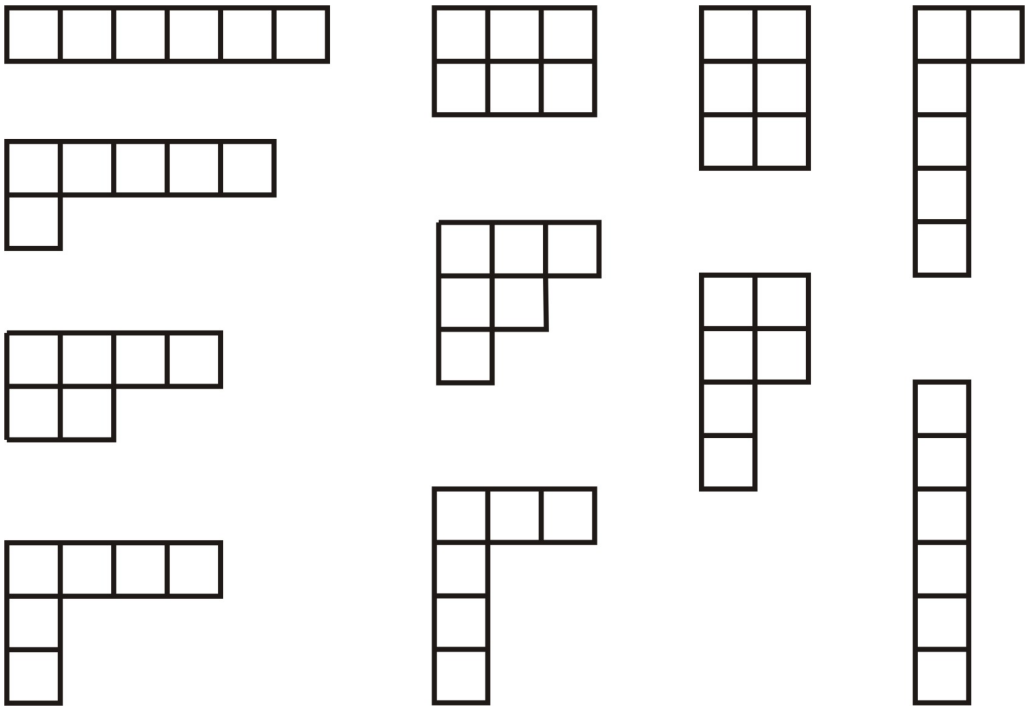
C&O 330

Introduction to

Combinatorial Enumeration

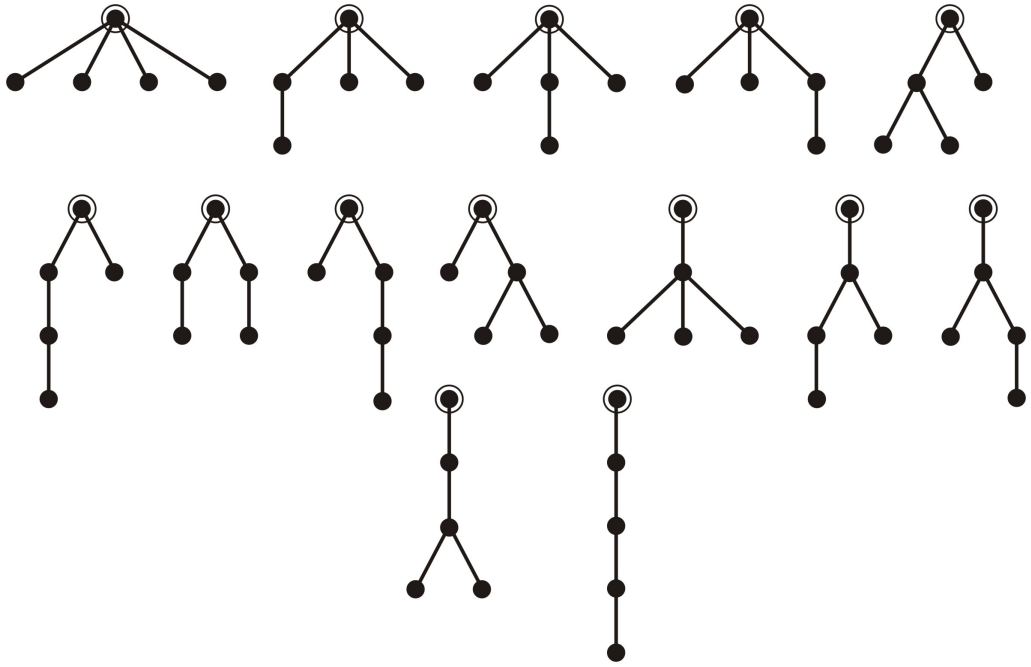


David G. Wagner

∞

for Kay

Contents.



0. Introduction.

These notes are an introduction to combinatorial enumeration intended specifically for the syllabus of C&O 330 at the University of Waterloo. Of course, in a 13 week semester one must be selective about the material presented and there is no possibility of treating the whole subject comprehensively. Additionally, care must be taken that the material begins at a level commensurate with the current knowledge of the students and does not ascend too rapidly. Therefore, I have chosen to concentrate on a few central topics. These are:
• bijective proofs of numerical identities and their q-analogues;
• the use of ordinary generating functions;
• recursive structure and the Lagrange Implicit Function Theorem;
• algebraic properties of formal power series;
• the theory of integer partitions;
• the use of exponential generating functions.
Usually, the explanation of these subjects has left me with between e and π days to play with in the semester.

For the gifted students (and to make writing these notes a more pleasant task for myself) there are a few topics sketched in the final chapters, most of which will never be covered in class during a normal semester. The foundations of the theory of exponential generating functions, following Joyal, and a combinatorial proof of the Lagrange Implicit Function Theorem, following Raney, are the most important (and really belong in C&O 430/630). Pólya's theory of enumeration and symmetry is hinted at, but I confess that these notes get no further than Burnside's Lemma. Conway's checker–jumping game thoroughly amused me as an undergraduate, and the temptation to share it with a new audience proved irresistable. The Matrix–Tree Theorem is a cornerstone of enumerative graph theory, and is one of my favourite facts. Building upon that is Kirchhoff's formula for the effective admittance of an electrical network, for which I give two proofs – one close to Kirchhoff's own (using Cramer's rule), and one which can be described as "semi–combinatorial" or "fractionally bijective".

Each chapter concludes with a list of exercises, and sometimes a list of endnotes including references to the literature for further reading. These exercises and endnotes are in places a bit thin, but then again these notes do not have to stand up to the level of nitpicking required for a book intended to be published for a wider audience. At least not yet they don't.

As a rough guide, Chapters 1 through 5 serve as a warm–up and introduction to the "philosophy" of the subject, Chapters 6 through 11 constitute the heart of the

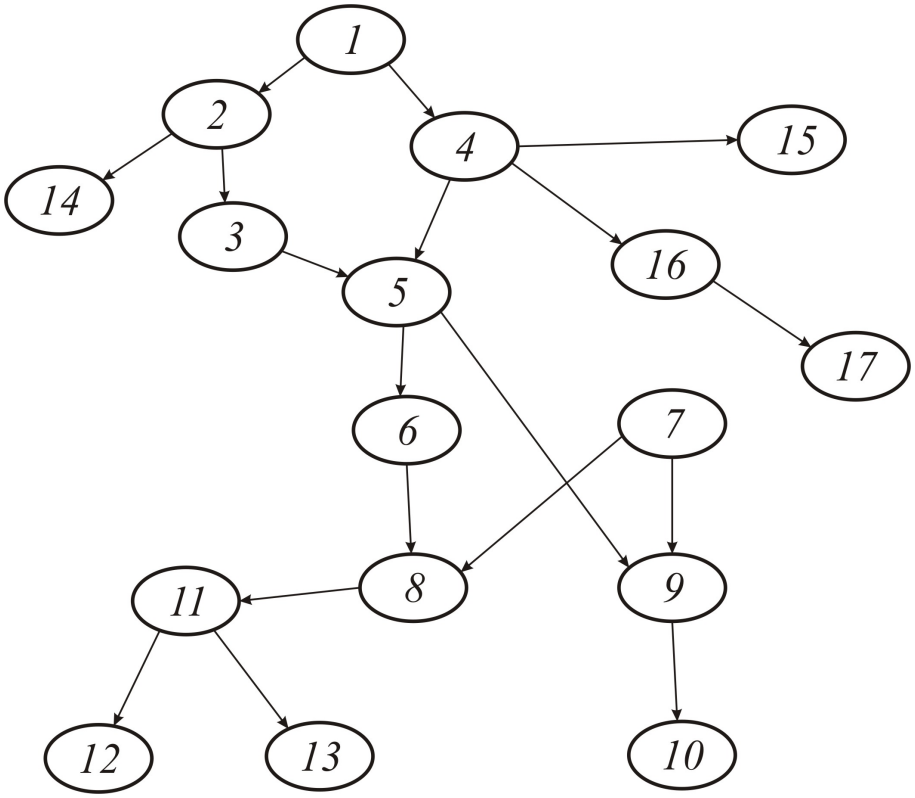


FIGURE 0.1. logical dependences among the chapters.

matter, and Chapters 12 through 17 hopefully provide a tasty smorgasbord to satisfy a curious student's healthy intellectual appetite. I am convinced that diagrams like Figure 0.1 are more amusing for the author than instructive to the reader, but I include one anyway. It is meant to indicate the main logical dependences among the chapters: an arrow from node i to node j means that Chapter i should be read before Chapter j. For example, after reading Chapters 1 and 2, a reader could skip directly to Chapter 14.

As far as prerequisites go, these are not extensive. The students in C&O 330 have two semesters of calculus (MATH 137, 138), two semesters of linear algebra (MATH 136, 235), one course on elementary abstract algebra (MATH 135), and one course (MATH 239) which introduces enumeration in six weeks (ordinary generating functions, compositions, rational binary languages, linear recurrence relations) and graph theory in seven weeks (isomorphism, connectedness, spanning trees, breadth–first–search, planarity, the 5–colour theorem, elementary matching theory, *et cetera*). We will not assume any familiarity with enumeration, but will from time to time require some familiarity with the language and basic results of calculus, algebra, and graph theory. Some of the constructions we describe by specifying an algorithm

in "pseudocode", but no specialized knowledge of topics in computer science is required.

We use the following more–or–less standard notational conventions.

$\mathbb{N}$	the natural numbers $\{0, 1, 2, \ldots\}$
$\mathbb{Z}$	the integers
$\mathbb{Q}$	the rational numbers
$\mathbb{R}$	the real numbers
$\mathbb{C}$	the complex numbers
$:=$	equality by definition
$[x^n]f(x)$	the coefficient of x^n in $f(x)$
LHS	left hand side
RHS	right hand side

Finally, I have many people to acknowledge for helping me to understand this material well enough to present it however well I have managed to do so. In particular, many thanks go to Brian Alspach, László Babai, Chris Godsil, Ian Goulden, David Jackson, and Richard Stanley, both as mentors and as colleagues.

1. Sets and Bijections.

"Combinatorial Enumeration" is all about counting things. In the most general terms, the type of problem we address is the following: having defined a finite set S of gadgets, we want to determine the cardinality (or size) of the set S. We will see many sophisticated techniques for solving several varieties of such problems, but at heart the most fundamental objects in the theory are exceedingly simple – finite sets and bijections. In this section we develop the basic principles governing finite sets and their cardinalities. Many of these principles will be familiar to you already, but we might as well begin at the beginning.

Let S and T be two sets. A function $f : S \to T$ is called a *surjection* if it satisfies the following condition:

- For every $t \in T$ there exists an $s \in S$ such that $f(s) = t$.

A function $f : S \to T$ is called an *injection* if it satisfies the following condition:

- For every $s, s' \in S$, if $f(s) = f(s')$ then $s = s'$.

(An older terminology – which is now out of fashion – is that a surjection is an "onto map" and an injection is a "one–to–one map".) A function which is both a surjection and an injection is called a *bijection* (or a "one–to–one and onto map"). If $f : S \to T$ is a bijection then for each $t \in T$ there is a unique $s \in S$ such that $f(s) = t$. In this case the *inverse function* $f^{-1} : T \to S$ is well–defined by saying that $f^{-1}(t) := s$ if and only if $f(s) = t$. Notice that f^{-1} is also a bijection, and that $(f^{-1})^{-1} = f$. In fact, a function $f : S \to T$ is a bijection **if and only if** f has an inverse function.

Two sets S and T are said to be *equicardinal* if and only if there is a bijection $f : S \to T$. Exercise 1.1 shows that this is an equivalence relation, which we denote by $S \rightleftharpoons T$. For each natural number $n \in \mathbb{N}$, let

$$N_n := \{1, \ldots, n\}.$$

In particular, $N_0 = \varnothing$. A set S is *finite* if it is equicardinal with N_n for some $n \in \mathbb{N}$, otherwise S is *infinite*. By Exercise 1.2, if S is finite then it is equicardinal with N_n for exactly one value of $n \in \mathbb{N}$. We say that S is a finite set of *cardinality* or *size* n, and write $\#S = n$ (or $|S| = n$) to denote this fact. For finite sets we may also use the expression

$$\#S = \sum_{s \in S} 1,$$

but for infinite sets the sum on the RHS does not converge. (In fact, the theory of cardinality for infinite sets is very interesting, but that's another story.)

Proposition 1.1. *Let S and T be finite sets. There is a bijection $f : S \to T$ if and only if $\#S = \#T$.*

Proof. Let $\#S = n$ and $\#T = m$, so that there are bijections $g : S \to N_n$ and $h : T \to N_m$. If there is a bijection $f : S \to T$ then, by Exercise 1.1, $h \circ f \circ g^{-1} : N_n \to N_m$ is a bijection, so that by Exercise 1.2 we conclude that $n = m$. Conversely, if $n = m$ then $h^{-1} \circ g : S \to T$ is a bijection, by Exercise 1.1. □

For any function $f : S \to T$ and element $t \in T$ we let

$$f^{-1}(t) := \{s \in S : f(s) = t\}$$

denote the *preimage of t under f*. This is the subset of elements of S which are mapped by f to the element t of T. Notice that f is a bijection if and only if $\#f^{-1}(t) = 1$ for all $t \in T$. (This notation is slightly at odds with the notation for inverse functions, but no confusion should arise.)

Proposition 1.2. *Let S and T be finite sets, and let $f : S \to T$ be any function. Then*

$$\#S = \sum_{t \in T} \#f^{-1}(t).$$

Proof.

$$\#S = \sum_{s \in S} 1 = \sum_{t \in T} \sum_{s \in f^{-1}(t)} 1 = \sum_{t \in T} \#f^{-1}(t).$$

□

Proposition 1.3. *Let S and T be finite sets, and let k be a positive integer. Suppose that $f : S \to T$ is a surjection such that for each $t \in T$, there are exactly k elements $s \in S$ such that $f(s) = t$. Then $\#S = k \cdot \#T$.*

Proof. Since $\#f^{-1}(t) = k$ for all $t \in T$, Proposition 1.2 implies that

$$\#S = \sum_{t \in T} k = k \sum_{t \in T} 1 = k \cdot \#T.$$

□

Recall that the *Cartesian product* of sets S and T is defined to be the set

$$S \times T := \{(s, t) : \ s \in S \text{ and } t \in T\}$$

of all ordered pairs with first coordinate from S and second coordinate from T.

Proposition 1.4. *Let S and T be finite sets. Then*

$$\#(S \times T) = (\#S) \cdot (\#T).$$

Proof. Consider the function $f : S \times T \to T$ given by $f(s, t) := t$ for each $(s, t) \in S \times T$. For every $t \in T$ we have $\#f^{-1}(t) = \#S$, so that Proposition 1.3 implies that $\#(S \times T) = (\#S) \cdot (\#T)$. □

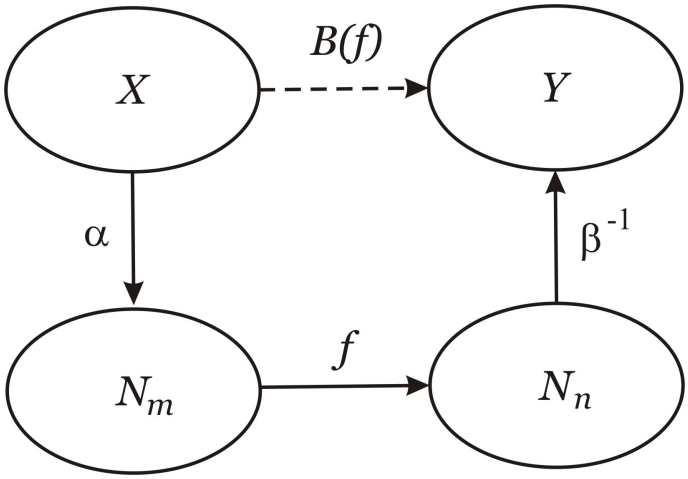


FIGURE 1.1. the bijection of Example 1.7.

Proposition 1.5. *Let S_1, S_2, ..., S_k be finitely many finite sets. Then*

$$\#(S_1 \times \cdots \times S_k) = (\#S_1) \cdots (\#S_k).$$

(The proof is left as an exercise.)

Example 1.6. Let $m, n \in \mathbb{N}$, and consider the set $\mathcal{F}(N_m, N_n)$ of all functions from N_m to N_n. What is $\#\mathcal{F}(N_m, N_n)$? A function $f : N_m \to N_n$ can be expressed alternatively as a sequence $(t_1, t_2, ..., t_m)$ in which $t_i := f(i) \in N_n$ for each $i \in N_m$. This construction describes a bijection $\mathcal{F}(N_m, N_n) \rightleftharpoons N_n \times \cdots \times N_n$ (with m factors), and by Proposition 1.5 we see that $\#\mathcal{F}(N_m, N_n) = (\#N_n)^m = n^m$.

Example 1.7. Let X and Y be two finite sets, with $\#X = m$ and $\#Y = n$, say. As in Example 1.6, let $\mathcal{F}(X, Y)$ be the set of all functions from X to Y. We claim that $\#\mathcal{F}(X, Y) = n^m$ as well. To prove this, it suffices to find a bijection $B : \mathcal{F}(N_m, N_n) \to \mathcal{F}(X, Y)$. Since X is equicardinal with N_m, there is a bijection $\alpha : X \to N_m$. Since Y is equicardinal with N_n, there is a bijection $\beta : Y \to N_n$. Now, for any function $f : N_m \to N_n$, we define $B(f)$ to be the composition of functions $B(f) := \beta^{-1} \circ f \circ \alpha$. (See Figure 1.1 for a schematic diagram of this situation.) Notice that $B(f)$ really is a function from X to Y. It is left as an exercise to show that $B : \mathcal{F}(N_m, N_n) \to \mathcal{F}(X, Y)$ is a bijection.

Example 1.8. Given a finite set V, let $\mathcal{P}(V)$ denote the set of all subsets of V. How many subsets does V have? That is, what is $\#\mathcal{P}(V)$? We may represent a subset $S \subseteq V$ by its *characteristic function* $f_S : V \to \{0, 1\}$, which is defined by

$$f_S(v) := \begin{cases} 1 & \text{if } v \in S, \\ 0 & \text{if } v \notin S, \end{cases}$$

for each $v \in V$. Conversely, given any function $f : V \to \{0, 1\}$ we may associate with it the subset $f^{-1}(1)$ of V. These two constructions $S \mapsto f_S$ and $f \mapsto f^{-1}(1)$ define mutually inverse bijections between the sets $\mathcal{P}(V)$ and $\mathcal{F}(V, \{0, 1\})$. (Verification of this claim is left as an exercise.) By applying Example 1.7, we see that $\#\mathcal{P}(V) = \#\mathcal{F}(V, \{0, 1\}) = 2^{\#V}$.

Example 1.9. We continue with Example 1.8, anticipating some of the ideas to be developed in Section 4. For a function $f : V \to \{0, 1\}$, let $|f| := \sum_{v \in V} f(v)$. Notice that in the bijection $\mathcal{P}(V) \rightleftharpoons \mathcal{F}(V, \{0, 1\})$, if the subset $S \subseteq V$ corresponds to the function $f : V \to \{0, 1\}$, then $\#S = |f|$. I like to display this kind of information in a little table, like this:

$$\begin{array}{rcl} \mathcal{P}(V) & \rightleftharpoons & \mathcal{F}(V, \{0, 1\}) \\ S & \leftrightarrow & f \\ \#S & = & |f| \end{array}$$

The first line names the two sets between which the bijection exists. The second line indicates that the bijection matches the item S on the left with the item f on the right. The third line indicates that when S corresponds to f we have $\#S = |f|$. Now let's introduce an indeterminate variable x, and consider the sum

$$\sum_{S \in \mathcal{P}(V)} x^{\#S}.$$

The terms in this summation are in bijective correspondence with the functions $f \in \mathcal{F}(V, \{0, 1\})$, so we may equivalently sum over these items instead. The term $x^{\#S}$ contributed by $S \in \mathcal{P}(V)$ is equal to the term $x^{|f|}$ contributed by the corresponding function $f \in \mathcal{F}(V, \{0, 1\})$. That is,

$$\sum_{S \in \mathcal{P}(V)} x^{\#S} = \sum_{f \in \mathcal{F}(V, \{0,1\})} x^{|f|}.$$

For any particular $v \in V$, the value $f(v)$ is either 0 or 1. Since $|f| = \sum_{v \in V} f(v)$, we may compute that

$$\sum_{f \in \mathcal{F}(V, \{0,1\})} x^{|f|} = \prod_{v \in V} \sum_{f(v)=0}^{1} x^{f(v)} = \prod_{v \in V} (1 + x) = (1 + x)^{\#V}.$$

In conclusion,

$$\sum_{S \subseteq V} x^{\#S} = (1 + x)^{\#V}.$$

Notice that both sides are polynomials in the indeterminate x, and that upon setting $x = 1$ we obtain the conclusion of Example 1.8. We could even introduce different

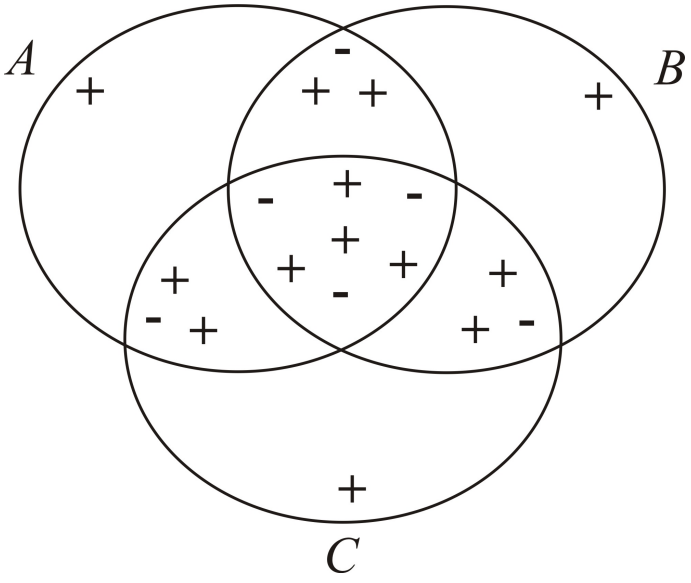


FIGURE 1.2. the union of three sets.

indeterminats x_v for each element $v \in V$. For a subset $S \subseteq V$, define $\mathbf{x}^S := \prod_{v \in S} x_v$. The bijection $\mathcal{P}(V) \rightleftharpoons \mathcal{F}(V, \{0,1\}$ we are discussing leads to the identity

$$\sum_{S \subseteq V} \mathbf{x}^S = \prod_{v \in V} (1 + x_v).$$

This will be used in Example 1.12.

