and $i\leq\lambda_j$. In the American version we put the indices in matrix format, with the i coordinates going across from left-to-right, and the j coordinates going vertically from top to bottom. In the French version the j coordinates go up, as in the positive quadrant of $\mathbb{R}^2$. Both have benefits. We will use the American version. See Figure 3.7 for some examples.

Another version of partitions may be described as follows. For each $k\in\mathbb{N}$, let $\nu_k:\mathcal{P}_n\to\{0,1,\dots\}$ be defined as

$$\forall\boldsymbol{\lambda}\in\mathcal{P}_n\,,\quad \text{we have}\ \ \nu_k(\boldsymbol{\lambda})\;=\;\big|\{j\in\mathbb{N}\,:\,\lambda_j=k\}\big|\,. \tag{3.91}$$

We define $\mathcal{N}_n$ to be the set of all $\boldsymbol{a}=(a_1,a_2,\dots)\in\{0,1,\dots\}^{\mathbb{N}}$ satisfying

$$\sum_{k=1}^{\infty} k\cdot a_k\;=\;n\,. \tag{3.92}$$

We can define a mapping $\mathcal{T}_n : \mathcal{P}_n \to \mathcal{N}_n$ such that

$$\forall \boldsymbol{\lambda} \in \mathcal{P}_n\,, \quad \text{we have} \quad \mathcal{T}_n(\boldsymbol{\lambda}) \;=\; (\nu_1(\boldsymbol{\lambda}), \nu_2(\boldsymbol{\lambda}), \dots)\,. \tag{3.93}$$

Then $\mathcal{T}_n$ is a bijection. Indeed, we may construct the inverse.

Given any $\boldsymbol{a} \in \mathcal{N}_n$ let $\mathcal{L}(\boldsymbol{a})$ be equal to $\sum_{k=1}^{\infty} a_k$. For any $\boldsymbol{a} \in \mathcal{N}_n$, the number $\mathcal{L}(\boldsymbol{a})$ is finite and is bounded above by n, since n equals $\sum_{k=1}^{\infty} k \cdot a_k$. Then, for each $j \in \mathbb{N}$, we define $\Lambda_j : \mathcal{N}_n \to \{0, 1, \dots\}$ by the formula

$$\forall \boldsymbol{a} \in \mathcal{N}_n\,, \quad \text{we let} \quad \Lambda_j(\boldsymbol{a}) \;=\; \begin{cases} 1 & \text{if } j \le a_1, \\ 2 & \text{if } a_1 < j \le a_1 + a_2, \\ \vdots & \\ k & \text{if } \sum_{i=1}^{k-1} a_i < j \le \sum_{i=1}^{k} a_i, \\ \vdots & \\ 0 & \text{if } j > \mathcal{L}(\boldsymbol{a}). \end{cases} \tag{3.94}$$

Then the mapping from $\mathcal{N}_n$ to $\mathcal{P}_n$ taking $\boldsymbol{a}$ to $(\Lambda_1(\boldsymbol{a}), \Lambda_2(\boldsymbol{a}), \dots)$ is the inverse of $\mathcal{T}_n$.

It is common to consider the partitions to just consist of the positive parts, not including the infinite tail of $\lambda_{\ell+1}$, $\lambda_{\ell+2}$, ... which are all zero.

Definition 6.1 *Given $n \in \mathbb{N}$, let us define* $\mathrm{Par}(n)$ *to be a collection (set) of all ordered tuples*

$$(\lambda_1, \dots, \lambda_\ell) \in \mathbb{N}^\ell\,, \quad \textit{such that:} \quad \ell \in \mathbb{N}\,, \quad \lambda_1 \ge \cdots \ge \lambda_\ell \quad \textit{and} \quad \lambda_1 + \cdots + \lambda_\ell = n\,. \tag{3.95}$$

We extend the definition to $n \in \{0, 1, \dots\}$ by defining $\mathrm{Par}(0)$ *to be the singleton set consisting of* $()$ *where ℓ is interpreted as* 0 *for this "empty partition."*

One writes partitions in $\mathrm{Par}(n)$ as just λ, so that

$$\lambda \;=\; (\lambda_1, \dots, \lambda_\ell)\,. \tag{3.96}$$

Also, in the literature sometimes one denotes the relation $\lambda \in \mathrm{Par}(n)$ instead as $\lambda \vdash n$.