Proposition 1.5 allows us to determine the cardinality of a Cartesian product of sets of known sizes. Determining the cardinality of a union of sets of known sizes is a bit more difficult, because the way in which the sets overlap is important. For example, for two sets we have

$$\#(A \cup B) = \#A + \#B - \#(A \cap B),$$

since the sum $\#A + \#B$ counts every element in $A \cap B$ exactly twice. Similarly, one can check that for three sets we have

$$\#(A \cup B \cup C) = \#A + \#B + \#C - \#(A \cap B) - \#(A \cap C) - \#(B \cap C) + \#(A \cap B \cap C).$$

See Figure 1.2 – the signs indicate contributions from each of the seven terms on the RHS above.

To generalize this to any (finite) number of sets A_1, A_2, ..., A_m we need to introduce some notation. For $\varnothing \neq S \subseteq N_m$, let

$$A_S := \bigcap_{i \in S} A_i.$$

So, for example, $A_{\{2,3,5\}} = A_2 \cap A_3 \cap A_5$. The following theorem is known as the *Principle of Inclusion/Exclusion.*

Theorem 1.10 (Inclusion/Exclusion)**.** *Let A_1, A_2, ..., A_m be finite sets. Then*

$$\#(A_1 \cup \cdots \cup A_m) = \sum_{\varnothing \neq S \subseteq N_m} (-1)^{\#S-1}(\#A_S).$$

Proof. Let $V := A_1 \cup \cdots \cup A_m$, and for each $v \in V$ let $T(v) := \{i \in N_m : \ v \in A_i\}$. Notice that $T(v) \neq \varnothing$, for all $v \in V$. Also notice that for $\varnothing \neq S \subseteq N_m$ we have $v \in A_S$ if and only if $\varnothing \neq S \subseteq T(v)$. Therefore

$$\sum_{\varnothing \neq S \subseteq N_m} (-1)^{\#S-1}(\#A_S) = \sum_{\varnothing \neq S \subseteq N_m} (-1)^{\#S-1} \sum_{v \in A_S} 1 = \sum_{v \in V} \sum_{\varnothing \neq S \subseteq T(v)} (-1)^{\#S-1}.$$

From Example 1.9 we have

$$1 + \sum_{\varnothing \neq S \subseteq T(v)} x^{\#S} = (1+x)^{\#T(v)}$$

for each $v \in V$. Upon setting $x = -1$ in this polynomial equation, we obtain

$$1 + \sum_{\varnothing \neq S \subseteq T(v)} (-1)^{\#S} = 0$$

since $\#T(v) \geq 1$. Therefore,

$$\sum_{\varnothing \neq S \subseteq N_m} (-1)^{\#S-1}(\#A_S) = \sum_{v \in V} \sum_{\varnothing \neq S \subseteq T(v)} (-1)^{\#S-1} = \sum_{v \in V} 1 = \#V,$$

as was to be shown. □

The case in which the sets A_i are pairwise disjoint is particularly simple, since then $A_S = \varnothing$ whenever $\#S \geq 2$.

Corollary 1.11. *Let $A_1, \ldots, A_m$ be finite sets. Assume that these sets are paiwise disjoint,* i.e. *that $A_i \cap A_j = \varnothing$ for all $1 \leq i < j \leq m$. Then*

$$\#(A_1 \cup \cdots \cup A_m) = \#A_1 + \cdots + \#A_m.$$

Example 1.12. For a positive integer n, the *Euler totient* of n is the number $\varphi(n)$ of integers b in the range $1 \leq b \leq n$ such that b and n are relatively prime. That is,

$$\varphi(n) := \#\{b \in N_n : \ \gcd(b,n) = 1\}.$$

We can use Inclusion/Exclusion to obtain a formula for $\varphi(n)$, as follows. Let the prime factorization of n be $n = p_1^{c_1} p_2^{c_2} \cdots p_m^{c_m}$, in which the p_i are pairwise distinct primes and the c_i are positive integers. For each $1 \leq i \leq m$, let

$$A_i := \{b \in N_n : \ p_i \text{ divides } b\}.$$

Then

$$\varphi(n) = \#(N_n \smallsetminus (A_1 \cup \cdots \cup A_m)) = n - \#(A_1 \cup \cdots \cup A_m).$$

Since the factors p_i are pairwise coprime, for any $\varnothing \neq S \subseteq N_m$ and $b \in N_n$ we have $b \in A_S$ if and only if $\prod_{i \in S} p_i$ divides b. Therefore,

$$\#A_S = \frac{n}{\prod_{i \in S} p_i}.$$

By Inclusion/Exclusion, it follows that

$$\#(A_1 \cup \cdots \cup A_m) = n \sum_{\varnothing \neq S \subseteq N_m} (-1)^{\#S-1} \prod_{i \in S} \frac{1}{p_i}.$$

Therefore

$$\begin{aligned} \varphi(n) &= n - n \sum_{\varnothing \neq S \subseteq N_m} (-1)^{\#S-1} \prod_{i \in S} \frac{1}{p_i} \\ &= n \sum_{S \subseteq N_m} (-1)^{\#S} \prod_{i \in S} \frac{1}{p_i} = n \prod_{i=1}^{m} \left(1 - \frac{1}{p_i}\right). \end{aligned}$$

In the last equality we have used Example 1.9 with $V = N_m$ and $x_i = -1/p_i$ for each $i \in N_m$.

---

1. Exercises.

---

1. Let S, T, and U be sets, and let $f : S \to T$ and $g : T \to U$ be functions. Prove the following implications:

f, g are both injections $\implies g \circ f$ is an injection $\implies f$ is an injection.

f, g are both surjections $\implies g \circ f$ is a surjection $\implies g$ is a surjection.

Deduce that equicardinality is an equivalence relation.

---

2. Let $n, m \in \mathbb{N}$ be natural numbers. Show that if N_n and N_m are equicardinal then $n = m$. (Hint: Suppose not. Then there is a smallest value of $n \in \mathbb{N}$ for which there exists $m \in \mathbb{N}$ with $m \neq n$ and $N_n \rightleftharpoons N_m$. Now derive a contradiction.)

---

3. Let S and T be two sets, and let $f : S \to T$ and $g : T \to S$ be functions. Consider the following two conditions:
(i) For all $s \in S$, $g(f(s)) = s$.
(ii) For all $t \in T$, $f(g(t)) = t$.
Show that if (i) holds then f is an injection and g is a surjection. Show that if (i) holds and f is a surjection then f and g are mutually inverse bijections. Deduce that both (i) and (ii) hold if and only if f and g are mutually inverse bijections.

---

4. Let S and T be finite sets, and assume that $S \rightleftharpoons T$. Let $f : S \to T$ be a function. Consider the following three conditions:
(i) f is a bijection.
(ii) f is a surjection.
(iii) f is an injection.
Show that these three conditions are equivalent. Give examples which show that that they are not equivalent if S and T are infinite.

---

5. Let S and T be finite sets, and let $f : S \to T$ and $g : T \to S$ be functions. Consider the following three conditions:
(i) Both f and g are surjections.
(ii) Both f and g are injections.
(iii) Both f and g are bijections.
Show that these three conditions are equivalent. Give an example of a pair of such functions for which $g \neq f^{-1}$. Give examples which show that if S and T are infinite then these conditions are not equivalent.

---

6. Prove Proposition 1.5.

---

7. Use Exercise 1.3 to verify that the function $B : \mathcal{F}(N_m, N_n) \to \mathcal{F}(X, Y)$ of Example 1.7 is a bijection.

---

8. Use Exercise 1.3 to verify that the constructions $S \mapsto f_S$ and $f \mapsto f^{-1}(1)$ of Example 1.8 are mutually inverse bijections between $\mathcal{P}(X)$ and $\mathcal{F}(X, \{0, 1\})$.

---

9. Let $G = (V, E)$ be a finite simple graph. For each $k \in \mathbb{N}$, let $P_G(k)$ be the number of functions $f : V \to N_k$ such that if $\{v, w\} \in E$ then $f(v) \neq f(w)$. For a subset $S \subseteq E$, let $\mathrm{comp}(S)$ denote the number of connected components of the spanning subgraph (V, S) of G. Prove that for all $k \in \mathbb{N}$,

$$P_G(k) = \sum_{S \subseteq E} (-1)^{\#S} k^{\mathrm{comp}(S)}.$$

This is a polynomial function of k, called the *chromatic polynomial* $P_G(x)$ of the graph G.

---

1. Endnotes.

The axiomatic theory of sets provides the logical foundation for all of mathematics. The subtleties of the theory arise only for infinite sets, and in this course we can avoid these difficulties since our focus is on finite sets. Just to indicate the kind of problems which arise, imagine that Ω is the set of all sets that do not contain themselves as elements. Is Ω an element of Ω, or not? Something funny is going on... we can not admit Ω as a set without contradiction!

As references to begin with, I recommend the following:

- P.R. Halmos, "Naive Set Theory," D. van Nostrand, Princeton, NJ, 1960.
- Y.N. Moschovakis, "Notes on Set Theory," Springer–Verlag, Ney York, 1994.

As a general–purpose self–help guide to combinatorial theory, one can do no better than

- L. Lovász, "Combinatorial Problems and Exercises," North–Holland, Amsterdam/New York, 1979.

Exercise 9 was first proved in

- H. Whitney, *A logical expansion in mathematics*, Bull. Amer. Math. Soc. **38** (1932), 572-579.

For an introduction to the theory of chromatic polynomials, see

- R.C. Read and W.T. Tutte, *Chromatic polynomials*, in "Selected Topics in Graph Theory 3" (L.W. Beineke and R.J. Wilson, ed.), Academic Press, San Diago, CA, 1988.

Regarding Example 1.12, the following amusing fact appears in a book by Gupta cited in Section 10: $\varphi(5186) = \varphi(5187) = \varphi(5188)$.

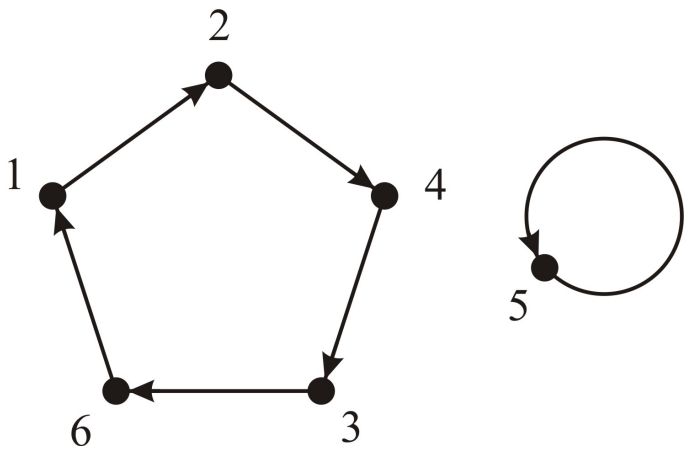


FIGURE 2.1. the functional directed graph of a permutation.

2. Permutations and Subsets.

Finite sets and bijections are the most basic objects in enumeration, but permutations and subsets are close in second place. (Indeed, a permutation is a special kind of bijection.) In this section we will prove some familiar results in order to illustrate the methods of Section 1. The details of these (seemingly over–complicated) constructions will yield extra dividends in Section 5.

For any natural number $n \in \mathbb{N}$, a *permutation of length* n is a bijection $\sigma : N_n \to N_n$. Permuations are traditionally written in *word notation*: that is, a permutation σ is represented by the sequence $a_1 a_2 ... a_n$ in which $a_i := \sigma(i)$. Thus, a permutation may be expressed as a word $a_1 a_2 ... a_n$ in which each number from 1 to n appears exactly once (in some order). For example, 2 4 6 3 5 1 is word notation for the permutation σ of length 6 for which $\sigma(1) := 2$, $\sigma(2) := 4$, $\sigma(3) := 6$, $\sigma(4) := 3$, $\sigma(5) := 5$, and $\sigma(6) := 1$. Figure 2.1 indicates the *functional directed graph* of this permutation σ, which is defined by drawing an arc $i \to \sigma(i)$ for each $i \in N_n$. Let $\mathcal{S}_n$ denote the set of all permutations of length n.

Theorem 2.1. *The number* $\#\mathcal{S}_n$ *of permutations of length* n *is*

$$n! := n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1.$$

Before giving a formal proof of this, here is an intuitive argument for the formula. Indeed, the proof which follows is a formalization of the idea embodied in this "simple proof". The permutations of length n are represented by words $a_1 a_2 ... a_n$ in which each number from 1 to n occurs exactly once in some order. We

can choose the first number a_1 in any of n ways. Having chosen a_1, we can choose a_2 in any of $n-1$ ways, since a_1 may not be repeated. Continuing in this fashion, we similarly have $n+1-i$ choices for a_i, since none of $a_1, ..., a_{i-1}$ may be repeated. This holds for all i from 1 to n. Every permutation is constructed exactly once in this manner, which "proves" the formula.

The above argument is conceptually clear, and has the advantage of being expressed almost entirely in English. However, the claim that each permutation is constructed exactly once, while obvious, has not been substantiated. The following formal proof corrects this flaw and illustrates the bijective technique of Exercise 1.3. It also gives us some extra information which will be developed in Section 5.

Proof of Theorem 2.1. We proceed by constructing a pair of mutually inverse bijections between the set $\mathcal{S}_n$ and the set $\mathcal{Q}_n := N_n \times N_{n-1} \times \cdots N_2 \times N_1$. Proposition 1.5 implies that $\#\mathcal{Q}_n = n!$ and then, since we will have shown that $\mathcal{S}_n \rightleftharpoons \mathcal{Q}_n$, Proposition 1.1 completes the proof.

We define a function $I_n : \mathcal{S}_n \to \mathcal{Q}_n$ as follows.

```
FUNCTION:  I_n from S_n to Q_n;
INPUT:  a_1 a_2 .... a_n;
repeat with i from 1 to n:
  let r_i be the number of j with i < j ≤ n and a_i > a_j;
end repeat;
OUTPUT:  (1 + r_1, 1 + r_2, ..., 1 + r_n).
```

For example, given $\sigma = 4\ 6\ 3\ 7\ 5\ 1\ 2 \in \mathcal{S}_7$ we get $I_7(\sigma) = (4, 5, 3, 4, 3, 1, 1) \in \mathcal{Q}_7$. In general, for each $1 \leq i \leq n$, the value of r_i is the number of entries in the permutation which are strictly to the right of a_i and are less than a_i.

To see that $I_n(\sigma) \in \mathcal{Q}_n$, we must check that $1 \leq 1 + r_i \leq n + 1 - i$ for every $1 \leq i \leq n$. But $r_i \geq 0$ since r_i is the cardinality of a set, and from $r_i \leq \#\{j : \ i < j \leq n\} = n - i$ we deduce the result. Thus I_n is a function from the set $\mathcal{S}_n$ to the set $\mathcal{Q}_n$, as claimed.

It remains to show that I_n is a bijection. To do this we construct the inverse function $J_n : \mathcal{Q}_n \to \mathcal{S}_n$ as follows.

```
FUNCTION:  J_n from Q_n to S_n;
INPUT:  (h_1, h_2, ..., h_n);
repeat for i from 1 to n;
  let b_i be the h_i-th smallest element of N_n \ {b_1, ..., b_{i-1}};
end repeat;
OUTPUT: b_1 b_2 ... b_n.
```

As an example, the sequence $\rho = (6, 6, 3, 1, 3, 1, 1) \in \mathcal{Q}_7$ produces the output $J_7(\rho) = 6\ 7\ 3\ 1\ 5\ 2\ 4\ \in \mathcal{S}_7$.

In general, since $(h_1, \ldots, h_n) \in \mathcal{Q}_n$, at step i we have $1 \leq h_i \leq n+1-i$, so that there is an h_i–th smallest element of the $(n+1-i)$–element set $N_n \smallsetminus \{b_1, \ldots, b_{i-1}\}$. Thus, the algorithm does not terminate improperly. When the `repeat` loop is finished the output $b_1 b_2 \ldots b_n$ will list the elements of N_n once each (in some order). That is, J_n does output a permutation in $\mathcal{S}_n$.

It remains to check that $J_n(I_n(\sigma)) = \sigma$ for every $\sigma \in \mathcal{S}_n$, and that $I_n(J_n(\rho)) = \rho$ for every $\rho = (h_1, \ldots, h_n) \in \mathcal{Q}_n$. By Exercise 1.3, this implies that I_n and J_n are mutually inverse bijections between $\mathcal{S}_n$ and $\mathcal{Q}_n$. We verify the first of these claims, leaving the other verification as an exercise.

Claim. For every $\sigma = a_1 a_2 \ldots a_n \in \mathcal{S}_n$, we have $J_n(I_n(\sigma)) = \sigma$.

Proof of Claim. By the definition of I_n we have $I_n(\sigma) = (1+r_1, \ldots, 1+r_n)$ in which $r_i = \#\{j : i < j \leq n \text{ and } a_i > a_j\}$ for each $1 \leq i \leq n$. From this it follows that a_i is the $(1 + r_i)$–th smallest element of the set $\{a_i, \ldots, a_n\}$, for each $1 \leq i \leq n$. Now apply J_n to $I_n(\sigma)$ to get the output permutation $b_1 b_2 \ldots b_n$; we will prove that $b_i = a_i$ for all $1 \leq i \leq n$ by induction on i.

By definition, b_1 is the $(1 + r_1)$–th smallest element of N_n. From the above paragraph, so is a_1; thus $b_1 = a_1$, establishing the basis of induction. Now assume that $b_i = a_i$ for all $1 \leq i \leq k$, where $1 \leq k < n$; we will show that $b_{k+1} = a_{k+1}$ as well, establishing the induction step. By definition, b_{k+1} is the $(1+r_{k+1})$–th smallest element of $N_n \smallsetminus \{b_1, \ldots, b_k\}$. Thus, since $b_i = a_i$ for all $1 \leq i \leq k$, it follows that b_{k+1} is the $(1 + r_{k+1})$–th smallest element of $N_n \smallsetminus \{a_1, \ldots, a_k\} = \{a_{k+1}, \ldots, a_n\}$. From the previous paragraph, this element is a_{k+1}; hence $b_{k+1} = a_{k+1}$, completing the induction step and finishing the proof of the claim. □

Verification that $I_n(J_n(\rho)) = \rho$ for all $\rho \in \mathcal{Q}_n$ is left as an exercise. This shows that I_n and J_n are mutually inverse bijections, so that $\mathcal{S}_n \rightleftharpoons \mathcal{Q}_n$ and hence $\#\mathcal{S}_n = \#\mathcal{Q}_n = n!$. This completes the proof of Theorem 2.1. □

Example 2.2. Let X be any finite set. A *permutation of* X is a bijection $\sigma : X \to X$. We claim that if $\#X = n$ then there are $n!$ permutations of the set X. To see this, let $\mathcal{S}_X$ be the set of all permutations of X – we now exhibit a bijection $\mathcal{S}_X \rightleftharpoons \mathcal{S}_n$. Since $\#X = n$ there is a bijection $f : X \to N_n$, by Proposition 1.1. Define a function $\alpha : \mathcal{S}_X \to \mathcal{S}_n$ by $\alpha(\sigma) := f \circ \sigma \circ f^{-1}$. By Exercise 1.1, $\alpha(\sigma)$ is a bijection from N_n to N_n – that is, a permutation in $\mathcal{S}_n$. The inverse function $\beta : \mathcal{S}_n \to \mathcal{S}_X$ is defined by $\beta(\tau) := f^{-1} \circ \tau \circ f$ for all $\tau \in \mathcal{S}_n$. Exercise 1.3 can now be used to show that α and β are mutually inverse bijections between $\mathcal{S}_X$ and S_n. Therefore, $\#\mathcal{S}_X = \#\mathcal{S}_n = n!$.

Having found a formula for the number of permutations of length n, we can use it to derive a formula for the number of k–element subsets of the n–element

set N_n. You quite likely already know that this number is a binomial coefficient, but the details of this bijection will again yield extra information in Section 5. I am trying to emphasize the "philosophy" of bijective proofs here – the point being that this way of thinking is very useful in situations that are not as clear–cut as the present one.

Let $\mathcal{B}(n,k)$ denote the set of all k–element subsets of the set N_n.

Theorem 2.3. *For any $n \in \mathbb{N}$ and $0 \leq k \leq n$, the number of k–element subsets of N_n is*

$$\#\mathcal{B}(n,k) = \binom{n}{k} := \frac{n!}{k!(n-k)!}.$$

Proof. We must show that $\#\mathcal{B}(n,k) = n!/(k!(n-k)!)$. Since we know that $\#\mathcal{S}_m = m!$, it is enough to show that

$$\#(\mathcal{B}(n,k) \times \mathcal{S}_k \times \mathcal{S}_{n-k}) = \#\mathcal{S}_n.$$

To do this we construct a bijection $\Psi_{n,k}$ from $\mathcal{S}_n$ to $\mathcal{B}(n,k) \times \mathcal{S}_k \times \mathcal{S}_{n-k}$. In order to construct $\Psi_{n,k}$ the following subroutine is useful. For any $m \in \mathbb{N}$, let $\mathcal{R}_m$ denote the set of all sequences $a_1 a_2 \dots a_m$ of m pairwise distinct positive integers.

```
FUNCTION: P_m from R_m to S_m;
INPUT: a_1 a_2 ... a_m;
repeat with i from 1 to m;
  let h_i be the number of j ∈ N_m such that a_i ≥ a_j;
end repeat;
OUTPUT: h_1 h_2 ... h_m.
```

For example, $P_6(3\ 9\ 2\ 8\ 6\ 5) := 2\ 6\ 1\ 5\ 4\ 3$. The effect of P_m is to replace the i–th smallest element of the input by the number i, for each $1 \leq i \leq m$. The result is thus a permutation in $\mathcal{S}_m$.

Now we define the function $\Psi_{n,k}$.

```
FUNCTION: Ψ_{n,k} from S_n to B(n,k) × S_k × S_{n-k};
INPUT: a_1 a_2 ... a_n;
let A := {a_1, a_2, ..., a_k};
let β := P_k(a_1 a_2 ... a_k);
let γ := P_{n-k}(a_{k+1} a_{k+2} ... a_n);
OUTPUT: (A, β, γ).
```

For example, $\Psi_{7,3}(2\ 5\ 4\ 7\ 1\ 6\ 3) := (\{2,4,5\},\ 1\ 3\ 2,\ 4\ 1\ 3\ 2)$. In general, it is clear that A is a k–element subset of N_n, so that the triple (A,β,γ) is in $\mathcal{B}(n,k) \times \mathcal{S}_k \times \mathcal{S}_{n-k}$, as required. To show that $\Psi_{n,k}$ is a bijection we construct the inverse function $\Phi_{n,k} : \mathcal{B}(n,k) \times \mathcal{S}_k \times \mathcal{S}_{n-k} \to \mathcal{S}_n$ as follows.

```
FUNCTION: Φ_{n,k} from B(n,k) × S_k × S_{n−k} to S_n;
INPUT: (A, β, γ);
sort A as s_1 < s_2 < ··· < s_k;
sort N_n ∖ A as t_1 < t_2 < ··· < t_{n−k};
repeat with i from 1 to k;
  let c_i := s_{β(i)};
end repeat;
repeat with j from 1 to n − k;
  let c_{k+j} := t_{γ(j)};
end repeat;
OUTPUT: c_1 c_2 ... c_n .
```

(You should compute a few arbitrary examples to get a feel for how $\Phi_{n,k}$ works.) Since $c_1c_2\dots c_n$ consists of the elements of A (in some order) followed by the elements of $N_n \smallsetminus A$ (in some order), it follows that $c_1c_2\dots c_n$ is a permutation in $\mathcal{S}_n$, as required.

Now we show that $\Psi_{n,k}$ and $\Phi_{n,k}$ are mutually inverse bijections.

Claim. For every $\sigma = a_1a_2\dots a_n \in \mathcal{S}_n$, we have $\Phi_{n,k}(\Psi_{n,k}(\sigma)) = \sigma$.

Proof of Claim. Let $\Psi_{n,k}(\sigma) = (A, \beta, \gamma)$. We have $A = \{a_1, a_2, \dots, a_k\}$, $\beta = P_k(a_1a_2\dots a_k)$, and $\gamma = P_{n-k}(a_{k+1}a_{k+2}\dots a_n)$. In the algorithm for $\Phi_{n,k}$, let's see what the first `repeat` loop does with A and β. For each i from 1 to k, c_i is the $\beta(i)$–th smallest element of A. Since $\beta = P_k(a_1a_2\dots a_k)$, the $\beta(i)$–th smallest element of A is a_i. Therefore, $c_i = a_i$ for all $1 \le i \le k$. Similarly, in the second `repeat` loop, for each $1 \le j \le n-k$ the output c_{k+j} is the $\gamma(j)$–th smallest element of $\{a_{k+1}, \dots, a_n\}$. As in the first case, since $\gamma = P_{n-k}(a_{k+1}a_{k+2}\dots a_n)$ it follows that $c_{k+j} = a_{k+j}$ for all $1 \le j \le n-k$. Hence, $c_1\dots c_n = a_1\dots a_n$, completing the proof of the claim. □

Verification that $\Psi_{n,k}(\Phi_{n,k}(A, \beta, \gamma)) = (A, \beta, \gamma)$ for all $(A, \beta, \gamma) \in \mathcal{B}(n,k) \times \mathcal{S}_k \times \mathcal{S}_{n-k}$ is left as an exercise. By Exercise 1.3, this shows that the functions $\Psi_{n,k}$ and $\Phi_{n,k}$ are mutually inverse bijections. Since $\mathcal{B}(n,k) \times \mathcal{S}_k \times \mathcal{S}_{n-k} \rightleftharpoons \mathcal{S}_n$, Propositions 1.1 and 1.5 and Theorem 2.1 imply that $(\#\mathcal{B}(n,k))k!(n-k)! = n!$. This completes the proof of Theorem 2.3. □

The identity

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

is frequently useful and we will employ it when necessary without further comment.

Example 2.4. Let X be a finite set, with $\#X = n$, say, and let $0 \le k \le n$. There are $\binom{n}{k}$ k–element subsets of the set X. To see this, let $\mathcal{B}(X,k)$ be the set of all

k–element subsets of X – we now describe a bijection $\mathcal{B}(X,k) \rightleftharpoons \mathcal{B}(n,k)$. Since $\#X = n$ there is a bijection $f : X \to N_n$, by Proposition 1.1. Define a function $\alpha : \mathcal{B}(X,k) \to \mathcal{B}(n,k)$ by putting $\alpha(S) := \{f(s) : \ s \in S\}$ for every $S \in \mathcal{B}(X,k)$. Also define a function $\beta : \mathcal{B}(n,k) \to \mathcal{B}(X,k)$ by putting $\beta(T) := \{f^{-1}(t) : \ t \in T\}$ for every $T \in \mathcal{B}(n,k)$. Exercise 1.3 can now be used to show that α and β are mutually inverse bijections, so that $\#\mathcal{B}(X,k) = \#\mathcal{B}(n,k) = \binom{n}{k}$.

Example 2.5. In Example 1.9 we derived the formula

$$\sum_{S \subseteq V} x^{\#S} = (1+x)^{\#V}$$

for any finite set V. By Example 2.4, V has $\binom{\#V}{k}$ k–element subsets, for each $0 \leq k \leq \#V$. Therefore

$$(1+x)^{\#V} = \sum_{S \subseteq V} x^{\#S} = \sum_{k=0}^{\#V} \binom{\#V}{k} x^k.$$

This is one proof of the Binomial Theorem.

Example 2.6. A permutation $\sigma \in \mathcal{S}_n$ is called a *derangement* if $\sigma(i) \neq i$ for all $i \in N_n$. Let $\mathcal{D}_n \subset \mathcal{S}_n$ be the set of derangements of length n. What is $\#\mathcal{D}_n$? We can solve this by using Inclusion/Exclusion. For each $i \in N_n$, let A_i be the set of those permutations $\sigma \in \mathcal{S}_n$ such that $\sigma(i) = i$. Then

$$\mathcal{D}_n = \mathcal{S}_n \smallsetminus (A_1 \cup \cdots \cup A_n).$$

For any $\varnothing \neq S \subseteq N_n$, a permutation $\sigma \in \mathcal{S}_n$ is in A_S if and only if $\sigma(i) = i$ for all $i \in S$. If $\#S = k$ then this fixes k of the values of σ. The remaining $n-k$ elements in $N_n \smallsetminus S$ must be permuted among themselves by σ – therefore $A_S \rightleftharpoons \mathcal{S}_{(N \smallsetminus S)}$ and so there are $(n-k)!$ permutations in A_S, whenever $\#S = k$. Now, by Inclusion/Exclusion,

$$\begin{aligned}
\#\mathcal{D}_n &= \#(\mathcal{S}_n \smallsetminus (A_1 \cup \cdots \cup A_n)) = n! - \#(A_1 \cup \cdots \cup A_n) \\
&= n! - \sum_{\varnothing \neq S \subseteq N_n} (-1)^{\#S-1} (\#A_S) \\
&= n! - \sum_{k=1}^{n} \binom{n}{k} (-1)^{k-1} (n-k)! = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}.
\end{aligned}$$

Notice that as $n \to \infty$, the ratio $\#\mathcal{D}_n / n!$ tends to the limit

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = \frac{1}{e}.$$

Since $\#\mathcal{S}_n = n!$ this can be interpreted as saying that as $n \to \infty$, the probability that a randomly chosen permutation in $\mathcal{S}_n$ is a derangement is asymptotically $1/e \approx 0.36787944\ldots$.

---

2. Exercises.

---

1. Complete the proof of Theorem 2.1 by showing that for all $\rho \in \mathcal{Q}_n$, $I_n(J_n(\rho)) = \rho$.

---

2. A permutation $\sigma \in \mathcal{S}_n$ is a *cyclic permutation* if $n \geq 1$ and for every $v, w \in N_n$, there is a $k \in \mathbb{N}$ such that $\sigma^k(v) = w$. (Here, σ^k refers to the k–fold functional iteration of σ.) Let $\mathcal{C}_n$ be the subset of cyclic permutations in $\mathcal{S}_n$. Define a function $f : \mathcal{S}_n \to \mathcal{C}_n$ such that $\#f^{-1}(\gamma) = n$ for all $\gamma \in \mathcal{C}_n$. By Proposition 1.3 this implies that $\#\mathcal{C}_n = (n-1)!$ for all $n \geq 1$.

---

3. Complete the proof of Theorem 2.3 by showing that $\Psi_{n,k}(\Phi_{n,k}(A, \beta, \gamma)) = (A, \beta, \gamma)$ for all $(A, \beta, \gamma) \in \mathcal{B}(n,k) \times \mathcal{S}_k \times \mathcal{S}_{n-k}$.

---

4. Let k_1, k_2, …, k_r be nonnegative integers, and let $n := k_1 + \cdots + k_r$. Define a *multinomial coefficient* by

$$\binom{n}{k_1, \ldots, k_r} := \frac{n!}{k_1! k_2! \cdots k_r!}.$$

Show that $\binom{n}{k_1,\ldots,k_r}$ is the number of ways to express N_n as a union

$$N_n = A_1 \cup A_2 \cup \cdots \cup A_r$$

of an ordered sequence $(A_1, \ldots, A_r)$ of r pairwise disjoint subsets such that $\#A_i = k_i$ for each $1 \leq i \leq r$.

---

5. Fix $n \in \mathbb{N}$ and a permutation $\pi \in \mathcal{S}_n$. Let $\mathcal{D}(\pi)$ be the set of permutations $\sigma \in \mathcal{S}_n$ such that $\sigma(v) \neq \pi(v)$ for all $v \in N_n$. (For example, if $\pi = \iota$ is the identity permutation, then $\mathcal{D}(\iota) = \mathcal{D}_n$ is the set of derangements in Example 2.6.) Show that for every $\pi \in \mathcal{S}_n$,

$$\#\mathcal{D}(\pi) = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

---

6. A *group* is a set Γ together with a binary operation $\circ : \Gamma \times \Gamma \to \Gamma$ which satisfies the following axioms:
(i) there is an element $\iota \in \Gamma$ such that $\iota \circ g = g \circ \iota = g$ for all $g \in \Gamma$;
(ii) for all $f, g, h \in \Gamma$, we have $(f \circ g) \circ h = f \circ (g \circ h)$;
(iii) for all $g \in \Gamma$ there exists an $h \in \Gamma$ such that $g \circ h = h \circ g = \iota$.

(a) Show that in a group the element ι is unique.

(b) Show that in a group, for every $g \in \Gamma$ the element h guaranteed by axiom (iii) is unique.

(c) Show that the set $\mathcal{S}_n$ of permutations with the operation $\circ$ of functional composition is a group.

---

3. Lattice Paths and Polynomial Identities.

For many people, geometric or visual reasoning is a helpful device for solving problems. The next example gives a useful geometric interpretation for binomial coefficients.

Example 3.1. Let $a, b \in \mathbb{N}$, and consider the set $\mathcal{L}(a,b)$ of all paths in the integer plane $\mathbb{Z} \times \mathbb{Z}$ which begin at $(0,0)$, end at (a,b), and consist of a sequence of steps either to the *east*: $(x,y) \to (x+1,y)$, or to the *north*: $(x,y) \to (x,y+1)$. What is $\#\mathcal{L}(a,b)$?

To each such path corresponds the sequence

$$P = s_1 s_2 ... s_{a+b}$$

in which for each $1 \leq i \leq a+b$ we have $s_i \in \{\mathsf{E}, \mathsf{N}\}$; here E denotes a step to the east and N denotes a step to the north. Such a sequence P represents a path in $\mathcal{L}(a,b)$ if and only if a of its elements are E and b of its elements are N. Thus the paths in $\mathcal{L}(a,b)$ are in bijective correspondence with the set of sequences of a E–s and b N–s. Now to any such sequence corresponds the set

$$S_P := \{i \in N_{a+b} : s_i = \mathsf{N}\}.$$

This is a b–element subset of N_{a+b}. Conversely, to any b–element subset $S \subseteq N_{a+b}$ corresponds the sequence $P_S := s_1 s_2 ... s_{a+b}$ defined by

$$s_i := \begin{cases} \mathsf{N} & \text{if } \ i \in S, \\ \mathsf{E} & \text{if } \ i \notin S. \end{cases}$$

This P_S is a path in $\mathcal{L}(a,b)$. Figure 3.1 illustrates this correspondence. These constructions $P \mapsto S_P$ and $S \mapsto P_S$ are mutually inverse bijections $\mathcal{L}(a,b) \rightleftharpoons \mathcal{B}(a+b,b)$. We conclude that $\#\mathcal{L}(a,b) = \#\mathcal{B}(a+b,b) = \binom{a+b}{b}$.

Since $\binom{a+b}{b} = \binom{a+b}{a}$, Proposition 1.1 implies that there is a bijection $\mathcal{L}(a,b) \rightleftharpoons \mathcal{L}(b,a)$. I leave it to you to describe such a bijection explicitly – it is quite easy.

Now we have another combinatorial interpretation for the binomial coefficients $\binom{n}{k}$: these numbers count not only k–element subsets of an n–set, but also lattice paths from the origin $(0,0)$ to $(n-k,k)$. Using either of these interpretations we can construct bijective proofs of many equations, known collectively as *binomial identities*.

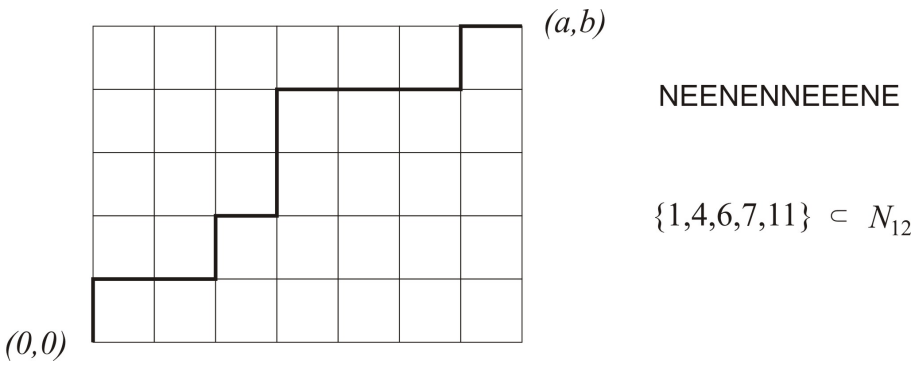


FIGURE 3.1. a path, sequence, and subset.

Example 3.2. For $0 \leq j \leq k \leq n$,

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

This identity is so easy that it can be verified numerically without difficulty:

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

However, this numerical identity is a consequence of a deeper fact – a bijection between two sets. Let X be a set with $\#X = n$. Let $\mathcal{A}$ be the set of all pairs (S,T) such that $S \subseteq T \subseteq X$ and $\#S = j$ and $\#T = k$. By considering the function $f : \mathcal{A} \to \mathcal{B}(X,j)$ given by $f(S,T) := T$, Proposition 1.3 and Example 2.4 imply that

$$\#\mathcal{A} = \binom{n}{k}\binom{k}{j},$$

the LHS of the identity. Similarly, the RHS is the cardinality of the set $\mathcal{B}$ of all pairs (P,Q) with $P \subseteq X$, $Q \subseteq X$, $\#P = j$, $\#Q = k-j$, and $P \cap Q = \varnothing$. We define mutually inverse bijections between $\mathcal{A}$ and $\mathcal{B}$ as follows:

$$\begin{array}{rcl} \mathcal{A} & \rightleftharpoons & \mathcal{B} \\ (S,T) & \mapsto & (S, T \smallsetminus S) \\ (P, P \cup Q) & \leftarrow & (P,Q) \end{array}$$

That these are mutually inverse bijections is easily seen, and it follows that $\#\mathcal{A} = \#B$, establishing the identity.

Example 3.3. For $a, b \in \mathbb{N}$,

$$\binom{a+1+b}{b} = \sum_{j=0}^{b} \binom{a+j}{j}.$$

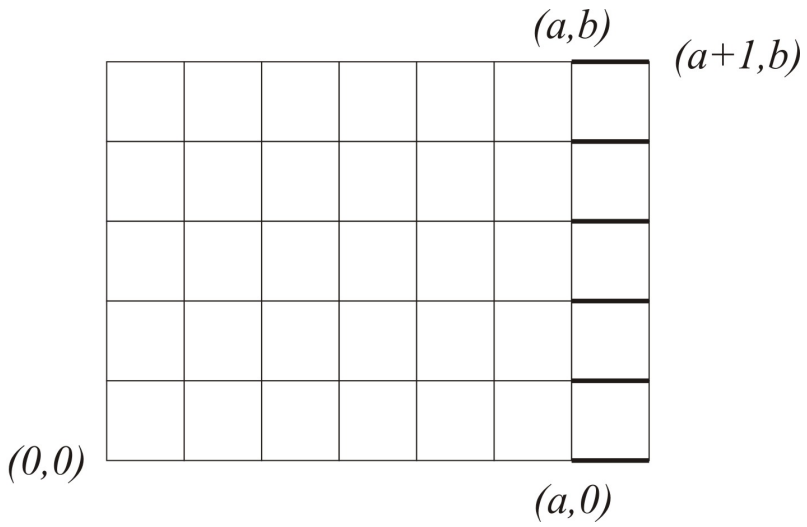


FIGURE 3.2. illustration of Example 3.3.

This identity is not so easy to derive numerically, but the bijective technique works very well. The key is to "decode" what the identity is really about.

By Example 3.1, $\#\mathcal{L}(a+1,b) = \binom{a+1+b}{b}$. Given $a, b \in \mathbb{N}$, each path in $\mathcal{L}(a+1,b)$ uses exactly one of the edges $(a,j) \to (a+1,j)$ for $0 \le j \le b$. (See Figure 3.2.) Such a path consists of a path in $\mathcal{L}(a,j)$ followed by EN^{b-j}. Thus there is a bijective correspondence

$$\mathcal{L}(a+1,b) \rightleftharpoons \bigcup_{j=0}^{b} \left(\mathcal{L}(a,j) \times \{\mathsf{EN}^{b-j}\}\right)$$

in which the union is disjoint. Taking cardinalities of both sides, we obtain the stated formula.

Example 3.4. For $n \in \mathbb{N}$,

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

To prove this, notice that each lattice path from $(0,0)$ to (n,n) passes through exactly one of the points $(k, n-k)$ for $0 \le k \le n$. (See Figure 3.3.) Such a path consists of a path in $\mathcal{L}(k, n-k)$ followed by a translation by $(k, n-k)$ of a path in $\mathcal{L}(n-k, k)$. Thus there is a bijective correspondence

$$\mathcal{L}(n,n) \rightleftharpoons \bigcup_{k=0}^{n} \left(\mathcal{L}(k,n-k) \times \mathcal{L}(n-k,k)\right)$$

in which the union is disjoint. Taking cardinalities of both sides yields the result.

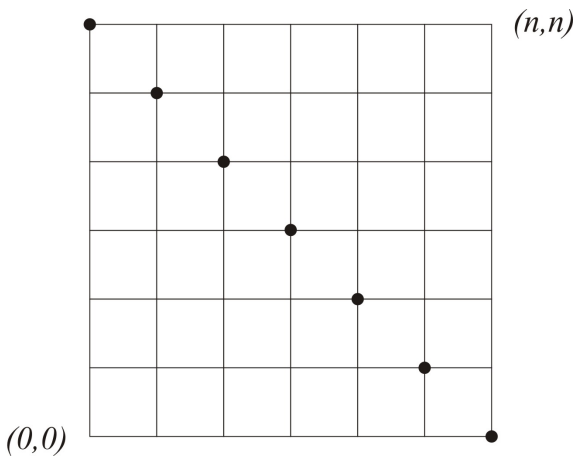


FIGURE 3.3. illustration of Example 3.4.

Example 3.5 (Stirling Numbers of the Second Kind)**.** Let X be a finite set. A *set partition* of X is a finite set $\pi = \{B_1, \ldots, B_k\}$ such that:
• for each $1 \leq i \leq k$: $\varnothing \neq B_i \subseteq X$,
• for each $1 \leq i < j \leq k$: $B_i \cap B_j = \varnothing$, and
• $B_1 \cup \cdots \cup B_k = X$.
Let $S(n,k)$ denote the number of set partitions of N_n of size k. (These are known as the *Stirling numbers of the second kind.*) Then $S(0,0) = 1$, $S(n,0) = 0$ for all $n \geq 1$, and for all $1 \leq k \leq n$ we have

$$S(n,k) = S(n-1,k-1) + kS(n-1,k).$$

To prove this, let $\Pi(n,k)$ denote the set of set partitions of N_n which have size k, so that $\#\Pi(n,k) = S(n,k)$. Let $\Pi'(n,k) \subseteq \Pi(n,k)$ be the subset of those $\pi \in \Pi(n,k)$ such that $\{n\} \in \pi$ (that is, n occurs by itself as a set in π), and let $\Pi''(n,k) := \Pi(n,k) \smallsetminus \Pi'(n,k)$. Certainly, we have

$$\Pi(n,k) = \Pi'(n,k) \cup \Pi''(n,k)$$

and the union is disjoint. Therefore,

$$S(n,k) = \#\Pi'(n,k) + \#\Pi''(n,k).$$

Moreover, we have a bijection

$$\begin{array}{rcl} \Pi'(n,k) & \rightleftharpoons & \Pi(n-1,k-1) \\ \pi & \mapsto & \pi \smallsetminus \{\{n\}\} \\ \mu \cup \{\{n\}\} & \leftarrow & \mu \end{array}$$

so that $\#\Pi'(n,k) = \#\Pi(n-1,k-1) = S(n-1,k-1)$. The function

$$\begin{aligned} \Pi''(n,k) &\to \Pi(n-1,k) \\ \pi &\mapsto \{B \setminus \{n\} : B \in \pi\} \end{aligned}$$

is such that the preimage of every $\mu \in \Pi(n-1,k)$ has size k. Proposition 1.3 implies that $\#\Pi''(n,k) = k(\#\Pi(n-1,k)) = kS(n-1,k)$. These observations suffice to prove the claim.

The recursion of Example 3.5 can be used to calculate the first several values of $S(n,k)$, illustrated in the following table for $0 \leq k \leq n \leq 7$.

$n \setminus k$	0	1	2	3	4	5	6	7
0	1	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0
2	0	1	1	0	0	0	0	0
3	0	1	3	1	0	0	0	0
4	0	1	7	6	1	0	0	0
5	0	1	15	25	10	1	0	0
6	0	1	31	90	65	15	1	0
7	0	1	63	301	350	140	21	1

Further properties of these numbers $S(n,k)$ are developed in the exercises.

We have seen a few examples of how bijective proofs can be used to establish numerical identities. If both sides depend polynomially on a particular value, then this establishes an identity **of polynomials**, for the following reason.