The partition number of n is p_n where

$$p_n \;=\; |\,\mathrm{Par}(n)|\,. \tag{3.97}$$

In particular, by our convention,

$$p_0 \;=\; |\,\mathrm{Par}(n)| \;=\; \big|\{()\}\big| \;=\; 1\,. \tag{3.98}$$

Then, denoting $\boldsymbol{p}$ to equal $(p_0, p_1, \dots)$ and taking $\boldsymbol{\gamma}$ such that $\gamma_0 = \gamma_1 = \cdots = 1$, the opsgf for the partition numbers is

$$f_{\boldsymbol{\gamma},\boldsymbol{p}}(z) \;=\; \sum_{n=0}^{\infty} p_n z^n\,. \tag{3.99}$$

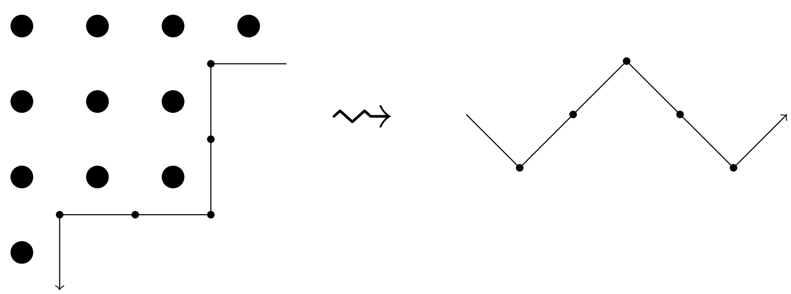


Figure 3.8: In this example, start with the Young diagram for $\lambda = (4,3,3,1)$. The 0^{th} step is not show, it is a down-step, starting from the top of the last dot in the first row. Then after that, let each step be a down-step or a left-step, following the outer boundary of the Young diagram. The 0^{th} step is not shown, and also the last step is not shown because it is always a left-step. After rotating by 135° we get a standard walk on $\mathbb{Z}$.

An elementary fact is that

$$\forall m \in \{0,1,\dots\}\,, \quad \text{we have} \ \sum_{n=1}^{\infty} \left|\left\{\lambda \in \mathrm{Par}(n)\,:\, \lambda_1 + \ell(\lambda) = m+2\right\}\right| \,=\, 2^m\,. \tag{3.100}$$

One way to see this is consider the Young diagrams for each partition $\lambda \in \mathrm{Par}(n)$, and then to draw a path following a portion of the outer boundary of the Young diagram. See Figure 3.8. The path may be easily transformed to be an element of $\mathcal{W}_0^{(m)}(\mathbb{Z})$, the set of all walks of length $n-1$ on $\mathbb{Z}$ starting at 0. (Rotate by 45° clockwise in the French convention or by 135° in the American convention.) Moreover, if one considers all partitions satisfying the conditions, it gives every path in $\mathcal{W}_0^{(m)}(\mathbb{Z})$. Since $|\mathcal{W}_0^{(m)}(\mathbb{Z})| = 2^m$, that gives the identity.

Note that $\lambda_1 + \ell(\lambda) \leq n+1$, with the the inequality obtained as an equality only for the special cases of $\lambda \in \{(n)\,,\ (n-1,1)\,,\ (n-2,1,1)\,,\ \dots\,,\ (n-j,1^j)\,,\ \dots\,,\ (2,1^{n-2})\,,\ (1^n)\}$. Here, we have used a notation which is common in the literature, combining pieces of the description from both $\mathcal{P}_n$ and $\mathcal{N}_n$. Therefore, if we define the sets $\mathcal{Q}_0, \mathcal{Q}_1, \dots$ such that

$$\forall m \in \{0,1,\dots\}\,, \quad \text{we have} \ \ \mathcal{Q}_m \,=\, \bigcup_{n=1}^{\infty} \left\{\lambda \in \mathrm{Par}(n)\,:\, \lambda_1 + \ell(\lambda) = m+2\right\}, \tag{3.101}$$

then it follows that

$$\forall n \in \mathbb{N}\,, \quad \text{we have} \ \ \mathrm{Par}(n) \,\subseteq\, \bigcup_{m=0}^{n-1} \mathcal{Q}_m\,. \tag{3.102}$$