Proposition 3.6. *Let $p(y)$ and $q(y)$ be polynomials in the variable y. If $p(n) = q(n)$ for infinitely many natural numbers n, then $p(y) = q(y)$ as polynomials.*

Proof. Under the hypotheses, $p(y)-q(y)$ is a polynomial which vanishes for infinitely many natural numbers. Therefore, this must be the zero polynomial. □

To illustrate this, reconsider Example 3.3. For $k \in \mathbb{N}$, we can interpret the binomial coefficient $\binom{y}{k}$ as a polynomial function of y by using the definition

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

This agrees with the definition in Theorem 2.3 when $y = n \in \mathbb{N}$ is a natural number. Since Example 3.3 holds for all $a \in \mathbb{N}$, we deduce the polynomial identities

$$\binom{y+1+b}{b} = \sum_{j=0}^{b} \binom{y+b}{b}$$

for each $b \in \mathbb{N}$.

Similarly, from Example 3.2 and Proposition 3.6 we deduce the polynomial identities

$$\binom{y}{k}\binom{k}{j} = \binom{y}{j}\binom{y-j}{k-j}$$

for all $0 \leq j \leq k$.

Example 3.7 (The Binomial Series Expansion)**.** Consider the Binomial Theorem: for every $n \in \mathbb{N}$,

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

If $n \in \mathbb{N}$ and $n < k$, then $\binom{n}{k} = 0$, so we might as well extend this summation to infinity:

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

Consider the coefficient of x^k on both sides – that is,

$$[x^k](1+x)^n = \binom{n}{k}.$$

This is a polynomial function of n of degree k. Since both sides agree for all $n \in \mathbb{N}$, they must be equal as polynomials. That is,

$$[x^k](1+x)^y = \binom{y}{k}$$

for all $k \in \mathbb{N}$. (You might feel uncomfortable about the fact that we have not shown that the LHS is a polynomial function of y. Fair enough – see Definition 6.3 and Example 7.8.) In summary, we have the power series expansion

$$(1+x)^y = \sum_{k=0}^{\infty} \binom{y}{k} x^k,$$

in which both x and y are indeterminates. Since $[x^k](1+x)^y$ is a polynomial function of y for each $k \in \mathbb{N}$, we may specialize the indeterminate y to any particular complex number and the result will remain a well–defined power series with coefficients in $\mathbb{C}$. That is,

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

for every $\alpha \in \mathbb{C}$.

3. Exercises.

For Exercises 1 to 8, prove the polynomial identity by using Proposition 3.6 and constructing a bijection to prove the corresponding numerical identity. Example 1.7 is relevant for some of the exercises.

---

1. For all $k \geq 1$:

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

---

2. For $n \in \mathbb{N}$:

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

---

3. For $n \in \mathbb{N}$:

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

---

4. For $n \in \mathbb{N}$:

$$(x+y+z)^n = \sum_{i+j+k=n} \binom{n}{i,j,k} x^i y^j z^k.$$

(This is the "Trinomial Theorem".)

---

5. For $k \in \mathbb{N}$:

$$\binom{x+y}{k} = \sum_{j=0}^{k} \binom{x}{j}\binom{y}{k-j}.$$

(This is the "Vandermonde Convolution Formula".)

---

6. For $a, n \in \mathbb{N}$:

$$\binom{x+1+n}{n} = \sum_{j=0}^{n} \binom{a+j}{j}\binom{x-a+n-j}{n-j}.$$

---

7. For $i, j \in \mathbb{N}$:

$$\binom{x}{i}\binom{x}{j} = \sum_{k=0}^{i+j} \binom{k}{k-j, k-i, i+j-k}\binom{x}{k}.$$

---

8. For $n \in \mathbb{N}$:

$$x^n = \sum_{k=0}^{n} k! S(n,k) \binom{x}{k}.$$

---

9. For $n \in \mathbb{N}$:

$$\binom{2n}{n} = 2\sum_{j=0}^{n-1} \binom{n-1+j}{j}.$$

(This is not a polynomial function of n, so that Proposition 3.6 does not apply.)

---

10. Let $a_0,\ a_1,\ a_2, \ldots$ and $b_0,\ b_1,\ b_2, \ldots$ be sequences of numbers. Consider the following two conditions.
(i) for all $n \in \mathbb{N}$:

$$b_n = \sum_{j=0}^{n} \binom{n}{j} a_j.$$

(ii) for all $n \in \mathbb{N}$:

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

Prove that these two conditions are equivalent.

---

11. Deduce from Exercises 8 and 10 that for all $n, k \in \mathbb{N}$,

$$S(n,k) = \frac{1}{k!} \sum_{j=1}^{k} \binom{k}{j} (-1)^{k-j} j^n.$$

---

12. Show that if $2 \leq k \leq p-1$ and p is prime, then p divides $S(p,k)$.

---

13. For a graph $G = (V, E)$ and $k \in \mathbb{N}$, let $e_k(G)$ denote the number of k–element subsets $S \subseteq V$ of vertices such that no two vertices of S are joined by an edge. Let $P_G(x)$ be the chromatic polynomial defined in Exercise 1.9.

(a) Prove the polynomial identity

$$P_G(x) = \sum_{k=0}^{\#V} k! e_k(G) \binom{x}{k}.$$

(Exercise 8 is the special case of this for the graph $\overline{K}_n$ with n vertices and no edges.)

(b) Prove that

$$e_k(G) = \frac{1}{k!} \sum_{j=1}^{k} \binom{k}{j} (-1)^{k-j} P_G(j).$$

(Exercise 11 is the special case of this for the graph $\overline{K}_n$ with n vertices and no edges.)

---

4. Ordinary Generating Functions.

Ordinary generating functions are used to solve whole families of related enumeration problems all at the same time. Before developing the theory it might help to go through a simple example somewhat informally.

Consider two six–sided dice, one red and one white. When rolling these dice, the outcome can be described by an ordered pair (a, b) in which the number of pips on the top face of the red die is a and the number of pips on the top face of the white die is b. Of course, this ordered pair (a, b) is in the set $N_6 \times N_6$, and we assume that the dice are *fair* – that is, every outcome in $N_6 \times N_6$ is equally likely.

One enumeration question that can be asked about these dice is the following: in how many ways can one roll a nine? That is, how many outcomes $(a, b) \in N_6 \times N_6$ are such that $a + b = 9$? Of course, there is nothing special about "nine" here, and we might ask more generally – for each $n \in \mathbb{N}$, how many outcomes $(a, b) \in N_6 \times N_6$ are such that $a + b = n$? This is a family of related enumeration problems of the type which can be solved by ordinary generating functions.

Consider the expression

$$\sum_{a=1}^{6} \sum_{b=1}^{6} x^{a+b},$$

in which x is an **indeterminate**, or a variable which does not satisfy any special identities. (In particular, x does not have any particular real or complex value.) Collecting terms with the same power of x, we see that for all $n \in \mathbb{N}$, the coefficient of x^n in this polynomial is the number of outcomes $(a, b) \in N_6 \times N_6$ such that $a + b = n$. That is, this polynomial solves all of the related enumeration problems we are considering by presenting the solutions as the coefficients of the various powers of x. After a little algebra, we see that the polynomial is

$$\begin{aligned} & (x + x^2 + x^3 + x^4 + x^5 + x^6)^2 \\ = \; & x^2 + 2x^3 + 3x^4 + 4x^5 + 5x^6 + 6x^7 + 5x^8 + 4x^9 + 3x^{10} + 2x^{11} + x^{12}. \end{aligned}$$

In particular, there are four ways to roll a nine on two six–sided dice, as is easily verified – $(3, 6)$, $(4, 5)$, $(5, 4)$, and $(6, 3)$.

It is now time to introduce the formal definition of a generating function. In contrast with the simple example above, there may be more than one variable. Accordingly, let us say that $x_1, x_2, ..., x_r$ are indeterminates which commute with one another; that is, $x_i x_j = x_j x_i$ for all i and j. Thus, when we multiply a finite number of the x_i together we may write their product in the form $x_1^{a_1} x_2^{a_2} \cdots x_r^{a_r}$ in which each a_i is a natural number. It helps to have a handy notation for these

monomials. Let $\alpha := (a_1, a_2, ..., a_r) \in \mathbb{N}^r$. If $\mathbf{x} := (x_1, x_2, ..., x_r)$ is a sequence of pairwise commuting indeterminates, then we use the notation

$$\mathbf{x}^\alpha := x_1^{a_1} x_2^{a_2} \cdots x_r^{a_r}$$

for the corresponding monomial in the $\mathbf{x}$ variables. If $\alpha := (a_1, a_2, ..., a_r)$ and $\beta := (b_1, b_2, ..., b_r)$ are in $\mathbb{N}^r$, then the sum of α and β is

$$\alpha + \beta := (a_1 + b_1, a_2 + b_2, ..., a_r + b_r),$$

the usual vector sum. We also write $\alpha \leq \beta$ when $a_i \leq b_i$ for all $1 \leq i \leq r$. With this notation we have the familiar exponent law: $\mathbf{x}^\alpha \cdot \mathbf{x}^\beta = \mathbf{x}^{\alpha+\beta}$. Also, $\mathbf{x}^\alpha$ divides $\mathbf{x}^\beta$ if and only if $\alpha \leq \beta$.

One very general situation in which generating functions are applicable is as follows. We are given a set S (which need not be finite) and a function $\omega : S \to \mathbb{N}^r$ which has the following property:

[WF] For each $\alpha \in \mathbb{N}^r$, the preimage $\omega^{-1}(\alpha) := \{s \in S : \ \omega(s) = \alpha\}$ of α under ω is finite.

Such a function $\omega : S \to \mathbb{N}^r$ is called a *weight function* on S.

Definition 4.1 (Ordinary Generating Functions)**.** Let S be a set with a weight function $\omega : S \to \mathbb{N}^r$. The *ordinary generating function of S with respect to ω* is defined to be

$$\Phi_S^\omega(\mathbf{x}) := \sum_{s \in S} \mathbf{x}^{\omega(s)}.$$

Here are a few examples to consider, just to get started. It may help to return to them after reading the rest of this section.

Example 4.2. Let $S := \{1, 2, 3, ...,\}$, and define a function $\omega : S \to \mathbb{N}$ by saying that $\omega(s) := k$ if and only if 2^k divides s but 2^{k+1} does not divide s. What is the generating function $\Phi_S^\omega(x) := \sum_{s \in S} x^{\omega(s)}$ in this case? Well, this is a trick question! In fact, the summation $\sum_{s \in S} x^{\omega(s)}$ is not well–defined, because there are infinitely many $s \in S$ such that $\omega(s) = 0$, for example (namely, the odd positive integers). Therefore ω is not a weight function on S. If we try to sum this series we end up with the coefficient of x^0 diverging to ∞, which is not acceptable for algebraic manipulations. (In fact, in this example **every** coefficient diverges to ∞.) In general, the condition that ω is a weight function is exactly what we need to guarantee that every coefficient in $\Phi_S^\omega(\mathbf{x})$ is a finite number.

Example 4.3. Let $\mathbb{N}$ be the set of natural numbers, and define a weight function $\omega : \mathbb{N} \to \mathbb{N}$ by $\omega(n) = n$. Then

$$\Phi_{\mathbb{N}}^\omega(x) = \sum_{n \in \mathbb{N}} x^n = x^0 + x^1 + x^2 + x^3 + \cdots = \frac{1}{1 - x}$$

is the familiar geometric series.

Example 4.4. Let $S = N_6^3$ be the set of outcomes (a, b, c) when rolling three six–sided dice, one red, one white, and one green, where a, b, and c are the numbers of pips showing on the upper faces of the red, white, and green dice, respectively. Define a weight function $\omega : S \to \mathbb{N}^3$ as follows: for an outcome $(a, b, c) \in S$, let $\omega((a, b, c)) = (a, b, c)$. Then each outcome (a, b, c) contributes $x^a y^b z^c$ to the generating function $\Phi_S^\omega(x, y, z)$, and we have

$$\begin{aligned}\Phi_S^\omega(x,y,z) &= \sum_{(a,b,c)\in S} x^a y^b z^c = \sum_{a=1}^{6}\sum_{b=1}^{6}\sum_{c=1}^{6} x^a y^b z^c \\ &= \left(\sum_{a=1}^{6} x^a\right)\left(\sum_{b=1}^{6} y^b\right)\left(\sum_{c=1}^{6} z^c\right) = D(x)D(y)D(z)\end{aligned}$$

in which

$$D(t) = t + t^2 + \cdots + t^6 = \frac{t - t^7}{1 - t}.$$

Example 4.5. Let $S = N_6^3$ be as in the previous example, but now define the weight function $\nu : S \to \mathbb{N}^4$ by $\nu((a, b, c)) = (a, a + b, b, b + c)$. To keep track of the four pieces of information in our weight function ν we need four variables: w, x, y, and z. Now each outcome $(a, b, c) \in S$ contributes the monomial

$$(w, x, y, z)^{\nu(a,b,c)} = (w, x, y, z)^{(a,a+b,b,b+c)} = w^a x^{a+b} y^b z^{b+c} = (wx)^a (xyz)^b z^c$$

to the generating function. Hence

$$\Phi_S^\nu(w, x, y, z) = \sum_{(a,b,c)\in S} (wx)^a (xyz)^b z^c = D(wx)D(xyz)D(z)$$

with $D(t)$ as in Example 4.4.

The usefulness of the generating function approach to enumeration is that certain combinatorial properties of sets are reflected in the algebraic properties of generating functions. We now prove the most basic and applicable of these relations. The application of this technique to some interesting examples begins in Section 5.

Proposition 4.6. *Let S be a set with a weight function $\omega : S \to \mathbb{N}^r$. Then*

$$\Phi_S^\omega(\mathbf{x}) := \sum_{\alpha\in\mathbb{N}^r} (\#\omega^{-1}(\alpha))\mathbf{x}^\alpha.$$

Proof. This follows immediately by collecting all the terms which contribute $\mathbf{x}^\alpha$ in the definition of $\Phi_S^\omega(\mathbf{x})$, for each $\alpha \in \mathbb{N}^r$:

$$\begin{aligned}\Phi_S^\omega(\mathbf{x}) &= \sum_{s\in S} \mathbf{x}^{\omega(s)} = \sum_{\alpha\in\mathbb{N}^r} \sum_{s\in\omega^{-1}(\alpha)} \mathbf{x}^{\omega(s)} = \sum_{\alpha\in\mathbb{N}^r} \sum_{s\in\omega^{-1}(\alpha)} \mathbf{x}^{\alpha} \\ &= \sum_{\alpha\in\mathbb{N}^r} \mathbf{x}^\alpha \sum_{s\in\omega^{-1}(\alpha)} 1 = \sum_{\alpha\in\mathbb{N}^r} (\#\omega^{-1}(\alpha))\mathbf{x}^\alpha.\end{aligned}$$

□

For two sets S and T, Proposition 1.1 states that if they are finite then there is a bijection from S to T if and only if $\#S = \#T$. To lift this observation up to the level of generating functions we need a little more structure. Assume that S and T are sets equipped with weight functions $\omega : S \to \mathbb{N}^r$ and $\nu : T \to \mathbb{N}^r$. A function $f : S \to T$ is said to be *weight–preserving* when it satisfies the following condition:

[WP] For all $s \in S$, $\nu(f(s)) = \omega(s)$.

Weight–preserving bijections provide the most natural concept of equivalence when dealing with ordinary generating functions, as the following proposition shows.

Proposition 4.7. *Let S and T be sets with weight functions $\omega : S \to \mathbb{N}^r$ and $\nu : T \to \mathbb{N}^r$. There is a weight–preserving bijection $f : S \to T$ if and only if*

$$\Phi_S^\omega(\mathbf{x}) = \Phi_T^\nu(\mathbf{x}).$$

Proof. Assume that $f : S \to T$ is a weight–preserving bijection. By the definitions of ordinary generating function and weight–preserving bijection we calculate that

$$\Phi_S^\omega(\mathbf{x}) = \sum_{s\in S} \mathbf{x}^{\omega(s)} = \sum_{s\in S} \mathbf{x}^{\nu(f(s))} = \sum_{t\in T} \mathbf{x}^{\nu(t)} = \Phi_T^\nu(\mathbf{x}).$$

Conversely, assume that $\Phi_S^\omega(\mathbf{x}) = \Phi_T^\nu(\mathbf{x})$. Comparing the coefficient of $\mathbf{x}^\alpha$ on both sides of this equation, we see that $\#\omega^{-1}(\alpha) = \#\nu^{-1}(\alpha)$ for each $\alpha \in \mathbb{N}^r$. By Proposition 1.1, there is a bijection $f_\alpha : \omega^{-1}(\alpha) \to \nu^{-1}(\alpha)$. Now define a function $f : S \to T$ as follows: for $s \in S$ we let $f(s) := f_\alpha(s)$, in which $\alpha := \omega(s)$. One easily checks that $f : S \to T$ is a weight–preserving bijection. □

The following proposition is the analogue of Corollary 1.11 for ordinary generating functions.

Proposition 4.8 (The Sum Lemma)**.** *Let S be a set with a weight function $\omega : S \to \mathbb{N}^r$. Suppose that $S = S_1 \cup S_2 \cup \cdots$ is an expression of S as a union of pairwise disjoint subsets. Then*

$$\Phi_S^\omega(\mathbf{x}) = \sum_{i=1}^{\infty} \Phi_{S_i}^\omega(\mathbf{x}).$$

Proof. We calculate that

$$\Phi_S^\omega(\mathbf{x}) = \sum_{s\in S} \mathbf{x}^{\omega(s)} = \sum_{i=1}^{\infty} \sum_{s\in S_i} \mathbf{x}^{\omega(s)} = \sum_{i=1}^{\infty} \Phi_{S_i}^\omega(\mathbf{x}).$$

□

Similarly, we have the following analogue of Proposition 1.5.

Proposition 4.9 (The Product Lemma)**.** *Let S and T be sets with weight functions $\omega : S \to \mathbb{N}^r$ and $\nu : T \to \mathbb{N}^r$. Define a weight function φ on the Cartesian product of sets $S \times T$ by $\varphi(s,t) := \omega(s) + \nu(t)$. Then*

$$\Phi_{S\times T}^\varphi(\mathbf{x}) = \Phi_S^\omega(\mathbf{x}) \cdot \Phi_T^\nu(\mathbf{x}).$$

Proof.

$$\begin{aligned} \Phi_{S\times T}^\varphi(\mathbf{x}) &= \sum_{(s,t)\in S\times T} \mathbf{x}^{\varphi(s,t)} = \sum_{s\in S}\sum_{t\in T} \mathbf{x}^{\omega(s)+\nu(t)} \\ &= \left(\sum_{s\in S} \mathbf{x}^{\omega(s)}\right)\left(\sum_{t\in T} \mathbf{x}^{\nu(t)}\right) = \Phi_S^\omega(\mathbf{x}) \cdot \Phi_T^\nu(\mathbf{x}). \end{aligned}$$

□

(This may be extended to the Cartesian product of any finite number of sets – see Exercise 4.3.)

Example 4.10. We give another proof of the "Trinomial Theorem" of Exercise 3.4:

$$(x+y+z)^n = \sum_{i+j+k=n} \binom{n}{i,j,k} x^i y^j z^k.$$

Define $\omega : N_3 \to \mathbb{N}^3$ by $\omega(1) := (1,0,0)$, $\omega(2) := (0,1,0)$, and $\omega(3) := (0,0,1)$. Thus

$$\Phi_{N_3}^\omega(x,y,z) = x+y+z.$$

By the Product Lemma (extended to the Cartesian product of n sets) the LHS is the generating function of the set N_3^n with respect to the weight function on N_3^n induced from ω. Let $\mathcal{A}$ be the set of all ordered triples (A_1, A_2, A_3) such that each $A_i \subseteq N_n$, the sets A_1, A_2, and A_3 are pairwise disjoint, and $A_1 \cup A_2 \cup A_3 = N_n$. Define the weight of such a triple to be $\nu(A_1, A_2, A_3) := (\#A_1, \#A_2, \#A_3)$. By Exercise 2.4, the RHS of the equation is the generating function of $\mathcal{A}$ with respect to ν. To prove the formula we exhibit a weight–preserving bijection:

$$\begin{aligned} N_3^n &\rightleftharpoons \mathcal{A} \\ (b_1,\ldots,b_n) &\leftrightarrow (A_1,A_2,A_3) \\ \omega(b_1,\ldots,b_n) &= \nu(A_1,A_2,A_3) \end{aligned}$$

Given $(b_1, \ldots, b_n) \in N_3^n$ we define

$$A_i := \{j \in N_n : \ b_j = i\}$$

for each $i \in N_3$. Conversely, given $(A_1, A_2, A_3) \in \mathcal{A}$ and $j \in N_n$, we define b_j to equal the unique index $i \in N_3$ such that $j \in A_i$. One easily checks that these are mutually inverse weight–preserving bijections $N_3^n \rightleftharpoons \mathcal{A}$, which completes the proof.

Propositions 4.8 and 4.9 have as a consequence Proposition 4.13. The proof is straightforward after ironing out one technical wrinkle, and is left as an exercise.

Definition 4.11. Let S be a set with a weight function $\omega : S \to \mathbb{N}^r$. The set of all *finite strings on* S is denoted by S^* and defined to be

$$S^* := \bigcup_{k=0}^{\infty} S^k,$$

where $S^k = S \times \cdots \times S$ (k times) is the Cartesian product of k copies of S. In particular, $S^0 = \{\varepsilon\}$ in which ε denotes the *empty string* of length zero. Each $\sigma \in S^*$ is in S^k for exactly one $k \in \mathbb{N}$, which is called the *length* of σ and denoted by $\ell(\sigma)$. We define a function ω^* on S^* as follows. Suppose that $\sigma \in S^*$ has length k, so that $\sigma = (s_1, s_2, ..., s_k)$ and each $s_i \in S$. Then we put

$$\omega^*(\sigma) := \omega(s_1) + \cdots + \omega(s_k).$$

Lemma 4.12. *Let S be a set with weight function $\omega : S \to \mathbb{N}^r$. Then ω^* is a weight function on S^* if and only if $\omega(s) \neq \mathbf{0}$ for all $s \in S$.*

Proof. First, suppose that $\omega(z) = \mathbf{0}$ for some $z \in S$. Then for each $k \in \mathbb{N}$, the sequence $z^k = (z, z, ..., z)$ (k times) is in S^*. But by definition, $\omega^*(z^k) = k\omega(z) = \mathbf{0}$. Hence there are infinitely many $\sigma \in S^*$ for which $\omega^*(\sigma) = \mathbf{0}$, which contradicts the condition [WF] defining a weight function. Thus, in this case ω^* is not a weight function on S^*.

Conversely, suppose that $\omega(s) \neq \mathbf{0}$ for all $s \in S$. For any $\beta \in \mathbb{N}^r$, say $\beta = (b_1, b_2, ..., b_r)$, let $|\beta| := b_1 + b_2 + \cdots + b_r$. Since $\omega(s) \neq \mathbf{0}$ for all $s \in S$, we have for each $\sigma = (s_1, s_2, ..., s_k)$ in S^* that

$$|\omega^*(\sigma)| = |\omega(s_1)| + |\omega(s_2)| + \cdots + |\omega(s_k)| \geq k = \ell(\sigma).$$

That is, the length of σ is at most $|\omega^*(\sigma)|$.

To show that ω^* is a weight function on S^*, consider any $\alpha \in \mathbb{N}^r$; we will show that $\{\sigma \in S^* : \ \omega^*(\sigma) = \alpha\}$ is finite. From the previous paragraph, if $\omega^*(\sigma) = \alpha$ then the length of σ is at most $|\alpha|$. On the other hand, if $s \in S$ is an entry in σ and $\omega^*(\sigma) = \alpha$, then $\omega(s) \leq \alpha$. That is, s is an element of the set

$$U_\alpha := \bigcup_{\beta \leq \alpha} \omega^{-1}(\beta).$$

Since ω is a weight function, this is a finite union of finite sets, so U_α is finite. Now if $\sigma \in S^*$ is such that $\omega^*(\sigma) = \alpha$ then σ is a sequence of at most $|\alpha|$ letters from U_α. That is, σ is an element of the set

$$W_\alpha := \{\varnothing\} \cup U_\alpha \cup U_\alpha^2 \cup \cdots \cup U_\alpha^{|\alpha|}.$$

This is a finite union of finite sets, so W_α is finite. Finally, since $(\omega^*)^{-1}(\alpha) = \{\sigma \in S^* : \omega^*(\sigma) = \alpha\}$ is a subset of W_α, it is also a finite set, which completes the proof. □

Proposition 4.13 (The Finite String Lemma)**.** *Let S be a set with weight function ω such that $\omega(s) \neq \mathbf{0}$ for all $s \in S$. Then*

$$\Phi_{S^*}^{\omega^*}(\mathbf{x}) = \frac{1}{1 - \Phi_S^\omega(\mathbf{x})}.$$

---

4. Exercises.

---

1. Show that if S is a set and $\omega : S \to \mathbb{N}^r$ is a weight function then S is either finite or countably infinite.

---

2. Let S and T be as in the Product Lemma. Show that the function φ defined there is in fact a weight function on $S \times T$.

---

3. Extend the Product Lemma to the Cartesian product of any finite number of sets.

---

4. Prove the Finite String Lemma.

---

5. Let S be a set with a weight function ω, and let A_1, A_2, …, A_m be subsets of S. Generalize the Principle of Inclusion/Exclusion to give a formula for the ordinary generating function of $A_1 \cup \cdots \cup A_m$ with respect to ω.

---

6. Let S be a set with weight function $\omega : S \to \mathbb{N}^2$, and consider the generating function

$$\Phi_S^\omega(x, y) := \sum_{s \in S} x^{\omega_1(s)} y^{\omega_2(s)}.$$

By Proposition 4.6, for any $n \in \mathbb{N}$ the number of $s \in S$ with $\omega_1(s) = n$ equals $B_n := [x^n]\Phi_S^\omega(x, 1)$. Show that the average value of $\omega_2(s)$, among all B_n elements $s \in S$ such that $\omega_1(s) = n$, is equal to A_n/B_n in which

$$A_n := [x^n]\frac{\partial}{\partial y}\Phi_S^\omega(x, y)\Big|_{y=1}.$$

---

5. The q–Binomial Theorem.

In this section we re–examine some of the bijections constructed in Sections 2 and 3. These bijections are weight–preserving with respect to particular weight functions on the sets involved. Therefore, the material of Section 4 gives us some nice examples of generating function identites – the q–Binomial Theorem in particular.

Example 5.1. As in Examples 1.9 and 2.5, consider the set $\mathcal{P}(N_n)$ of all subsets of N_n. We define a weight function $\omega : \mathcal{P}(N_n) \to \mathbb{N}^2$ as follows: $\omega(S) := (\#S, \mathrm{sum}(S))$, in which $\mathrm{sum}(S) := \sum_{s\in S} s$. The generating function is

$$\Phi^{\omega}_{\mathcal{P}(N_n)}(x,q) := \sum_{S\subseteq N_n} x^{\#S} q^{\mathrm{sum}(S)}.$$

We can obtain a product formula for this summation as in the earlier examples, by noting that if $f = f_S$ is the characteristic function of $S \subseteq N_n$ then

$$\mathrm{sum}(S) = 1f(1) + 2f(2) + \cdots + nf(n).$$

Now we calculate that

$$\begin{aligned}
\sum_{S\subseteq N_n} x^{\#S} q^{\mathrm{sum}(S)} &= \sum_{(f_1,\ldots,f_n)\in\{0,1\}^n} x^{f_1+\cdots+f_n} q^{f_1+2f_2+\cdots+nf_n} \\
&= \left(\sum_{f_1\in\{0,1\}} x^{f_1} q^{f_1}\right)\left(\sum_{f_2\in\{0,1\}} x^{f_2} q^{2f_2}\right)\cdots\left(\sum_{f_n\in\{0,1\}} x^{f_n} q^{nf_n}\right) \\
&= (1+xq)(1+xq^2)\cdots(1+xq^n).
\end{aligned}$$

To obtain a proper generalization of the Binomial Theorem (Example 2.5) in the two–variable situation of Example 5.1, we should collect like powers of x: this will give

$$(1+xq)(1+xq^2)\cdots(1+xq^n) = \sum_{k=0}^{n} B_{n,k}(q) x^k$$

for some polynomials $B_{n,k}(q)$ in the variable q. Our goal in this section is to derive an algebraic formula for these polynomials. First of all, notice that by definition

$$B_{n,k}(q) := \sum_{A\in\mathcal{B}(n,k)} q^{\mathrm{sum}(A)} = \Phi^{\mathrm{sum}}_{\mathcal{B}(n,k)}(q).$$

Upon substituting $q = 1$ we see that

$$B_{n,k}(1) = \sum_{A \in \mathcal{B}(n,k)} 1 = \#\mathcal{B}(n,k) = \binom{n}{k}.$$

These polynomials $B_{n,k}(q)$ are therefore generalizations of binomial coefficients in a very natural sense. In fact, by keeping track of a little more information we can adapt the proof of Theorem 2.3 to obtain a formula for $B_{n,k}(q)$. As a first step we obtain a similar q–**analogue** of Theorem 2.1.

Definition 5.2. Let $\sigma = a_1 a_2 ... a_n \in \mathcal{S}_n$ be a permutation of length n. An *inversion in* σ is a pair (i, j) with $1 \leq i < j \leq n$ and $a_i > a_j$. That is, (i, j) is a pair of **places** in σ for which the corresponding entries are "in the wrong order". Let $\mathrm{inv}(\sigma)$ denote the number of inversions of σ. Notice that for any $\sigma \in \mathcal{S}_n$ we have $0 \leq \mathrm{inv}(\sigma) \leq \binom{n}{2}$.

Example 5.3. Consider $\sigma =$ 2 5 3 6 1 7 4 in $\mathcal{S}_7$. The inversions of this permutation are the pairs $(1,5)$, $(2,3)$, $(2,5)$, $(2,7)$, $(3,5)$, $(4,5)$, $(4,7)$, $(6,7)$. (Remember that the coordinates of an inversion (i, j) are indices or positions in σ, not the elements themselves!) Therefore, $\mathrm{inv}(\sigma) = 8$ in this case. For any $n \in \mathbb{N}$ there is exactly one $\sigma \in \mathcal{S}_n$ with $\mathrm{inv}(\sigma) = 0$, namely $\sigma =$ 1 2 ... $(n-1)$ n. There is also exactly one $\pi \in \mathcal{S}_n$ such that $\mathrm{inv}(\pi) = \binom{n}{2}$, namely $\pi = n\ (n-1)$ ... 2 1.

Theorem 5.4. *Let $n \in \mathbb{N}$. Then*

$$\Phi^{\mathrm{inv}}_{\mathcal{S}_n}(q) := \sum_{\sigma \in \mathcal{S}_n} q^{\mathrm{inv}(\sigma)} = [n]_q \cdot [n-1]_q \cdots [2]_q \cdot [1]_q,$$

in which for $k \in \mathbb{N}$ we define $[k]_q := 1 + q + \cdots + q^{k-1}$.

Proof. Recall the bijective correspondence $\mathcal{S}_n \rightleftharpoons \mathcal{Q}_n$ of Theorem 2.1, in which $\mathcal{Q}_n := N_n \times \cdots \times N_2 \times N_1$. We claim that if $I_n(\sigma) = (1 + r_1, 1 + r_2, ..., 1 + r_n)$ then

$$\mathrm{inv}(\sigma) = r_1 + r_2 + \cdots + r_n.$$

To see this, consider any index $1 \leq i \leq n$. By definition of the function I_n, then r_i is the number of indices j such that $i < j \leq n$ and $a_i > a_j$. That is, r_i is the number of inversions of σ which have i in the first coordinate. Summing over all $1 \leq i \leq n$ gives the total number of inversions of σ, establishing the claim. Thus we calculate that

$$\begin{aligned}
\sum_{\sigma \in \mathcal{S}_n} q^{\mathrm{inv}(\sigma)} &= \sum_{(h_1,...,h_n) \in \mathcal{Q}_n} q^{(h_1-1)+\cdots+(h_n-1)} \\
&= \left(\sum_{h_1 \in N_n} q^{h_1-1} \right) \left(\sum_{h_2 \in N_{n-1}} q^{h_2-1} \right) \cdots \left(\sum_{h_n \in N_1} q^{h_n-1} \right) \\
&= (1 + q + \cdots + q^{n-1})(1 + q + \cdots + q^{n-2}) \cdots (1),
\end{aligned}$$

as was to be shown. □

It is convenient to use the notation

$$[n]!_q := [n]_q[n-1]_q \cdots [2]_q[1]_q$$

for the polynomial appearing in Theorem 5.4, called the *q–factorial* of n. Now we can obtain a formula for $B_{n,k}(q)$.

Theorem 5.5. *Let $n \in \mathbb{N}$ and $0 \leq k \leq n$. Then*

$$B_{n,k}(q) := \sum_{A \in \mathcal{B}(n,k)} q^{\mathrm{sum}(A)} = q^{k(k+1)/2} \frac{[n]!_q}{[k]!_q[n-k]!_q}.$$

Proof. Recall the bijective correspondence $\mathcal{S}_n \rightleftharpoons \mathcal{B}(n,k) \times \mathcal{S}_k \times \mathcal{S}_{n-k}$ of Theorem 2.3. We claim that if $\Psi_{n,k}(\sigma) = (A, \beta, \gamma)$ then

$$\mathrm{inv}(\sigma) = \left[\mathrm{sum}(A) - \frac{k(k+1)}{2}\right] + \mathrm{inv}(\beta) + \mathrm{inv}(\gamma).$$

To see this, separate the inversions (i,j) of σ into three sets:

$$\begin{aligned} E_1 &:= \{ \text{ inversions } (i,j) : \ i < j \leq k\} \\ E_2 &:= \{ \text{ inversions } (i,j) : \ k+1 \leq i < j\} \\ E_3 &:= \{ \text{ inversions } (i,j) : \ i \leq k \text{ and } k+1 \leq j\} \end{aligned}$$

Thus $\mathrm{inv}(\sigma) = \#E_1 + \#E_2 + \#E_3$. An inversion (i,j) in E_1 corresponds to a pair of elements in σ such that $a_i > a_j$ and $i < j \leq k$. Now consider $\beta = b_1 b_2 ... b_k := P_k(a_1 a_2 ... a_k) \in \mathcal{S}_k$. Since the function P_k preserves the relative order of the entries of its input, we have $b_i > b_j$ and $i < j$ as well. Furthermore, every inversion of β corresponds to exactly one inversion of σ in the set E_1. Therefore, $\#E_1 = \mathrm{inv}(\beta)$. A similar argument shows that $\#E_2 = \mathrm{inv}(\gamma)$.

It remains to show that $\#E_3 = \mathrm{sum}(A) - k(k+1)/2$. An inversion (i,j) in E_3 corresponds to a pair of elements in σ such that $a_i > a_j$ and $i \leq k$ and $k+1 \leq j$. Since $A := \{a_1, ..., a_k\}$, this is a pair of elements $(a,z) \in N_n \times N_n$ such that $a \in A$, $z \in N_n \setminus A$, and $a > z$. Thus $\#E_3$ is the size of this set.

Lemma 5.6. *Fix $n \in \mathbb{N}$ and $0 \leq k \leq n$, and let $A \in \mathcal{B}(n,k)$. If*

$$E_A := \{(a,z) \in N_n \times N_n : \ a \in A, \ z \notin A, \text{ and } a > z\}$$

then $\#E_A = \mathrm{sum}(A) - k(k+1)/2$.

Proof. Let A be a k–element subset of N_n, and sort the elements of $A = \{s_1, \ldots, s_k\}$ so that $s_1 < s_2 < \cdots < s_k$. For each $1 \leq i \leq k$, there are i elements of A which are less than or equal to s_i, and there are s_i elements of N_n which are less than or equal to s_i. Thus, for each $1 \leq i \leq k$ there are $s_i - i$ elements of $N_n \setminus A$ which are less than s_i. It follows that

$$\#E_A = (s_1 - 1) + (s_2 - 2) + \cdots + (s_k - k) = \mathrm{sum}(A) - \frac{k(k+1)}{2}.$$

□

Now that we have expressed $\mathrm{inv}(\sigma)$ in terms of (A, β, γ) we calculate that

$$\begin{aligned}\sum_{\sigma\in\mathcal{S}_n} q^{\mathrm{inv}(\sigma)} &= \sum_{(A,\beta,\gamma)} q^{[\mathrm{sum}(A)-k(k+1)/2]+\mathrm{inv}(\beta)+\mathrm{inv}(\gamma)} \\ &= \left(\sum_{A\in\mathcal{B}(n,k)} q^{\mathrm{sum}(A)-k(k+1)/2}\right)\left(\sum_{\beta\in\mathcal{S}_k} q^{\mathrm{inv}(\beta)}\right)\left(\sum_{\gamma\in\mathcal{S}_{n-k}} q^{\mathrm{inv}(\gamma)}\right).\end{aligned}$$

By Theorem 5.4 and the definition of $B_{n,k}(q)$ we see that

$$[n]!_q = q^{-k(k+1)/2} B_{n,k}(q)[k]!_q[n-k]!_q$$

from which the result follows. □

The *Gaussian polynomials*, or *q–binomial coefficients* are defined for $n \in \mathbb{N}$ and $0 \le k \le n$ to be the polynomials

$$\left[\begin{array}{c} n \\ k \end{array}\right]_q := \frac{[n]!_q}{[k]!_q[n-k]!_q}.$$

By combining Theorem 5.5 with the remarks after Example 5.1 we obtain the q–Binomial Theorem.

Theorem 5.7 (The q–Binomial Theorem)**.** *For each $n \in \mathbb{N}$,*

$$(1+qx)(1+q^2x)\cdots(1+q^nx) = \sum_{k=0}^{n} q^{k(k+1)/2}\left[\begin{array}{c} n \\ k \end{array}\right]_q x^k.$$

The Gaussian polynomials have two other very nice combinatorial interpretations, which is why they are used instead of the polynomials $B_{n,k}(q)$.

Theorem 5.8. *Fix $a, b \in \mathbb{N}$, and consider the set $\mathcal{L}(a,b)$ of lattice paths from $(0,0)$ to (a,b). For any such path P let $\mathrm{area}(P)$ denote the area of the compact region enclosed by P and the line segments $(0,0) \to (0,b)$ and $(0,b) \to (a,b)$. Then*

$$\Phi^{\mathrm{area}}_{\mathcal{L}(a,b)} := \sum_{P\in\mathcal{L}(a,b)} q^{\mathrm{area}(P)} = \left[\begin{array}{c} a+b \\ b \end{array}\right]_q.$$

Theorem 5.9. *Fix $n \in \mathbb{N}$ and $0 \le k \le n$, and let $q = p^c$ be a prime power. Let $\mathbb{F}_q$ denote the finite field with q elements. Then the number of k–dimensional subspaces of an n–dimensional vector space over $\mathbb{F}_q$ is $\left[\begin{array}{c} n \\ k \end{array}\right]_q$.*

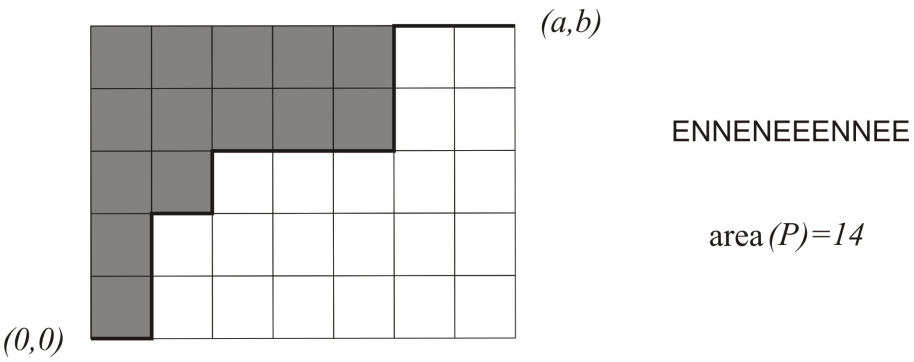


FIGURE 5.1. the area of a lattice path.

5. Exercises.

1. Show that for $a, b \in \mathbb{N}$:

$$\left[\begin{array}{c} a+1+b \\ b \end{array} \right]_q = \sum_{j=0}^{b} q^{(a+1)(b-j)} \left[\begin{array}{c} a+j \\ j \end{array} \right]_q .$$

2. Show that for $n \in \mathbb{N}$:

$$\left[\begin{array}{c} 2n \\ n \end{array} \right]_q = \sum_{k=0}^{n} q^{k^2} \left[\begin{array}{c} n \\ k \end{array} \right]_q^2 .$$

3. Show that for $k, m, n \in \mathbb{N}$:

$$\left[\begin{array}{c} m+n \\ k \end{array} \right]_q = \sum_{j=0}^{k} q^{(m-j)(k-j)} \left[\begin{array}{c} m \\ j \end{array} \right]_q \left[\begin{array}{c} n \\ k-j \end{array} \right]_q .$$

(This generalizes Exercise 3.5.)

4. Exercise 3.6 implies that for all $a, b, n \in \mathbb{N}$:

$$\binom{b+1+n}{n} = \sum_{j=0}^{n} \binom{a+j}{j} \binom{b-a+n-j}{n-j}.$$

Prove a generalization of this involving q–binomial coefficients.

5. Prove a generalization of Exercise 3.9 involving q–binomial coefficients.

6(a) Let $0 \leq a \leq c \leq b$ be integers. Show that

$$\binom{a+b}{b} = \sum_{j=0}^{a} \binom{c}{c-j} \binom{a+b-c}{j+b-c}.$$

6(b) State (without proof) a generalization of the formula in part (a) which involves q–binomial coefficients.

---

7. Prove Theorem 5.8.

---

8. Prove Theorem 5.9. (To get started, first show that the number of ordered bases of an m–dimensional vector space over $\mathbb{F}_q$ is

$$(q^m - 1)(q^m - q)(q^m - q^2) \cdots (q^m - q^{m-1}),$$

then mimic the proof of Theorem 2.3.)

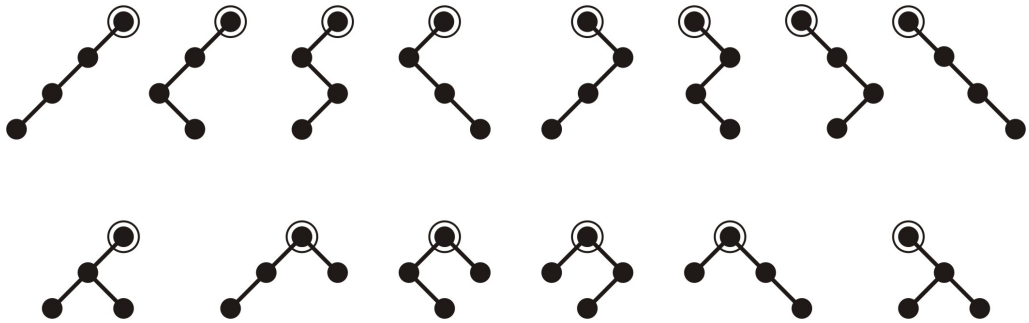


FIGURE 6.1. the BRTs with four nodes.

6. Recursive Structure.

Recursive structure is a term used to describe objects that can be built up out of smaller pieces, each piece resembling the larger object in the way it is constructed. This kind of of structure leads to a powerful method for enumerating such objects. In this section we will see a few simple examples of this method.

Definition 6.1. A *binary rooted tree* is a tree, with a designated node $\odot$ called the *root node*, which is drawn in the plane such that each node has at most two children, one to the left and one to the right, if they exist. If a node has only one child then it is still labelled either left or right. A *terminal* is a node with no children. A *leaf* is a node of degree at most one. Let $n(T)$ denote the number of nodes of the binary rooted tree T, let $\tau(T)$ denote the number of terminals of T, and let $\ell(T)$ denote the number of leaves of T. See Figure 6.1.

Notice that in a binary rooted tree (BRT) every terminal node is a leaf, and that if there is a leaf node which is not a terminal then it is the root node and has degree one. In short, the concepts of leaf node and terminal node are almost synonymous, the only difference being possibly at the root node.

BRTs are commonly used as data structures in computer science, especially in sorting algorithms and other applications in which pairwise operations between elements are important. Unfortunately, we don't have time to go into that. Anyway, here is a question which is kind of interesting.

Question 6.2. Given a random binary rooted tree T with n nodes, what is the expected number of terminals of T?

By "random", I mean that among all the BRTs with n nodes, choose one in such a way that each tree is equally likely. The answer to such a question might be

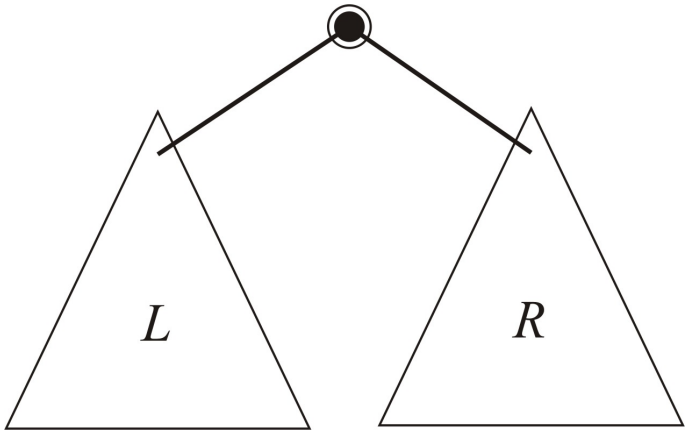


FIGURE 6.2. the recursive structure of BRTs.

important in determining the average case running–time for some algorithm which takes input based on a BRT data structure and spends most of its time processing at the terminals of the tree. (See Figure 16.1, for example.)

Let T_n denote the number of BRTs with n nodes, and let A_n denote the total number of terminals among all BRTs with n nodes. The expected number of terminals among all BRTs with n nodes is A_n/T_n, by elementary probability theory. For example, with $n = 4$ the expected number of terminals is $20/14$, which is roughly 1.4286 terminals per tree. Now we'll figure out what T_n and A_n are, and hence what this average number of terminals is, for all $n \in \mathbb{N}$.