Hence, we have the bound

$$\forall n \in \mathbb{N}\,, \quad \text{we have} \ \ p_n \,\leq\, 2^{n-1} + 2^{n-2} + \cdots + 1 \,=\, 2^n - 1\,. \tag{3.103}$$

So, the radius of convergence for the opsgf satisfies $R \geq 1/2$.

Theorem 6.2 *For the opsgf of the partition numbers, $f_{\boldsymbol{\gamma},\boldsymbol{p}}(z) = \sum_{n=0}^{\infty} p_n z^n$, we have*

$$\forall z \in \mathbb{C} \ \ \textit{such that } |z| < R\,, \qquad f_{\boldsymbol{\gamma},\boldsymbol{p}}(z) \ = \ \prod_{n=1}^{\infty} \frac{1}{1-z^n}\,. \tag{3.104}$$

Proof: Inside the radius of convergence we are justified in rearranging the series. Start with

$$f_{\boldsymbol{\gamma},\boldsymbol{p}}(z) \ = \ \sum_{n=0}^{\infty} p_n z^n \ = \ \sum_{n=0}^{\infty} \sum_{\boldsymbol{\lambda} \in \mathcal{P}_n} z^n\,. \tag{3.105}$$

Then this may also be written as

$$f_{\boldsymbol{\gamma},\boldsymbol{p}}(z) \ = \ \sum_{n=0}^{\infty} \sum_{\boldsymbol{a} \in \mathcal{N}_n} z^{\sum_{k=1}^{\infty} k \cdot a_k}\,, \tag{3.106}$$

and by rearrangeability, this may be written as the sum of $z^{1\cdot a_1 + 2\cdot a_2 + \cdots}$ summed over the union of all the $\mathcal{N}_n$ sets. Note that

$$\bigcup_{n=0}^{\infty} \mathcal{N}_n \ = \ \big\{\boldsymbol{a} = (a_1, a_2, \dots) \in \{0,1,\dots\}^{\mathbb{N}} \ : \ |\{k : a_k > 0\}| < \infty\big\}\,. \tag{3.107}$$

Let us define a new set $\mathcal{N}^{(r)}$ to be $\mathcal{N}^{(0)} = \{(0,0,\dots)\}$ and

$$\forall r \in \mathbb{N}\,, \quad \mathcal{N}^{(r)} \ = \ \big\{\boldsymbol{a} = (a_1, a_2, \dots) \in \{0,1,\dots\}^{\mathbb{N}} \ : \ a_r > 0 \ \text{ and } (\forall s > r \text{ we have } a_r = 0)\big\}\,. \tag{3.108}$$

Then we can see that $\bigcup_{n=0}^{\infty} \mathcal{N}_n = \bigcup_{r=0}^{\infty} \mathcal{N}^{(r)}$ and the second union is a disjoint union. Also,

$$\forall r \in \mathbb{N}\,, \quad \text{we have } \ \mathcal{N}^{(r)} \ = \ \big\{(a_1, \dots, a_r, 0, 0, \dots) \ : \ a_1, \dots, a_{r-1} \in \{0,1,\dots\}\,, \ a_r \in \mathbb{N}\big\}\,. \tag{3.109}$$

So, from this formula,

$$\sum_{\boldsymbol{a} \in \mathcal{N}^{(r)}} z^{1\cdot a_1 + \cdots + r \cdot a_r} \ = \ \left(\prod_{k=1}^{r-1} \left(\sum_{a_k=0}^{\infty} (z^k)^{a_k}\right)\right) \cdot \sum_{a_r=1}^{\infty} (z^r)^{a_r}\,. \tag{3.110}$$

But this equals $\prod_{k=1}^{r-1}(1-z^k)^{-1}$ times the factor $\big((1-z^r)^{-1} - 1\big)$. So we have a telescoping sum,

$$f_{\boldsymbol{\gamma},\boldsymbol{p}}(z) \ = \ \sum_{r=0}^{\infty} \sum_{\boldsymbol{a} \in \mathcal{N}^{(r)}} z^{1\cdot a_1 + \cdots + r\cdot a_r} \ = \ z^0 + \sum_{r=1}^{\infty} \left(\prod_{k=1}^{r} \frac{1}{1-z^k} - \prod_{k=1}^{r-1} \frac{1}{1-z^k}\right)\,, \tag{3.111}$$

where we separated out the $r=0$ term accounted for by the summand $z^0 = 1$. Also, the empty product (in the case that $r = 1$ so $r - 1 = 0$) is equal to 1 as well. The formula follows. □

We used $R \geq 1/2$ to get a non-trivial result in the theorem. But actually, a consequence of the formula in (3.104) is that the radius of convergence, R, is actually equal to

$$R = 1. \tag{3.112}$$

For a positive power series, the radius of convergence will be equal to the supremum of the set of all $t \in [0, \infty)$ such that the series is convergent for all $x \in [0, t]$. This is closely related to Pringsheim's theorem. But we can see that for $x < 1$ we have

$$\prod_{k=1}^{n} \frac{1}{1-x^k} = \exp\left(-\sum_{k=1}^{n} \ln(1-x^k)\right), \tag{3.113}$$

and an easy inequality is $-\ln(1-y) = \int_0^y (1-t)^{-1}\, dt$ is bounded by $y/(1-y)$. So

$$-\sum_{k=1}^{n} \ln(1-x^k) \leq \sum_{k=1}^{n} \frac{x^k}{1-x^k} \leq \frac{\sum_{k=1}^{n} x^k}{1-x}, \tag{3.114}$$

because, for all $k \in \mathbb{N}$, we have $1/(1-x) > 1/(1-x^k)$. So, from this, for every $x \in [0, 1)$,

$$\forall n \in \mathbb{N}, \quad \text{we have} \quad \prod_{k=1}^{n} \frac{1}{1-x^k} \leq \exp\left(1/(1-x)^2\right), \tag{3.115}$$

proving that the positive power series converges (to the supremum of its partial sums) for that value of x. But clearly the formula in equation (3.104) does not converge at $z = 1$ because even the first factor diverges: $\lim_{x\to 1^-} 1/(1-x) = +\infty$ as an extended real number.

Theorem 6.3 (Hardy and Ramanujan's preliminary upper bound) *For each n in $\{0, 1, \dots\}$, we have*

$$p_n \leq e^{\pi\sqrt{2n/3}}. \tag{3.116}$$

Proof: Using Cauchy's integral formula, or Fourier series, or just the fact that $\int_0^{2\pi} e^{inx}\, dx$ is equal to 2π if $n = 0$ and is equal to 0 if $n \in \mathbb{Z} \setminus \{0\}$, and rearrangements, we may see that

$$\forall r \in (0, 1), \ \forall n \in \{0, 1, \dots\}, \quad \text{we have} \quad p_n = \frac{1}{2\pi r^n} \sum_{k=0}^{\infty} p_k r^k \int_0^{2\pi} e^{ikx} e^{-inx}\, dx, \tag{3.117}$$

and this may be rewritten (interchanging the sum and the integral inside the radius of convergence)

$$\forall r \in (0, 1), \ \forall n \in \{0, 1, \dots\}, \quad \text{we have} \quad p_n = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{r^n e^{inx}} \prod_{k=1}^{\infty} \frac{1}{(1 - r^k e^{ikx})}\, dx. \tag{3.118}$$