Let $\mathcal{T}$ denote the set of all BRTs, and consider the generating function

$$T(x) := \Phi^n_{\mathcal{T}}(x) = \sum_{T \in \mathcal{T}} x^{n(T)} = \sum_{n=0}^{\infty} T_n x^n.$$

We'll get a formula for $T(x)$ and use this to extract the coefficients T_n. To get this formula we use the recursive structure of binary rooted trees.

Given a BRT T, we remove the root node $\odot$ of T to get an ordered pair of "subtrees" (L, R) which are either empty or are themselves BRTs. (See Figure 6.2.) The subtree L is rooted at the left child of $\odot$ if it exists, otherwise L is empty. Similarly, the subtree R is rooted at the right child of $\odot$ if it exists, otherwise R is empty. Conversely, given any ordered pair (L, R) in which each of L and R is either $\varnothing$ or is a BRT, adjoin a new root node $\odot$ which has as its children the root of L (to the left) and the root of R (to the right) if these subtrees are not empty. These constructions give a bijection

$$\mathcal{T} \rightleftharpoons \{\odot\} \times (\mathcal{T} \cup \{\varnothing\}) \times (\mathcal{T} \cup \{\varnothing\}).$$

In this bijection, if T corresponds to $(\odot, L, R)$ then

$$n(T) = 1 + n(L) + n(R),$$

since the root node contributes 1, L contributes $n(L)$, and R contributes $n(R)$ to the number of nodes of T. Therefore,

$$\begin{aligned} T(x) = \sum_{T \in \mathcal{T}} x^{n(T)} &= \sum_{(L,R) \in (\mathcal{T} \cup \{\varnothing\})^2} x^{1+n(L)+n(R)} \\ &= x \left(\sum_{L \in \mathcal{T} \cup \{\varnothing\}} x^{n(L)} \right) \left(\sum_{R \in \mathcal{T} \cup \{\varnothing\}} x^{n(R)} \right) \\ &= x(1 + T(x))(1 + T(x)). \end{aligned}$$

Therefore, the generating function $T(x)$ for binary rooted trees with respect to number of nodes satisfies the functional equation

$$xT^2 + (2x - 1)T + x = 0.$$

We want to solve this equation for $T = T(x)$. To do this, we use the Quadratic Formula just as we would if we were solving it for a complex value. (The justification of such manipulations with formal power series is the subject of Section 7.) The solutions are thus

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

Notice that we get two solutions, depending on a choice for the $\pm$ sign. These cannot both be the true generating function for BRTs, so we'll have to rule one of them out later.

To finish finding the number T_n of BRTs with n nodes, we want to expand $T(x)$ in powers of x. To do this, we have to expand $\sqrt{1 - 4x}$ in powers of x. The Binomial Series expansion of Example 3.7 provides this, and it is worthwhile to see a second derivation of it.

Definition 6.3 (The Binomial Series Expansion)**.** For $\alpha \in \mathbb{C}$ and $k \in \mathbb{N}$, let

$$\binom{\alpha}{k} := \frac{\alpha(\alpha - 1) \cdots (\alpha - k + 1)}{k!}.$$

(Notice that if $\alpha \in \mathbb{N}$ then this agrees with our earlier definition of $\binom{\alpha}{k}$ in Theorem 2.3.) By taking the Taylor series expansion of the function $(1 + x)^\alpha$ about the point $x = 0$ (which is valid since $(1 + x)^\alpha$ is analytic in the disc $|x| < 1$) we obtain the *binomial series expansion*

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

This is true since the Taylor series coefficients are in this case

$$\frac{1}{k!}\frac{d^k}{dx^k}(1+x)^\alpha\bigg|_{x=0} = \binom{\alpha}{k}$$

for all $k \in \mathbb{N}$. This identity remains valid when x is an indeterminate, or variable. The only time that this expansion may not be used is when x is a complex number with $|x| \geq 1$. Thus, it is safe to use the binomial series expansion in a generating function, since our variables are not complex numbers but are formal indeterminates instead.

Now, back to the task at hand. To find $T(x)$ we need to expand $\sqrt{1-4x}$. From the Binomial Series expansion we have

$$\sqrt{1-4x} = (1+(-4x))^{1/2} = \sum_{k=0}^{\infty}\binom{1/2}{k}(-4x)^k.$$

For $k \geq 1$ we can simplify the coefficient of x^k as follows:

$$\begin{aligned}
(-4)^k\binom{1/2}{k} &= (-1)^k 2^{2k}\frac{(\frac{1}{2})(\frac{1}{2}-1)\cdots(\frac{1}{2}-k+1)}{k!} \\
&= (-1)^k 2^k \frac{(1)(1-2)(1-4)\cdots(1-2k+2)}{k!} \\
&= -2^k\frac{(1)(3)(5)\cdots(2k-3)}{k!} \\
&= -2\left(\frac{2\cdot 4\cdot 6\cdots(2k-2)}{(k-1)!}\right)\left(\frac{1\cdot 3\cdot 5\cdots(2k-3)}{k!}\right) \\
&= -\frac{2}{k}\binom{2k-2}{k-1}.
\end{aligned}$$

Of course, for $k=0$ we have $\binom{1/2}{0}(-4)^0 = 1$. Therefore,

$$\sqrt{1-4x} = \sum_{k=0}^{\infty}\binom{1/2}{k}(-4x)^k = 1 - 2\sum_{n=0}^{\infty}\frac{1}{n+1}\binom{2n}{n}x^{n+1}.$$

Finally, this gives us the formula we desire for $T(x)$:

$$\begin{aligned}
T(x) &= \frac{1}{2x} - 1 \pm \frac{1}{2x}\sqrt{1-4x} \\
&= \frac{1}{2x} - 1 \pm \frac{1}{2x}\left(1 - 2\sum_{n=0}^{\infty}\frac{1}{n+1}\binom{2n}{n}x^{n+1}\right).
\end{aligned}$$

Now it is clear which sign to choose in this formula for $T(x)$. If we choose the $+$ sign then the first few terms of the series are

$$\frac{1}{x} - 2 - x - 2x^2 - 5x^3 - 14x^4 - \cdots .$$

This makes no sense as a generating function, for two reasons. First, some of the coefficients of powers of x are negative – these coefficients are supposed to be cardinalities of finite sets, and hence should be nonnegative integers. Secondly, the term x^{-1} appears, which in this case would indicate that there is one BRT which has the number -1 of nodes. This is obviously absurd! Thus, we must take the $-$ sign in our expansion of $T(x)$, and we get

$$\begin{aligned} T(x) &= \sum_{n=1}^{\infty} \frac{1}{n+1}\binom{2n}{n} x^n \\ &= x + 2x^2 + 5x^3 + 14x^4 + 42x^5 + 132x^6 + \cdots . \end{aligned}$$

This makes much better sense. We have proved the following result.

Theorem 6.4. *For each $n \geq 1$, the number of binary rooted trees with n nodes is*

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

The numbers $\frac{1}{n+1}\binom{2n}{n}$, which arise frequently, are called *Catalan numbers*.

We are halfway towards answering Question 6.2; it remains to find A_n, the total number of terminals among all BRTs with n nodes. To do this, consider the two–variable generating function

$$T(x, y) := \sum_{T \in \mathcal{T}} x^{n(T)} y^{\tau(T)}$$

which encodes information about both the number of nodes and the number of terminals of BRTs. We can find A_n from this series, as follows. Notice that by definition

$$A_n := \sum_{T \in \mathcal{T}:\ n(T)=n} \tau(T).$$

Thus,

$$\left.\frac{\partial}{\partial y} T(x, y)\right|_{y=1} = \sum_{T \in \mathcal{T}} x^{n(T)} \tau(T) 1^{\tau(T)-1} = \sum_{n=0}^{\infty} \left(\sum_{T \in \mathcal{T}:\ n(T)=n} \tau(T) \right) x^n = \sum_{n=0}^{\infty} A_n x^n.$$

To find the number A_n we'll get an algebraic formula for $T(x, y)$ and then compute $[x^n]\partial T(x, y)/\partial y|_{y=1}$.

Referring to the recursive structure of BRTs in Figure 6.2, we see that if T has left subtree L and right subtree R then

$$\tau(T) = \begin{cases} \tau(L) + \tau(R) & \text{if either } L \neq \varnothing \text{ or } R \neq \varnothing, \\ 1 & \text{if } L = \varnothing \text{ and } R = \varnothing. \end{cases}$$

Now, using the bijection $\mathcal{T} \rightleftharpoons \{\odot\} \times (\mathcal{T} \cup \{\varnothing\})^2$ we calculate that

$$\begin{aligned} T(x,y) &= \sum_{T \in \mathcal{T}} x^{n(T)} y^{\tau(T)} \\ &= xy + x \sum_{L \in \mathcal{T}} x^{n(L)} y^{\tau(L)} + x \sum_{R \in \mathcal{T}} x^{n(R)} y^{\tau(R)} + x \sum_{(L,R) \in \mathcal{T} \times \mathcal{T}} x^{n(L)+n(R)} y^{\tau(L)+\tau(R)} \\ &= x(y + 2T(x,y) + T(x,y)^2). \end{aligned}$$

Now we solve for $T(x,y)$ using the Quadratic Formula. The equation is

$$xT^2 + (2x-1)T + xy = 0$$

so that

$$\begin{aligned} T(x,y) &= \frac{1 - 2x \pm \sqrt{(1-2x)^2 - 4x^2 y}}{2x} \\ &= \frac{1}{2x} - 1 \pm \frac{1}{2x}\sqrt{(1-2x)^2 - 4x^2 y} \end{aligned}$$

When we substitute $y = 1$ in $T(x,y)$ we get the generating function $T(x)$ which we considered earlier; thus, we must take the $-$ sign in this formula.

$$T(x,y) = \frac{1}{2x} - 1 - \frac{1}{2x}\sqrt{(1-2x)^2 - 4x^2 y}.$$

Now,

$$\begin{aligned} \frac{\partial}{\partial y} T(x,y) &= \left(\frac{-1}{2x}\right)\left(\frac{1}{2}\right)\left((1-2x)^2 - 4x^2 y\right)^{-1/2}(-4x^2) \\ &= \frac{x}{\sqrt{(1-2x)^2 - 4x^2 y}}. \end{aligned}$$

We set $y = 1$ to obtain

$$\left.\frac{\partial}{\partial y} T(x,y)\right|_{y=1} = \frac{x}{\sqrt{1-4x}}.$$

Expanding $(1-4x)^{-1/2}$ as a Binomial Series we get

$$(1-4x)^{-1/2} = \sum_{k=0}^{\infty} \binom{-1/2}{k} (-4x)^k.$$

We can simplify the coefficient of x^k in this as follows:

$$\begin{aligned}\binom{-1/2}{k}(-4)^k &= (-1)^k 2^{2k}\frac{(-\frac{1}{2})(-\frac{1}{2}-1)\cdots(-\frac{1}{2}-k+1)}{k!}\\ &= 2^k\frac{(1)(1+2)(1+4)\cdots(1+2k-2)}{k!}\\ &= \left(\frac{2\cdot 4\cdot 6\cdots(2k)}{k!}\right)\left(\frac{1\cdot 3\cdot 5\cdots(2k-1)}{k!}\right)\\ &= \binom{2k}{k}.\end{aligned}$$

Therefore,

$$\frac{1}{\sqrt{1-4x}} = \sum_{k=0}^{\infty}\binom{2k}{k}x^k.$$

These series expansions for $(1-4x)^{1/2}$ and $(1-4x)^{-1/2}$ are worth remembering.

Proposition 6.5.

$$\text{(a)}\qquad \sqrt{1-4x} = 1 - 2\sum_{n=0}^{\infty}\frac{1}{n+1}\binom{2n}{n}x^{n+1}.$$

$$\text{(b)}\qquad \frac{1}{\sqrt{1-4x}} = \sum_{n=0}^{\infty}\binom{2n}{n}x^n.$$

Referring to the above calculations, we see that

$$\left.\frac{\partial}{\partial y}T(x,y)\right|_{y=1} = \frac{x}{\sqrt{1-4x}} = \sum_{k=0}^{\infty}\binom{2k}{k}x^{k+1} = \sum_{n=1}^{\infty}\binom{2n-2}{n-1}x^n.$$

That is, for all $n \geq 1$ the total number of terminals among all BRTs with n nodes is $A_n = \binom{2n-2}{n-1}$.

Finally, we can answer Question 6.2. The expected number of terminals among all BRTs with n nodes is

$$\frac{A_n}{T_n} = \frac{\binom{2n-2}{n-1}}{\frac{1}{n+1}\binom{2n}{n}} = \frac{(n+1)(2n-2)!n!n!}{(n-1)!(n-1)!(2n)!} = \frac{n^2+n}{4n-2}.$$

As $n \to \infty$ this ratio is asymptotic to $n/4$, in the sense that

$$\lim_{n\to\infty}\left(\frac{n^2+n}{4n-2}\right)\Big/\left(\frac{n}{4}\right) = 1.$$

To be vague and intuitive about this statement, it says that in a large random binary rooted tree one expects roughly $1/4$ of the nodes to be terminals. That is something which you might have guessed based on intuition. I doubt, however, that you would have guessed the exact answer $(n^2+n)/(4n-2)$.

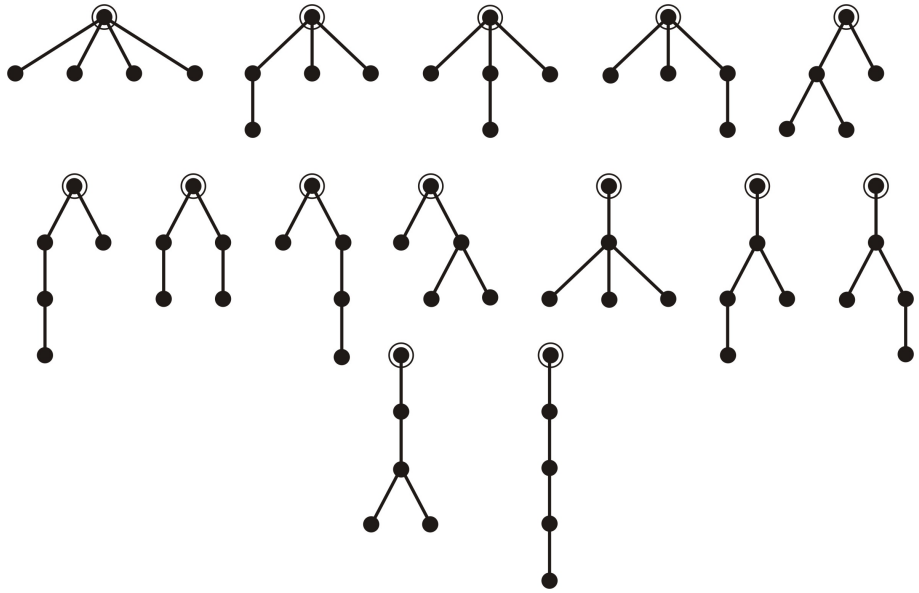


FIGURE 6.3. the PPTs with five nodes.

In the answer to Question 6.2 we used the recursive structure of BRTs, the Quadratic Formula, and some manipulations with Binomial Series. The recursive structure was used solely to derive functional equations satisfied by the generating functions $T(x)$ and $T(x, y)$. After that, the argument was purely algebraic. Now let's consider some more examples in which recursive structure is relevant, concentrating on the combinatorics rather than on the algebra.

Definition 6.6. A *plane planted tree* T is a tree drawn in the plane, which has a root node $\odot$, and is such that the children of each node are ordered from left to right. In particular, if a node has only one child then it is not specified to be a left child or a right child. See Figure 6.3.

Theorem 6.7. *For each $n \geq 1$, the number of plane planted trees with n nodes is*

$$\frac{1}{n}\binom{2n-2}{n-1}.$$

Proof. Let $\mathcal{U}$ denote the set of all plane planted trees (PPTs), and let

$$U(x) := \Phi^n_{\mathcal{U}}(x) = \sum_{T \in \mathcal{U}} x^{n(T)}.$$

The recursive structure of PPTs is described as follows. Consider any PPT T, and let $\odot$ be the root node of T. If we remove $\odot$ from T then we get an ordered sequence of subtrees $(T_1, T_2, ..., T_d)$ for some $d \in \mathbb{N}$. For each $1 \leq i \leq d$ the subtree T_i is rooted at the i–th child of $\odot$, starting from the left. (See Figure 6.4.) Each of these subtrees

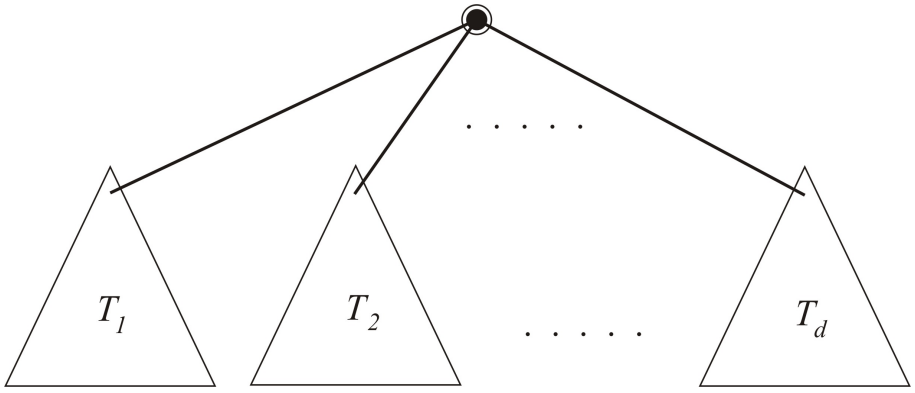


FIGURE 6.4. the recursive structure of PPTs.

is not empty, and is in fact a PPT. Conversely, given any finite sequence of PPTs $(T_1, T_2, ..., T_d)$ we can adjoin a new root node with children being the root nodes of $T_1, T_2, ..., T_d$ in that order from left to right.

These constructions show that we have a bijective correspondence

$$\mathcal{U} \rightleftharpoons \{\odot\} \times \left(\mathcal{U}^0 \cup \mathcal{U}^1 \cup \mathcal{U}^2 \cup \cdots\right) = \{\odot\} \times \mathcal{U}^*.$$

In this bijection, if the PPT T has a root node with d descendants then it corresponds to a $(d+1)$–tuple $(\odot, T_1, T_2, ..., T_d)$ in which $\odot$ is the root node and the T_i are the PPTs rooted at the children of $\odot$. The number of nodes of T is

$$n(T) = 1 + n(T_1) + n(T_2) + \cdots + n(T_d).$$

Thus we calculate that

$$\begin{aligned} U(x) = \sum_{T \in \mathcal{U}} x^{n(T)} &= \sum_{d=0}^{\infty} \sum_{(T_1, T_2, ..., T_d) \in \mathcal{U}^d} x^{1+n(T_1)+\cdots+n(T_d)} \\ &= x \sum_{d=0}^{\infty} \left(\sum_{T_1 \in \mathcal{U}} x^{n(T_1)}\right)\left(\sum_{T_2 \in \mathcal{U}} x^{n(T_2)}\right) \cdots \left(\sum_{T_d \in \mathcal{U}} x^{n(T_d)}\right) \\ &= x \sum_{d=0}^{\infty} U(x)^d = \frac{x}{1 - U(x)}. \end{aligned}$$

Notice that this could have been deduced directly from the bijection $\mathcal{U} \rightleftharpoons \{\odot\} \times \mathcal{U}^*$ by using Propositions 4.9 and 4.13.

This functional equation for $U(x)$ may be written as

$$U^2 - U + x = 0,$$

which we solve by the Quadratic Formula to get

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

(we will decide on the correct choice of signs in a moment). By Proposition 6.5 we expand this as

$$U(x) = \frac{1}{2} \pm \frac{1}{2}\left(1 - 2\sum_{k=0}^{\infty} \frac{1}{k+1}\binom{2k}{k}x^{k+1}\right).$$

Since every PPT has at least one node (*i.e.* the root node) the coefficient of x^0 in $U(x)$ is 0. Thus, we must take the $-$ sign in this equation, resulting in

$$U(x) = \sum_{k=0}^{\infty} \frac{1}{k+1}\binom{2k}{k}x^{k+1} = \sum_{n=1}^{\infty} \frac{1}{n}\binom{2n-2}{n-1}x^n.$$

Therefore, for each $n \geq 1$ the number of PPTs with n nodes is $\frac{1}{n}\binom{2n-2}{n-1}$, as claimed. □

Definition 6.8. A *super–diagonal lattice path* (or *Dyck path*) is a path in $\mathbb{N} \times \mathbb{N}$ which starts at $(0,0)$ and ends at (n,n) for some $n \in \mathbb{N}$, which uses steps of the form $\mathsf{E} = (1,0)$ and $\mathsf{N} = (0,1)$, and which always stays on or above the line $x = y$. We think of P as being a sequence in $\{\mathsf{E},\mathsf{N}\}^*$ that describes a path staying above the line $x = y$. If P is a super–diagonal lattice path then $n(P)$ denotes half the length of P, so that P ends at the point $(n(P), n(P))$.

Theorem 6.9. *For each $n \geq 0$, the number of super–diagonal lattice paths from $(0,0)$ to (n,n) is*

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

Proof. Let $\mathcal{V}$ denote the set of all super–diagonal lattice paths (SDLPs), and let

$$V(x) := \Phi^n_{\mathcal{V}}(x) = \sum_{P \in \mathcal{V}} x^{n(P)}.$$

The recursive structure of SDLPs is described as follows. Let P be any SDLP, and consider the points at which P meets the line $x = y$: call them

$$(0,0) = (k_0,k_0), (k_1,k_1), ..., (k_{r-1},k_{r-1}), (k_r,k_r) = (n(P),n(P)).$$

If P is not the empty path ε then $r \geq 1$, and between any two consecutive diagonal points (k_{i-1},k_{i-1}) and (k_i,k_i) that portion of P looks like $\mathsf{N}Q_i\mathsf{E}$, in which Q_i is also a SDLP. The reason that Q_i is a SDLP is that it must not touch the line $x = y$, since the points (k_{i-1},k_{i-1}) and (k_i,k_i) were supposed to be consecutive diagonal points of P. Note that Q_i may be the empty walk ε. In the remaining case that $P = \varepsilon$ is the empty walk, $r = 0$ and there is nothing else to say.

Conversely, given any finite sequence $(Q_1, Q_2, ..., Q_r)$ of SDLPs, we can form the sequence

$$P := \mathsf{N}Q_1\mathsf{E}\mathsf{N}Q_2\mathsf{E}\cdots\mathsf{N}Q_r\mathsf{E},$$

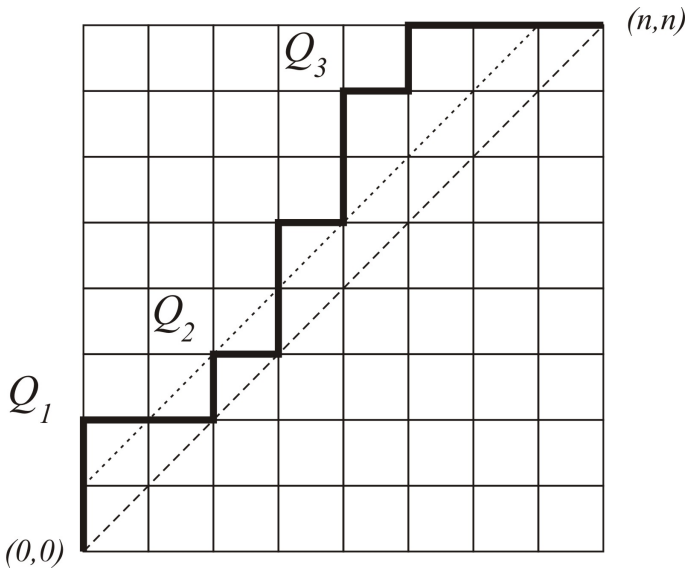


FIGURE 6.5. the recursive structure of SDLPs.

which describes a SDLP. (In the case $r = 0$ this construction produces the empty walk $P = \varepsilon$. If $r = 1$ and $Q_1 = \varepsilon$ then $P = \mathsf{NE}$.) Every SDLP is constructed exactly once in this way, and

$$n(P) = (n(Q_1) + 1) + (n(Q_2) + 1) + \cdots + (n(Q_r) + 1).$$

That is, we have a weight–preserving bijection

$$\mathcal{V} \rightleftharpoons \bigcup_{r=0}^{\infty} (\{\mathsf{N}\} \times \mathcal{V} \times \{\mathsf{E}\})^r = (\{\mathsf{N}\} \times \mathcal{V} \times \{\mathsf{E}\})^* .$$

Now, we calculate that

$$\begin{aligned} V(x) = \sum_{P \in \mathcal{V}} x^{n(P)} &= \sum_{r=0}^{\infty} \left(\sum_{Q_1 \in \mathcal{V}} x^{n(Q_1)+1} \right) \left(\sum_{Q_2 \in \mathcal{V}} x^{n(Q_2)+1} \right) \cdots \left(\sum_{Q_r \in \mathcal{V}} x^{n(Q_r)+1} \right) \\ &= \sum_{r=0}^{\infty} (xV(x))^r = \frac{1}{1 - xV(x)}. \end{aligned}$$

Notice that this could have been deduced directly from Propositions 4.9 and 4.13.

This functional equation for $V(x)$ may be written as

$$xV^2 - V + 1 = 0,$$

which we solve by the Quadratic Formula to get

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

Now we expand $\sqrt{1-4x}$ by Proposition 6.5, and see that we must choose the $-$ sign in order to get the coefficient of x^{-1} to be zero. The result is that

$$V(x) = \sum_{n=0}^{\infty} \frac{1}{n+1}\binom{2n}{n}x^n,$$

which shows that for each $n \in \mathbb{N}$ there are $\frac{1}{n+1}\binom{2n}{n}$ SDLPs from $(0,0)$ to (n,n). $\square$

We have now seen three seperate situations in which the Catalan numbers $\frac{1}{n+1}\binom{2n}{n}$ arise: BRTs, PPTs, and SDLPs. Since sets of these objects have the same cardinality, Proposition 1.1 ensures that there exist bijections between these sets of objects, but our algebraic method gives no hint as to how such bijections might be constructed. This situation is a bit of a tease to the combinatorially inclined, since the most satisfactory explanation for two sets being equicardinal is to exhibit a bijection between the sets. Thus, for completeness, we will briefly describe without proof some bijections which connect the sets of BRTs, PPTs, and SDLPs.

Definition 6.10. A *well–formed parenthesization* is a sequence of the symbols (and) which "group together" using the usual rules for closing pairs of parentheses; for example, (()()))(() is not well–formed, but (()(()))() is. In particular, in a well–formed parenthesization (WFP) there must be the same number of (s as there are)s. If π is a WFP, then let $n(\pi)$ denote the number of (s in π.

The proof of the following lemma is left as an exercise.

Lemma 6.11. *A sequence $\pi = s_1 s_2 ... s_{2n}$ in which each s_i is either (or) is a well–formed parenthesization if and only if:*
(i) *each of (and) occurs n times in π, and*
(ii) *for each $1 \leq k \leq 2n$, among the symbols $s_1 s_2 ... s_k$ there are at least as many (s as there are)s.*

We will now describe the three weight–preserving bijections depicted in Figure 6.6. These explain the equicardinality of sets of BRTs, PPTs, and SDLPs which we have observed above. The first two bijections are described directly. The third is described by a recursive algorithm – another aspect of recursive structure.

Example 6.12 (A bijection between SDLPs and WFPs)**.** Given a SDLP $P = s_1 s_2 ... s_{2n}$, each s_i is either E or N. Replace each N in P by (, and replace each E in P by). Conversely, given a WFP $\pi = s_1 s_2 ... s_{2n}$, each s_i is either (or). Replace each (in π by N, and replace each) in π by E.

Lemma 6.11 implies that these give functions between the set of SDLPs and the set of WFPs. These functions are in fact mutually inverse bijections. Furthermore, it is obvious that if the SDLP P corresponds to the WFP π through these bijections, then $n(P) = n(\pi)$.

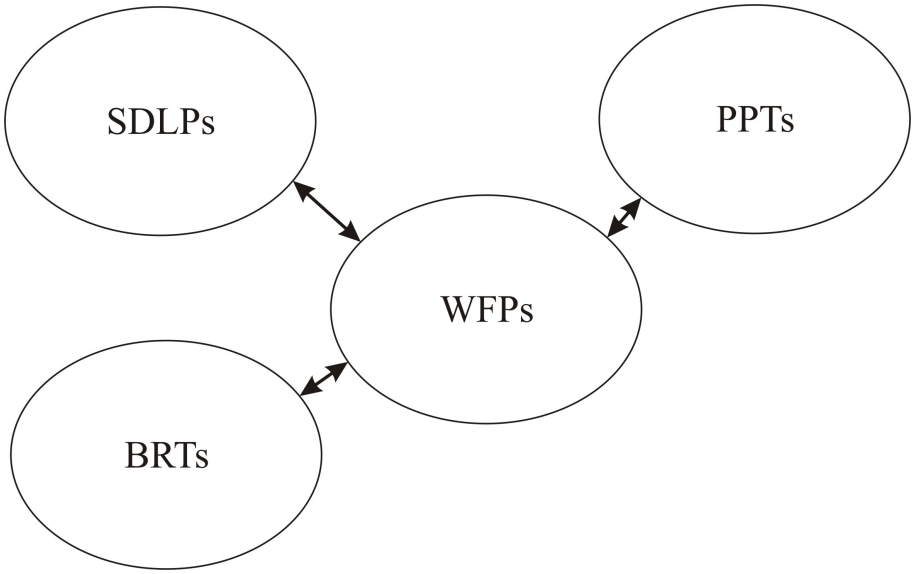


FIGURE 6.6. three bijections.

Example 6.13 (A bijection between PPTs and WFPs)**.** Let T be a PPT. Start at the root vertex $\odot$ and travel counterclockwise around the outside of the tree T. Each time you traverse an edge of T downwards, write down the symbol (. Each time you traverse an edge of T upwards, write down the symbol). Let π_T be the sequence of symbols constructed this way. Notice that we travel along each edge once in each direction, so the length of π is twice the number of edges of T. Moreover, for each edge we write down a (before we write down a). Lemma 6.11 then shows that π_T is a WFP. Notice that $n(\pi_T) = n(T) - 1$.

Conversely, let π be a WFP. Draw the root node of a tree, and call it the *current node*. Now start at the left end of π and read to the right end. Whenever you read a (, draw a child of the current node to the right of all children it may already have, and redefine the current node to be the new node. Whenever you read a), redefine the current node to be the unique parent of the present current node in the tree constructed so far, if this parent exists. If the parent does not exist then stop and call the result T_π. Notice that $n(T_\pi) = n(\pi) + 1$.

Since π is a WFP, Lemma 6.11 implies that when a) is read the current node always has an ancestor, except when the last symbol of π – which is necessarily a) – is read. Since the children of each node in T_π are ordered left–to–right, it follows that T_π is a PPT. The number of edges of T_π is easily seen to be $n(\pi)$.

These constructions $T \mapsto \pi_T$ and $\pi \mapsto T_\pi$ give mutually inverse bijections between the sets of PPTs and WFPs.

Example 6.14 (A bijection between BRTs and WFPs.)**.** Given a (BRT or $\varnothing$) T we associate with it a WFP π_T as follows. The definition of π_T is recursive. If $T = \varnothing$ then $\pi_\varnothing := \varepsilon$. If $T \neq \varnothing$ then let L and R be the left and right subtrees of T (note that L and R may be empty). In this case we let $\pi_T := (\pi_L)\pi_R$. It can be shown by induction on the number of nodes of T that π_T is a WFP.

Conversely, given a WFP π we associate with it a (BRT or $\varnothing$) called T_π as follows. The definition of T_π is recursive. If $\pi = \varepsilon$ then $T_\varepsilon := \varnothing$. If $n(\pi) \geq 1$ then the first symbol of π is a (. Since π is a WFP there is a unique symbol) of π with which it closes. Thus π has the form $\pi = (\alpha)\beta$, and both α and β are WFPs. We define T_π to be the BRT which has a root vertex connected to left subtree T_α and right subtree T_β. (Note that these subtrees may be empty.) By induction on $n(\pi)$ it can be seen that T_π is a BRT.

These constructions give mutually inverse bijections between the set of WFPs and the set of BRTs together with $\varnothing$. If π corresponds to T in this bijection then $n(\pi) = n(T)$.

There is only one more point which I would like to address in this section, and that is to dispell any misconception that recursive structure always leads to Catalan numbers. The examples above were chosen because they are solvable with the only technique currently at hand – at an essential stage, we used the Quadratic Formula in each case. Most interesting problems can not be solved by such a simple device. The following is an accessible example.

Example 6.15. A *ternary rooted tree* is a tree, with a designated *root node* $\odot$, which is drawn in the plane such that each node has at most three children, one to the left, one in the middle, and one to the right, if they exist. If a node has only one or two children then they are still labelled either left, middle, or right, with at most one of each. Let $n(T)$ denote the number of nodes of the ternary rooted tree T.

How many ternary rooted trees (TRTs) have n nodes, for each $n \in \mathbb{N}$? Let $\mathcal{W}$ be the set of TRTs. By reviewing the recursive structure for **binary** rooted trees, you should not be surprised to see that the recursive structure of the set of TRTs is described by

$$\mathcal{W} \rightleftharpoons \{\odot\} \times (\mathcal{W} \cup \{\varnothing\})^3 .$$

Moreover, if

$$T \leftrightarrow (\odot, L, M, R)$$

in this bijection (where L, M, R are the left, middle, and right subtrees of T, respectively), then

$$n(T) = 1 + n(L) + n(M) + n(R).$$

Defining the generating function

$$W(x) := \Phi_{\mathcal{W}}^{n}(x) = \sum_{T \in \mathcal{W}} x^{n(T)},$$

a calculation analogous to that in the case of BRTs yields the functional equation

$$W = x(1+W)^3$$

for the generating function $W(x)$. Obviously, the Quadratic Formula is useless for this problem! (There is a formula for solving cubic equations by radicals, but you can expect that calculation to get pretty horrible pretty fast.)

Okay, we are stuck for now. In Section 8, however, we will see absolutely **the most important theorem** in the subject of enumeration and recursive structure. This theorem will allow us to solve Example 6.15 with a one–line calculation! However, the proof of this result – the Lagrange Implicit Function Theorem – is a bit intricate, so in the next section we develop some of the theory of formal power series that will be needed to prove it.

---

6. Exercises.

---

1. What is the expected number of terminals among all PPTs with n nodes?

---

2. Among all PPTs with n nodes, what is the expected number of nodes which have exactly one child?

---

3. In a SDLP, a *peak* is a substring of the form NE. What is the expected number of peaks among all SDLPs with n nodes?

---

4. Let $F(x,y) := \sum_{T \in \mathcal{T}} x^{n(T)} y^{\ell(T)}$ be the generating function for BRTs with respect to both number of nodes and number of leaves. Let $T(x,y) := \sum_{T \in \mathcal{T}} x^{n(T)} y^{\tau(T)}$ be the generating function for BRTs with respect to both number of nodes and number of terminals. Explain why

$$F(x,y) = xy + 2xyT(x,y) + xT(x,y)^2$$

and use this to compute the average number of leaves among all BRTs with n nodes. [Hint: use the functional equation satisfied by $T(x,y)$.]

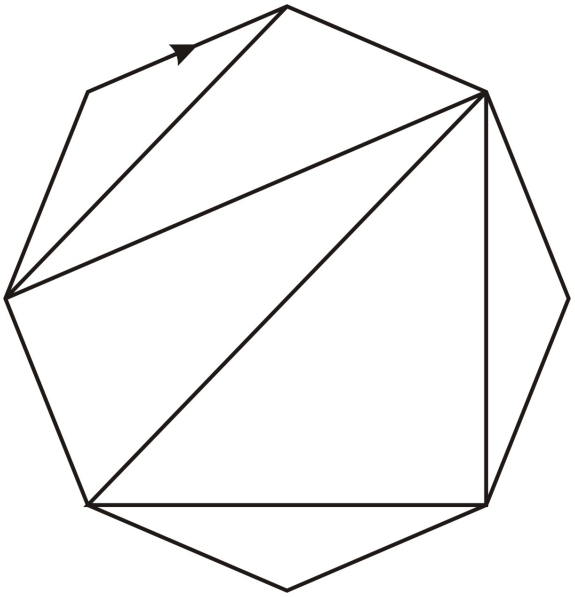


FIGURE 6.7. a triangulation of P_8.

---

5. What is the expected number of leaves among all PPTs with n nodes?

---

6. Let $\mathcal{A}$ denote the set of all plane planted trees (PPTs) such that every node has at most two children. Show that the number of PPTs in $\mathcal{A}$ with n nodes is

$$\sum_{j=0}^{\lfloor (n-1)/2 \rfloor} \frac{(n-1)!}{j!(j+1)!(n-1-2j)!}.$$

---

7. Let $\mathcal{H}$ be the set of all $PPTs$ in which no nodes have exactly one child.

(a) Explain why the generating function $H(x) := \Phi^n_{\mathcal{H}}(x)$ satisfies the equation

$$(x+1)H^2 - (x+1)H + x = 0.$$

(b) For each $n \in \mathbb{N}$, obtain a formula for the number of PPTs in $\mathcal{H}$ with n nodes.

---

8. Let P_n be regular polygon with n sides drawn in the plane, and with one of its sides marked with an arrow (to break the dihedral symmetry). A *triangulation* of P_n is a division of the interior of P_n into internally disjoint triangles, all the vertices of which are vertices of P_n. (See Figure 6.7 for an example.) Give a formula for τ_n, the number of triangulations of P_n, which is valid for all $n \geq 3$.

---

9. For $n \geq 1$, a 2–*by*–n *Standard Young tableau (SYT)* is a 2–by–n matrix in which the numbers $1, 2, \ldots, 2n$ each occur exactly once, the rows increase from left to right, and the columns increase from top to bottom. Show that for all $n \geq 1$, the number of 2–by–n SYTs is $\frac{1}{n+1}\binom{2n}{n}$.

10(a) Use Proposition 6.5(b) to show that for all $n \in \mathbb{N}$:

$$\sum_{k=0}^{n} \binom{2k}{k}\binom{2n-2k}{n-k} = 4^n.$$

10(b)* Prove the formula in part (a) combinatorially.

---

7. Formal Power Series.

In Sections 4 through 6 we have been manipulating infinite power series in one or more indeterminates without concerning ourselves that such manipulations are justified. So far we have not run into any problems, but perhaps that was just a matter of good luck. In this section we will see which algebraic manipulations are valid for formal power series (and more generally for formal Laurent series), as well as seeing some manipulations which are invalid. Special attention should be paid to the concept of convergence of a sequence of formal power series. Many students consistently confuse this with the concept of convergence of a sequence of real numbers (familiar from calculus), but the two concepts are in fact quite different.

First we recall some basic concepts and terminology of abstract algebra. (These are covered in MATH 135, but some review is warranted.) A *ring* is a set R which has two special elements, a *zero* $0 \in R$ and a *one* $1 \in R$, and is equipped with two binary operations, *multiplication* $\cdot : R \times R \to R$ and *addition* $+ : R \times R \to R$. A long list of axioms completes the definition, but suffice it here to say that the axioms state that the usual rules of integer arithmetic hold for $(R; \cdot, +; 0, 1)$ with one exception. In general, the multiplication in a ring is not required to be commutative: that is, the rule $ab = ba$ for all $a, b \in R$ is **not** in general required. When multiplication in R **is** commutative we say that R is a *commutative ring*. (The ring of 2–by–2 matrices with real entries is an example of a ring that is not commutative.) Some noncommutative rings are in fact useful in combinatorial enumeration, but all of the rings of importance in these notes are commutative.

The point of the previous paragraph is that when we say "R is a commutative ring" we mean that the familiar rules of arithmetic continue to hold for R.

Let $(R; \cdot, +; 0, 1)$ be a commutative ring. An element $a \in R$ is a *zero–divisor* if $a \neq 0$ and there is an element $b \in R$ with $b \neq 0$ such that $ab = 0$. For example, in the ring $\mathbb{Z}_{15}$ of integers modulo 15, we have $[3][5] = [15] = [0]$, so that $[3]$ and $[5]$ are zero–divisors in $\mathbb{Z}_{15}$. If R has no zero–divisors then R is called an *integral domain*. An element $a \in R$ is *invertible* if there is an element $b \in R$ such that $ab = 1$. Such an element is unique if it exists, for if $ac = 1$ as well, then

$$b = b1 = b(ac) = (ba)c = (ab)c = 1c = c.$$

Here we have used several of the axioms, including associativity and commutativity of multiplication. If $a \in R$ is invertible, then the unique element $b \in R$ such that $ab = 1$ is denoted by a^{-1}, and is called the *multiplicative inverse* of a. Notice that

$(a^{-1})^{-1} = a$. Finally, a commutative ring R is called a *field* if every $a \in R$ with $a \neq 0$ is invertible.

Proposition 7.1. *Let R be a commutative ring. If R is a field then R is an integral domain.*

Proof. Arguing for a contradiction, suppose not – thus, assume that R is a field but that $a \in R$ is a zero–divisor. Then $a \neq 0$ so that a^{-1} exists, and there is a $0 \neq b \in R$ such that $ab = 0$. Now we calculate that

$$b = b1 = b(aa^{-1}) = (ba)a^{-1} = (ab)a^{-1} = 0a^{-1} = 0,$$

which is the desired contradiction. □

There are several ways to construct a new ring starting from a ring which is already known. We will just give the four constructions which are useful for our purposes, and only for commutative rings. So, for the next little while, let R be a commutative ring.

Definition 7.2 (The Polynomial Ring)**.** The *ring of polynomials in x with coefficients in R* is denoted by $R[x]$, and is defined as follows. Here x is an *indeterminate*, meaning a symbol which is not in the set R, and is not a solution of any algebraic equation with coefficients in R. The elements of $R[x]$ are expressions of the form

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

for some $n \in \mathbb{N}$, in which $a_i \in R$ for all $0 \leq i \leq n$. Addition and multiplication are defined as you would expect, using the definitions of addition and multiplication in R for the coefficients, the distributive and associative laws, and the exponent rule for x – that is, $x^i \cdot x^j = x^{i+j}$. Since R is commutative, $R[x]$ is also commutative, but $R[x]$ is never a field. The invertible elements of $R[x]$ are just the constant polynomials a_0 with a_0 invertible in R. In particular, $x \in R[x]$ is not invertible. If R is an integral domain then so is $R[x]$ (this is Exercise 7.2(a)).

Definition 7.3 (The Ring of Rational Functions)**.** The *ring of rational functions in x with coefficients in R* is denoted by $R(x)$, and is defined as follows. There is some subtlety if R contains zero–divisors, so we will only consider the case in which R is an integral domain. The elements of $R(x)$ are of the form $f(x)/g(x)$ with $f(x), g(x) \in R[x]$ and $g(x) \neq 0$. Addition and multiplication are bootstrapped up from $R[x]$ by using the familiar means of manipulating fractions, and $R(x)$ is commutative because R is. Every nonzero element $f(x)/g(x)$ has a multiplicative inverse $g(x)/f(x)$, so that $R(x)$ is a field. In the expression $f(x)/g(x)$ we may take $g(x) = 1$, which shows that $R[x]$ is a subset of $R(x)$. The algebraic operations of these two rings agree on this subset $R[x]$, as is easily verified.

Definition 7.4 (The Ring of Formal Power Series)**.** The *ring of formal power series in x with coefficients in R* is denoted by $R[[x]]$, and is defined as follows. The elements of $R[[x]]$ are infinite expressions of the form

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n + \cdots$$

in which $a_n \in R$ for all $n \in \mathbb{N}$. Addition and multiplication are defined just as for the ring of polynomials $R[x]$, and $R[[x]]$ is commutative because R is. It is clear that $R[x]$ is a subset of $R[[x]]$, and that the algebraic operations of these two rings agree on this subset. The ring $R[[x]]$ is not a field because, for example, x is not invertible in $R[[x]]$. However, as the following proposition shows, there are quite a few invertible elements in $R[[x]]$.

Proposition 7.5. *Let R be a commutative ring, and let $f(x) = \sum_{i=0}^{\infty} a_ix^i$ be a formal power series in $R[[x]]$. Then $f(x)$ is invertible in $R[[x]]$ if and only if a_0 is invertible in R.*

Proof. We need to determine whether or not there exists a formal power series $g(x) = \sum_{j=0}^{\infty} b_jx^j$ in $R[[x]]$ such that $f(x)g(x) = 1$. Expanding the product, we have

$$\begin{aligned} f(x)g(x) &= \left(\sum_{i=0}^{\infty} a_ix^i\right)\left(\sum_{j=0}^{\infty} b_jx^j\right) \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_ib_jx^{i+j} = \sum_{k=0}^{\infty}\left(\sum_{i=0}^{k} a_ib_{k-i}\right)x^k. \end{aligned}$$

Comparing the coefficient of x^k on both sides of $f(x)g(x) = 1$, we see that $g(x)$ satisfies the equation if and only if $a_0b_0 = 1$ and $\sum_{i=0}^{k} a_ib_{k-i} = 0$ for all $k \geq 1$. If a_0 is not invertible in R then the equation $a_0b_0 = 1$ can not be solved for b_0, so that $g(x)$ does not exist and $f(x)$ is not invertible in $R[[x]]$. If a_0 is invertible in R then $b_0 := a_0^{-1}$ exists. Each of the remaining equations (for $k \geq 1$) can be rewritten as $a_0b_k = -\sum_{i=1}^{k} a_ib_{k-i}$, or, upon multiplying by b_0,

$$b_k = -b_0\sum_{i=1}^{k} a_ib_{k-i}.$$

These equations can be solved by induction on $k \geq 1$, yielding a solution for $g(x)$ which gives the multiplicative inverse of $f(x)$. Therefore, $f(x)$ is invertible in $R[[x]]$. □

Definition 7.6 (The Ring of Formal Laurent Series)**.** The *ring of formal Laurent series in x with coefficients in R* is denoted by $R((x))$, and is defined as follows. The elements of $R((x))$ are infinite expressions of the form

$$f(x) = a_rx^r + a_{r+1}x^{r+1} + a_{r+2}x^{r+2} + \cdots$$

in which $r \in \mathbb{Z}$ and $a_n \in R$ for all $n \geq r$. That is, a formal Laurent series is a generalization of a formal power series in which finitely many negative exponents are permitted. Addition and multiplication are defined just as for the ring $R[[x]]$ of formal power series, and $R((x))$ is commutative because R is. (I encourage you to check that when multiplying two formal Laurent series the coefficients of the product really are polynomial functions of the coefficients of the factors, and hence are in the ring R. This ensures that the multiplication in $R((x))$ is well–defined.) Note that the ring $R[[x]]$ is a subset of the ring $R((x))$, and that the algebraic operations of these rings agree on the subset $R[[x]]$. If $f(x) \in R((x))$ and $f(x) \neq 0$, then there is a smallest integer n such that $[x^n]f(x) \neq 0$; this is called the *index of* $f(x)$ and is denoted by $I(f)$. By convention, the index of 0 is $I(0) := +\infty$. Concerning the existence of multiplicative inverses in $R((x))$, we have the following proposition.