By the triangle inequality the right hand side is bounded by $r^{-n}\prod_{k=1}^{\infty}(1-r^k)^{-1}$, assuming $r \in (0,1)$. Then we note, trying $r = \exp(-\alpha)$,

$$\forall \alpha > 0\,, \quad \text{we have} \quad r^{-n}\prod_{k=1}^{\infty}(1-r^k)^{-1}\bigg|_{r=e^{-\alpha}} \;=\; e^{\alpha n - \sum_{k=1}^{\infty}\ln(1-e^{-\alpha k})}\,. \tag{3.119}$$

The integral comparison test shows that this is bounded above by $e^{\alpha n - \alpha^{-1}\int_0^\infty \ln(1-e^{-t})\,dt}$ (see for example Figure 3.9) where we have used $t_k = \alpha k$ and $\Delta t = \alpha$. Alternatively, we compared a sum on k to an integral on κ and used

$$\int_0^\infty \ln(1-e^{-\alpha\kappa})\,d\kappa \;=\; \frac{1}{\alpha}\int_0^\infty \ln(1-e^{-t})\,dt\,, \tag{3.120}$$

for $\alpha > 0$. But a famous exact formula is

$$-\int_0^\infty \ln(1-e^{-t})\,dt \;=\; \frac{\pi^2}{6}\,, \tag{3.121}$$

which may be verified with a symbolic computer program such as Mathematica. So, optimizing over α in the upper bound

$$e^{\alpha n - \alpha^{-1}\int_0^\infty \ln(1-e^{-t})\,dt} \;=\; \exp\left(\alpha n + \frac{\pi^2}{6\alpha}\right)\,, \tag{3.122}$$

we get the $\alpha^* = \pi/\sqrt{6n}$ (by taking the derivative of the exponent with respect to α, and setting it equal to 0). Substituting that in gives the upper bound $e^{2\pi\sqrt{n/6}}$ which is the same as $e^{\pi\sqrt{2n/3}}$. $\square$

A demonstration of the integral comparison test is shown in Figure 3.9. For the equation (3.121), note that positivity of all terms justifies the Fubini-Tonelli theorem

$$-\int_0^\infty \ln(1-e^{-t})\,dt \;=\; \int_0^\infty \sum_{n=1}^{\infty}\frac{e^{-nt}}{n}\,dx \;=\; \sum_{n=1}^{\infty}\frac{1}{n}\int_0^\infty e^{-nt}\,dt \;=\; \sum_{n=1}^{\infty}\frac{1}{n^2}\,. \tag{3.123}$$

The calculation of the Riemann zeta function at 2, $\zeta(2) = \sum_{n=1}^{\infty} n^{-2}$, is a well-known exact calculation:

$$\sum_{n=1}^{\infty}\frac{1}{n^2} \;=\; \frac{\pi^2}{6}\,. \tag{3.124}$$

Now let us begin to explain the difficulty posed in Hardy and Ramanujan's complex analysis approach.

Firstly, let us define a function $\psi_n : \mathbb{C} \to \mathbb{C}$ defined by the formula

$$\psi_n(z) \;=\; z^{-n}\,f_{\boldsymbol{\gamma},\boldsymbol{p}}(z)\,. \tag{3.125}$$

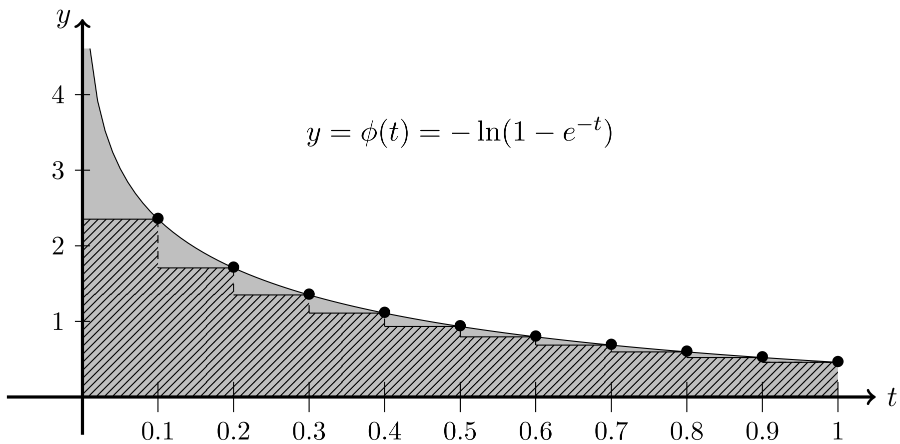


Figure 3.9: A visual representation of the integral comparison test for $\alpha = 0.1$. The area under the graph of $y = \phi(t) = -\ln(1-e^{-t})$ is shown in gray $\int_0^\infty \phi(t)\,dt$. Since the function is decreasing, that area is an upper bound to the area which is shaded with the diagonal-line-pattern. That area is the sum $\sum_{k=1}^\infty \phi(t_k)\Delta t$ which is $-\alpha \sum_{k=1}^\infty \ln(1-e^{-\alpha k})$.

Let us suppose that r_n^* is the choice of $r \in [0,1)$ which maximizes $\psi_n(r)$. We note that this is not quite exactly approximated by $\exp(-\pi\sqrt{n/6})$ as we used before.

Theorem 6.4 (G. N. Watson) *Defining the q-Pochhammer symbol (generalizing the rising factorial version of the Pochhammer symbol)*

$$\forall q \in \mathbb{C}\,, \quad \textit{such that} \quad |q| < 1\,, \quad \textit{we have} \quad (q;q)_\infty = \prod_{k=1}^{\infty}(1-q^k)\,, \tag{3.126}$$

its asymptotics for $q \to 1$ is as follows:

$$(q;q)_\infty \sim \sqrt{\frac{2\pi}{\ln(1/q)}}\,\exp\left(\frac{\pi^2}{6\ln(q)}\right)\,q^{-1/24}\,, \quad \textit{as } q \to 1. \tag{3.127}$$

The original reference for this result is Watson [8] who was an expert at special functions (and one of the first caretakers of Ramanujan's notebooks). Another good reference is [1] which seems to also be available on the arXiv at

`https://arxiv.org/abs/1602.01085`

The factor $q^{-1/24}$ converges to 1 when $q \to 1$. But Watson includes it probably because it indicates deeper properties, such as Ramanujan's discovery/invention of mock theta functions, related to the q-Pochhammer symbol, since his definitions do require that extra factor $q^{-1/24}$.

Alternatively, it seems to be the first correction beyond the order-1 behavior in the asymptotic expansion.

Note that, writing $q = \exp(-\alpha)$ for a small number $\alpha > 0$, we have

$$\ln\left((q;q)_\infty\right)\Big|_{q=\exp(-\alpha)} = \sum_{k=1}^{\infty} \ln\left(1 - e^{-\alpha k}\right) . \tag{3.128}$$

A good approximation is

$$\int_{1/2}^{\infty} \ln\left(1 - e^{-\alpha\kappa}\right)\, d\kappa \, = \, -\frac{\pi^2}{6\alpha} - \int_0^{1/2} \ln\left(1 - e^{-\alpha\kappa}\right)\, d\kappa\, . \tag{3.129}$$

Approximating $\ln(1 - e^{-\alpha\kappa})$ by $\ln(\alpha\kappa)$, we see that this quantity equals

$$-\frac{\pi^2}{6\alpha} - \frac{1}{2}\ln(\alpha) + \frac{1}{2} + \frac{1}{2}\ln(2) + O(\alpha\ln(\alpha))\, , \quad \text{as } \alpha \to 0. \tag{3.130}$$

However the approximation is not perfect. Instead the remainder is

$$\sum_{k=1}^{\infty} \ln\left(1 - e^{-\alpha k}\right) - \int_{1/2}^{\infty} \ln\left(1 - e^{-\alpha\kappa}\right)\, d\kappa \, = \, \sum_{k=1}^{\infty} \int_{k-(1/2)}^{k+(1/2)} \Big(\ln\left(1 - e^{-\alpha\kappa}\right) - \ln\left(1 - e^{-\alpha k}\right)\Big)\, d\kappa\, . \tag{3.131}$$

Using the fundamental theorem of calculus, this equals

$$\sum_{k=1}^{\infty} \int_{k-(1/2)}^{k+(1/2)} \left(\int_k^{\kappa} \frac{\alpha e^{-\alpha\kappa_1}}{1 - e^{-\alpha\kappa_1}}\, d\kappa_1\right) d\kappa\, . \tag{3.132}$$

Note that if we subtract off $\alpha e^{-\alpha\kappa}/(1-e^{-\alpha\kappa})$ from the inner integrand then we will get a constant times $(\kappa - k)$ as the extra term. But the integral of $(\kappa - k)$, integrated over $(k - \frac{1}{2}, k + \frac{1}{2})$ equals 0. So this is okay: subtracting that quantity from the inner integrand does not actually change the quantity. So we may rewrite it as

$$\sum_{k=1}^{\infty} \int_{k-(1/2)}^{k+(1/2)} \left(\int_k^{\kappa} \left(\frac{\alpha e^{-\alpha\kappa_1}}{1 - e^{-\alpha\kappa_1}} - \frac{\alpha e^{-\alpha\kappa}}{1 - e^{-\alpha\kappa}}\right) d\kappa_1\right) d\kappa\, . \tag{3.133}$$

Then, by the Fundamental theorem of calculus again, this equals

$$-\alpha^2 \sum_{k=1}^{\infty} \int_{k-(1/2)}^{k+(1/2)} \left(\int_k^{\kappa} \left(\int_k^{\kappa_1} \frac{e^{-\alpha\kappa_2}}{(1 - e^{-\alpha\kappa_2})^2}\, d\kappa_2\right) d\kappa_1\right) d\kappa\, . \tag{3.134}$$

Using the Fubini-Tonelli theorem, this may be rewritten as

$$-\frac{\alpha^2}{2} \sum_{k=1}^{\infty} \int_{k-(1/2)}^{k+(1/2)} \left(\frac{1}{2} - |\kappa_2 - k|\right)^2 \frac{e^{-\alpha\kappa_2}}{(1 - e^{-\alpha\kappa_2})^2}\, d\kappa_2\, . \tag{3.135}$$

We could approximate this by

$$-\frac{\alpha^2}{2} \int_{1/2}^{\infty} \frac{e^{-\alpha\kappa_2}}{(1 - e^{-\alpha\kappa_2})^2}\, d\kappa_2 \int_{-1/2}^{1/2} \left(\frac{1}{2} - |t|\right)^2 dt \tag{3.136}$$

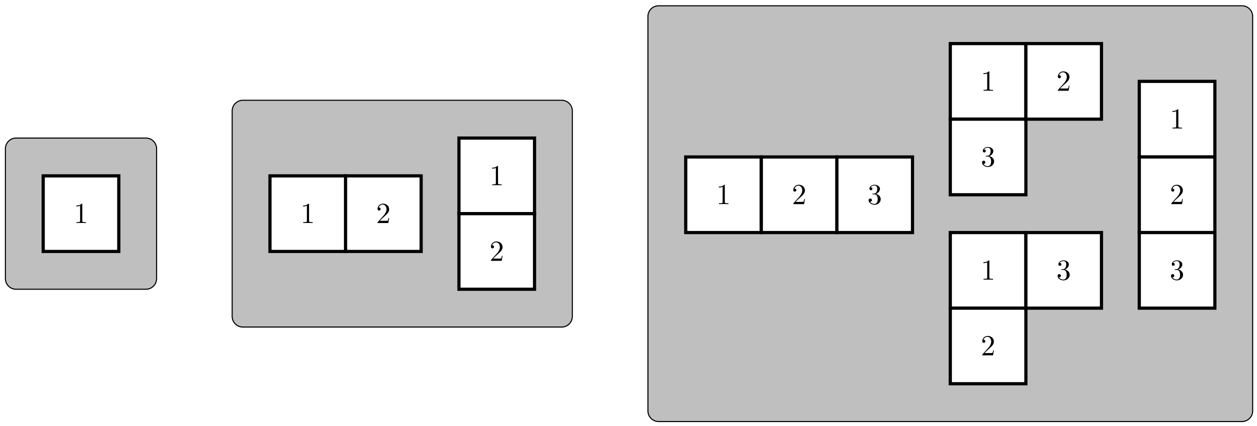


Figure 3.10: We have shown several examples of standard Young tableaux grouped according to n which their partition corresponds to.

which is

$$-\frac{\alpha^2}{24} \cdot \frac{1}{\alpha(e^{\alpha/2}-1)} \;\to\; \frac{1}{12}\,, \quad \text{as } \alpha \to 0. \tag{3.137}$$

Let us skip the proof of this easy lemma (which nevertheless requires a fair number of calculations to do). The main point is that we may use the approximation

$$-\sum_{k=1}^{\infty} \ln\left(1-e^{-\alpha k}\right) \;=\; -\int_{1/2}^{\infty} \ln\left(1-e^{-\alpha\kappa}\right)\, d\kappa + \frac{1}{12} + O\left(\frac{1}{\alpha}\right)\,, \tag{3.138}$$

which may be derived by applying a finite number of steps in the Euler-Maclaurin summation formula.

Note that we have essentially calculated the asymptotic limit

$$\lim_{n\to\infty} e^{-\pi\sqrt{2n/3}} \left(r^{-n} f_{\boldsymbol{\gamma},\boldsymbol{p}}(r)\Big|_{r=\exp(-\pi/\sqrt{6n})} \right) \;=\; 1\,. \tag{3.139}$$

of the dilogarithm and q-Pochhammer symbol. But it may be part of a potential final group project.)

The only drawback is that the length of the vector defining a partition, ℓ, may vary if one changes the partition. One frequently wants to understand something about the length of the partition, itself $\ell(\lambda)$. Let us define $\mathcal{P}_{n,\ell}$ to be the set of partitions of length ℓ:

$$\mathcal{P}_{n,\ell} \;=\; \{\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots) \in \mathcal{P}_n \;:\; \min(\{k : \lambda_{k+1} = 0\}) = \ell\}\,. \tag{3.140}$$

Indeed, in a different context, the problem was much in vogue. context of representation theory on the symmetric group, it became a major problem to understand the length of a partition, albeit one not chosen uniformly among all possible partitions, but rather a partition chosen uniformly among all pairs of identical partitions counted according to the number of ways to fill in the boxes of their Young tableaux in a standard way (increasing in both rows and columns).

That is important in representation theory because it gives the Haar measure to the space of representation/irreducible characters in the decomposition of the left-regular-representation of S_n. It also corresponds to a type of problem which is still in vogue in the field of mathematical statistical mechanics.

It is a difficult problem to determine the quantities p_n explicitly. It is even difficult to determine their asymptotic behavior.

Bibliography

[1] Subho Bannerjee and Blake Wilkerson. Asymptotic expansions of Lambert series and related q-series. *Int. J. Number Theory* **13**, no. 8, 2097–2113 (2017).

[2] Kai-Lai Chung and William Feller. On fluctuations in coin-tossing. *Proc. Nat. Acad. Sci.* **35**, 605–608 (1949).

[3] Herbert B.Enderton. *Elements of Set Theory, 1st Edition.* Academic Press, San Diego, CA 1977.

[4] William Feller. *An Introduction to Probability Theory and Its Applications, Volume 1, 3rd Edition.* John Wiley & Sons, USA 1968.

[5] Paul Lévy. Sur certains processus stochastiques homogènes. Compositio Mathematica **7**, 283–339 (1940).

[6] Alberto Grünbaum. A Feynman-Kac approach to a paper of Chung and Feller on fluctuations in the coin-tossing game. *Proc. Amer. Math. Soc. to appear.* `https://doi.org/10.1090/proc/14758`

[7] Henry McKean. *Probability: The Classical Limit Theorems.* Cambridge University Press, Cambridge, UK 2014.

[8] G. N. Watson. The Final Problem: An Account of the Mock Theta functions. *J. London Math. Soc.* **11**, 55–80 (1936).

[9] Herbert S. Wilf. *Generatingfunctionology.* Academic Press, Boston, MA 1990.