Proposition 7.7. *Let R be a commutative ring. If R is a field then $R((x))$ is a field.*

Proof. Consider a nonzero $f(x) = \sum_{n=I(f)}^{\infty} a_n x^n$ in $R((x))$. Then $a_{I(f)} \neq 0$ so that it is invertible in R, since R is a field. We may write $f(x) = x^{I(f)}g(x)$ with $g(x) = \sum_{n=0}^{\infty} a_{n+I(f)}x^n$, so that $g(x)$ is a formal power series in $R[[x]]$. The coefficient of x^0 in $g(x)$ is $a_{I(f)}$ and, by Proposition 7.5, it follows that $g(x)$ is invertible in $R[[x]]$, and hence in $R((x))$. Let $h(x) := x^{-I(f)}g^{-1}(x)$. Then

$$f(x)h(x) = x^{I(f)}g(x)x^{-I(f)}g^{-1}(x) = 1,$$

so that $h(x) = f^{-1}(x)$ and $f(x)$ is invertible in $R((x))$. Therefore, $R((x))$ is a field. □

The inclusions $R[x] \subset R[[x]] \subset R((x))$ and $R[x] \subset R(x)$ have been remarked upon already. In fact, if R is a field then $R(x) \subset R((x))$ as well. Also, the rings $R[[x]]$ and $R(x)$ have a nontrivial intersection, but neither one contains the other. Since we have no pressing need for these facts we will not pause to prove them, but instead relegate them to Exercise 7.2.

The constructions above may be combined and iterated, since the commutative ring R was quite general. For example, $R[x, y, z]$ denotes the ring of polynomials in three indeterminates x, y, and z. Similarly, $R[y][[x]]$ denotes the ring of formal power series in the indeterminate x with coefficients which are polynomials in the indeterminate y.

Example 7.8 (The Binomial Series)**.** In the polynomial ring $\mathbb{Q}[y]$, we define the polynomials $\binom{y}{n}$ for every $n \in \mathbb{N}$ by

$$\binom{y}{n} := \frac{y(y-1)\cdots(y-n+1)}{n!}.$$

The *binomial series* is then defined in $\mathbb{Q}[y][[x]]$ to be

$$(1+x)^y := \sum_{n=0}^{\infty} \binom{y}{n} x^n.$$

Notice that in the ring $\mathbb{Q}[y,z][[x]]$ we have the following identity:

$$\begin{aligned}
(1+x)^{y+z} &= \sum_{n=0}^{\infty} \binom{y+z}{n} x^n \\
&= \sum_{n=0}^{\infty} \sum_{j=0}^{n} \binom{y}{j} \binom{z}{n-j} x^n \\
&= \left(\sum_{j=0}^{\infty} \binom{y}{j} x^j \right) \left(\sum_{k=0}^{\infty} \binom{z}{k} x^k \right) \\
&= (1+x)^y \cdot (1+x)^z.
\end{aligned}$$

In this calculation we have used the Vandermonde Convolution Formula (Exercise 3.5). Notice that y and z, as well as x, are also indeterminates. Any complex number $\alpha \in \mathbb{C}$ may be substituted for y in $(1+x)^y$, and the resulting $(1+x)^\alpha$ is a formal power series in $\mathbb{C}[[x]]$.

The usual rules of arithmetic hold for all of the rings constructed above, but there are other operations on these rings that have no analogues in $\mathbb{Z}$. Care must be taken with these operations to ensure that they produce well–defined power series. In other words, these operations are not universally defined.

The first of the new operations are *formal differentiation* and *formal integration*. Since $R((x))$ contains all the other rings above (if R is a field) we will just define these operations on a typical formal Laurent series $f(x) = \sum_{n=I(f)}^{\infty} a_n x^n$. The formal derivative is always defined as

$$f'(x) := \frac{d}{dx} f(x) := \sum_{n=I(f)}^{\infty} n a_n x^{n-1}.$$

The formal integral is defined only if $\mathbb{Q} \subseteq R$ and $a_{-1} = 0$, in which case

$$\int f(x)dx := \sum_{n \geq I(f),\ n \neq -1} a_n \frac{x^{n+1}}{n+1}.$$

In particular, the formal integral is defined on all of $R[[x]]$ when $\mathbb{Q} \subseteq R$. One can show algebraically from the definitions that the familiar rules of calculus (the Product Rule, Quotient Rule, Chain Rule, Integration by Parts, and so on) continue to hold when all the integrals involved are defined.

Example 7.9 (The Logarithmic and Exponential Series)**.** (a) From calculus, we know that as functions of a real variable t,

$$\frac{d}{dt}\log(t) = \frac{1}{t}.$$

Now let $x := 1 - t$, so that $dx = -dt$. Then

$$\frac{d}{dx}\log\left(\frac{1}{1-x}\right) = \frac{d}{dt}\log(t) = \frac{1}{t} = \frac{1}{1-x}.$$

Assuming that $\log(1-x)^{-1} = \sum_{n=I}^{\infty} c_n x^n$ has a Laurent series expansion, we obtain the equation

$$\sum_{n=I}^{\infty} n c_n x^{n-1} = \sum_{k=0}^{\infty} x^k.$$

This implies that $c_n = 1/n$ for all $n \geq 1$, and that $c_n = 0$ for all $n \leq -1$, but gives no information about c_0. However, substituting $x = 0$ we see that $c_0 = \log(1) = 0$. In summary, we have the expansion

$$\log\left(\frac{1}{1-x}\right) = \sum_{n=1}^{\infty} \frac{x^n}{n}.$$

(b) As another example, the defining properties of the exponential function $\exp(x)$ are that $\exp(0) = 1$ and $\exp'(x) = \exp(x)$. Expanding this as a formal Laurent series $\exp(x) = \sum_{n=I}^{\infty} a_n x^n$ we get the equation

$$\frac{d}{dx}\exp(x) = \sum_{n=I}^{\infty} n a_n x^{n-1} = \sum_{m=I}^{\infty} a_m x^m = \exp(x).$$

Comparing the coefficients of x^{n-1} we see that $n a_n = a_{n-1}$ for all $n \geq I$. By induction, this implies that $a_n = 0$ for all $I \leq n < 0$, so that in fact $I \geq 0$. From $\exp(0) = 1$ we see that $a_0 = 1$, and it then follows that $a_n = 1/n!$ for all $n \in \mathbb{N}$, so that

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

The Chain Rule for differentiation involves "composition" of formal Laurent series, which has not yet been defined (and is not universally defined). In order to discuss the operation of composition we must introduce the concept of convergence of a sequence of formal Laurent series. This must not be confused with the concept of convergence of the single series $f(x) \in \mathbb{R}((x))$ if x is assigned a particular real value! In our applications the case of formal power series is most important, and we will restrict attention to this case for some points.

Definition 7.10 (Convergence of a Sequence of Formal Laurent Series)**.** Let $f_1(x)$, $f_2(x)$, $f_3(x)$,... be a sequence of formal Laurent series in $R((x))$. This sequence is *convergent* in $R((x))$ provided that the following two conditions hold:
• there is an integer J such that $J \leq I(f_k)$ for all $k \in \mathbb{N}$, and
• for every $n \in \mathbb{N}$ there exists a $K_n \in \mathbb{N}$ and $A_n \in R$ such that if $k \geq K_n$ then $[x^n]f_k(x) = A_n$.

The first condition says that there is a uniform lower bound for the indices of all of the entries $f_k(x)$ of the sequence. The second condition says that if we focus attention on only the n–th power of x, and consider the sequence $[x^n]f_k(x)$ of coefficients of x^n in $f_k(x)$ as $k \to \infty$, then this sequence (of elements of R) is *eventally constant*, with the ultimate value A_n.

If $(f_1(x), f_2(x), f_3(x), \ldots)$ is a convergent sequence of formal Laurent series then, with the above notations,

$$f(x) := \sum_{n=J}^{\infty} A_n x^n$$

is a well–defined formal Laurent series, called the *limit* of the convergent sequence. We use the notation

$$\lim_{k\to\infty} f_k(x) = f(x)$$

to denote this relationship.

Example 7.11. Here are a few simple examples for which you can check the definition easily.

(a) $\lim_{k\to\infty} x^k = 0.$

(b) $\lim_{k\to\infty} x^{-k}$ does not exist.

(c) $\lim_{k\to\infty} (1 + x + \cdots + x^k) = \dfrac{1}{1-x}.$

(d) $\lim_{k\to\infty} \dfrac{1}{1 - x/k}$ does not exist.

This last example is particularly instructive. You might think that the limit should exist and equal 1. However, the coefficient of x^1 in $(1 - x/k)^{-1}$ is $1/k$, and we see that the sequence $(1/k : \ k \geq 1)$ is not eventually constant. Therefore, this sequence of formal power series does not converge according to Definition 7.10.

Two important special cases of this concept of limits are the interpretations of infinite sums and infinite products of formal Laurent series. For an infinite sequence $(f_k(x) : \ k \in \mathbb{N})$ of formal Laurent series, we define

$$\sum_{k=0}^{\infty} f_k(x) := \lim_{K\to\infty} \sum_{k=0}^{K} f_k(x)$$

and

$$\prod_{k=0}^{\infty} f_k(x) := \lim_{K\to\infty} \prod_{k=0}^{K} f_k(x),$$

provided that these limits exist.

Example 7.12. Consider the infinite summation

$$\sum_{k=1}^{\infty} \frac{x^k}{1-x^k}.$$

Does this converge? To check the definition, for each $K \in \mathbb{N}$ consider the partial sum

$$f_K(x) := \sum_{k=1}^{K} \frac{x^k}{1-x^k} = \sum_{k=1}^{K} \left(x^k + x^{2k} + x^{3k} + \cdots\right).$$

To show that $\lim_{K\to\infty} f_K(x)$ exists, we fix an $n \in \mathbb{N}$ and consider the sequence $[x^n]f_K(x)$ as $K \to \infty$. Is this sequence eventually constant? Yes it is! If $K > n$ then

$$[x^n]f_K(x) = [x^n]\sum_{k=1}^{K} \frac{x^k}{1-x^k} = [x^n]\sum_{k=1}^{n} \frac{x^k}{1-x^k} = [x^n]f_n(x)$$

since the terms with $k > n$ can only contribute to the coefficients of powers of x which are strictly greater than n. Since this number $[x^n]f_n(x)$ depends only on n and not on K, the sequence $[x^n]f_K(x)$ is eventually constant as $K \to \infty$. Therefore, the infinite summation converges. In fact,

$$\sum_{k=1}^{\infty} \frac{x^k}{1-x^k} = \sum_{n=1}^{\infty} d(n)x^n$$

in which $d(n)$ is the number of positive integers that divide n.

Example 7.13. Consider the infinite product

$$\prod_{k=1}^{\infty} \left(1 + \frac{x}{k^2}\right).$$

Does this converge? To check the definition, for each $K \in \mathbb{N}$ consider the partial product

$$f_K(x) := \prod_{k=1}^{K} \left(1 + \frac{x}{k^2}\right).$$

To show that $\lim_{K\to\infty} f_K(x)$ exists, fix an $n \in \mathbb{N}$ and consider the sequence $[x^n]f_K(x)$ as $K \to \infty$. Is this sequence eventually constant? No, not for all $n \in \mathbb{N}$ – in

particular not for $n = 1$. To be precise, the sequence

$$[x^1]f_K(x) = \sum_{k=1}^{K} \frac{1}{k^2}$$

is not eventually constant as $K \to \infty$. Therefore, this infinite product is not convergent in the ring of fomal power series $R[[x]]$.

Examples 7.11(d) and 7.13 might seem strange to you, since the sequence of coefficients of x^1 converges *as a sequence of real numbers*, as does every other sequence of coefficients in these examples. That is, you would like to make use of the concept of convergence in the coefficient ring R (in this case, the field of real numbers). A general coefficient ring R has no such "topological" structure, which is why we require that the coefficient sequences be eventually constant. (In Exercise 12.12 we have occasion to discuss and employ a more flexible definition of convergence for a sequence of formal power series over a "normed" ring.)

The proof of the following proposition is left as an important exercise.

Proposition 7.14. *Let $(f_k(x) : k \in \mathbb{N})$ be an infinite sequence of formal power series in $R[[x]]$. The following conditions (a) and (b) are equivalent:*
(a) *The infinite sum $\sum_{k=0}^{\infty} f_k(x)$ converges;*
(b) *For every $J \in \mathbb{N}$, there are only finitely many $k \in \mathbb{N}$ such that $I(f_k) \leq J$.*
Moreover, these equivalent conditions imply the following:
(c) *The infinite product $\prod_{k=0}^{\infty}(1 + f_k(x))$ converges.*

At last, we turn to a discussion of the operation of composition of formal Laurent series. We restrict ourselves to the case in which R is a field, so that $R((x))$ is also a field. Given $f(x) = \sum_{n=I(f)}^{\infty} a_n x^n$ and $g(x) = \sum_{j=I(g)}^{\infty} b_j x^j$ in $R((x))$, we will determine the conditions under which $f(g(x))$ is a well–defined Laurent series. Of course, the symbol $f(g(x))$ is to be interpreted as

$$f(g(x)) := \sum_{n=I(f)}^{\infty} a_n g(x)^n := \lim_{K\to\infty} \sum_{n=I(f)}^{K} a_n g(x)^n.$$

There are a few cases. If $g(x) = 0$ then $g(x)^n$ does not exist for any negative integer n. Thus, if $g(x) = 0$ then $f(g(x)) = f(0)$ exists if and only if $I(f) \geq 0$, in which case $f(0) = [x^0]f(x)$. Assume now that $g(x) \neq 0$, so that $g^{-1}(x)$ does exist in the field $R((x))$. If $f(x)$ has only finitely many nonzero coefficients then $f(g(x))$ is a polynomial function of $g(x)$ and $g^{-1}(x)$, and therefore is an element of $R((x))$. Finally, assume that $f(x)$ has infinitely many nonzero coefficients. Therefore, for every $N \in \mathbb{N}$ there is an $n \geq N$ such that $a_n \neq 0$. Notice that (since R is an integral domain) the index of $a_n g(x)^n$ is $nI(g)$, n times the index of $g(x)$. Therefore, if $I(g) \leq 0$ then condition (b) of Proposition 7.14 is violated for the sequence

$(a_n g(x)^n : \ n \geq I(f))$, so that the infinite summation

$$f(g(x)) = \sum_{n=I(f)}^{\infty} a_n g(x)^n$$

is not defined. Conversely, if $I(g) > 0$ then condition (b) of Proposition 7.14 is satisfied, and $f(g(x))$ does exist. In summary, we have proved the following.

Proposition 7.15. *Let $f(x)$ and $g(x)$ be formal Laurent series in $R((x))$, in which R is a field. Then $f(g(x))$ is well–defined if and only if one of the following cases holds:*
(i) *$g(x) = 0$ and $I(f) \geq 0$;*
(ii) *$g(x) \neq 0$ and $f(x)$ has only finitely many nonzero coefficients;*
(iii) *$g(x) \neq 0$ and $I(g) > 0$.*

Under some circumstances, it is useful to think of a formal power series $f(x) \in R[[x]]$ as a "change of variables" by considering $u := f(x)$ as another variable itself. Of course, this u does depend on x, but one can consider the ring $R[[u]]$, which will be a subring of $R[[x]]$. The most interesting case is that in which $R[[u]] = R[[x]]$, which occurs exactly when there is a formal power series $g(u) \in R[[u]]$ such that $x = g(u)$. These two equations $u = f(x)$ and $x = g(u)$ imply that $x = g(f(x))$ and $u = f(g(u))$, which is called an *invertible change of variables*.

Proposition 7.16. *Let R be a field, and let $f(x)$ be a formal power series in $R[[x]]$. There exists $g(u) \in R[[u]]$ such that $x = g(f(x))$ and $u = f(g(u))$ if and only if either $I(f) = 1$, or $f(x) = a_0 + a_1 x$ with both a_0 and a_1 nonzero. If such a $g(u)$ exists then it is unique.*

Proof. Consider a pair of formal power series such that $x = g(f(x))$ and $u = f(g(u))$.

First, if $f(x) = 0$ then there is clearly no $g(u) \in R[[u]]$ such that $u = f(g(u))$, so that this case does not arise.

Second, consider the case in which $f(x) = \sum_{n=0}^{\infty} a_n x^n$ has a nonzero constant term $a_0 \neq 0$. In order for the composition $g(f(x))$ to be well-defined, $g(u) = \sum_{j=0}^{\infty} b_j x^j$ must have only finitely many nonzero coefficients, by Proposition 7.15. Since $f(g(x)) = x$ and $a_0 \neq 0$, it follows that $b_0 \neq 0$ as well (for otherwise, if $b_0 = 0$ then $[x^0]f(g(x)) = a_0 \neq 0$, contradicting $f(g(x)) = x$). Now, since $b_0 \neq 0$, for the composition $f(g(u))$ to be well-defined, $f(x)$ must have only finitely many nonzero coefficients, by Proposition 7.15. That is, $f(x)$ and $g(u)$ are polynomials with nonzero constant terms. Now both $f(g(u))$ and $g(f(x))$ are polynomials of degree $\deg(f) \cdot \deg(g)$, and hence $\deg(f) = \deg(g) = 1$. From $f(x) = a_0 + a_1 x$ and $g(u) = b_0 + b_1 u$ we see that $x = g(f(x)) = b_0 + b_1 a_0 + b_1 a_1 x$, so that $b_0 + b_1 a_0 = 0$ and $b_1 a_1 = 1$. That is, $b_1 = a_1^{-1}$ and $b_0 = -a_0 a_1^{-1}$, so that $g(u)$ is uniquely determined. One easily checks that $f(g(u)) = u$ as well, finishing this case.

Finally, consider a nonzero $f(x) \in R[[x]]$ with index at least one. Let $f(x) = \sum_{n=1}^{\infty} a_n x^n$ be given – we are assuming that $a_0 = 0$. To begin with, we seek a formal power series $g(u) = \sum_{j=0}^{\infty} b_j u^j$ such that $g(f(x)) = x$. Expanding this, we have

$$\begin{aligned} x &= g(f(x)) = \sum_{j=0}^{\infty} b_j f(x)^j = \sum_{j=0}^{\infty} b_j \left(\sum_{n=1}^{\infty} a_n x^n \right)^j \\ &= \sum_{m=0}^{\infty} x^m \sum_{j=0}^{\infty} b_j \sum_{n_1+n_2+\cdots+n_j=m} a_{n_1} a_{n_2} \cdots a_{n_j}. \end{aligned}$$

In the inner summation, since $a_0 = 0$ we need only consider those j–tuples $(n_1, \ldots, n_j)$ such that each $n_i \geq 1$. Comparing coefficients of like powers of x, we see that for x^0 we have

$$0 = b_0,$$

for x^1 we have

$$1 = b_1 a_1,$$

and for x^m with $m \geq 2$ we have

$$0 = \sum_{j=1}^{m} b_j \sum_{n_1+n_2+\cdots+n_j=m} a_{n_1} a_{n_2} \cdots a_{n_j}.$$

(The outer summation may be terminated at $j = m$ because each $n_i \geq 1$ and $n_1 + n_2 + \cdots + n_j = m$, and when $j = 0$ the inner summation is empty.)

Now let's solve these equations for the coefficients b_j of $g(u)$. That $b_0 = 0$ is immediate, and b_1 exists if and only if $a_1 \neq 0$, since R is a field. This shows that if $g(u)$ exists then $I(f) = 1$. Conversely, assume that $I(f) = 1$, so that $a_1 \neq 0$ – then $b_1 := a_1^{-1}$ exists. For all $m \geq 2$ we have

$$b_m = -b_1^m \sum_{j=1}^{m-1} b_j \sum_{n_1+n_2+\cdots+n_j=m} a_{n_1} a_{n_2} \cdots a_{n_j}.$$

The RHS is a polynomial function of $\{a_1, \ldots, a_m, b_1, \ldots, b_{m-1}\}$, so that these equations can be solved uniquely by induction on $m \geq 2$ to determine the coefficients of $g(u)$ in the case $I(f) = 1$. This establishes existence and uniqueness of a power series $g(u)$ such that $g(f(x)) = x$ when $I(f) = 1$.

Now, in the case $I(f) = 1$ the power series $g(u)$ we have constructed also has index one: $I(g) = 1$. We want to show that $f(g(u)) = u$ as well. By the preceeding argument, there is a unique formal power series $h(x)$ such that $h(g(u)) = u$. Let's substitute $u = f(x)$ into this – we obtain $h(g(f(x))) = f(x)$. Since $g(f(x)) = x$ this reduces to $h(x) = f(x)$, so that $f(g(u)) = u$, as desired. This completes the proof. $\square$

The unique formal power series $g(u)$ guaranteed by Proposition 7.16 is referred to as the *compositional inverse* of $f(x)$, and is sometimes denoted by $f^{<-1>}(u)$.

Proposition 7.16 is only part of the truth, as the following example shows.

Example 7.17. Consider the rational function $f(x) = (1+x)/(1-x)$. As a formal power series we have

$$\frac{1+x}{1-x} = 1 + 2x + 2x^2 + 2x^3 + 2x^4 + \cdots$$

Algebraically, we can solve $u = f(x)$ for x as follows: $u(1-x) = 1+x$, so that $u - 1 = x(1+u)$, so that $x = (-1+u)/(1+u)$. Hence, $x = g(u)$, where

$$g(u) = \frac{-1+u}{1+u} = -1 + 2u - 2u^2 + 2u^3 - 2u^4 + \cdots$$

Now, even though $u = f(x)$ and $x = g(u)$, neither of the compositions $f(g(u))$ nor $g(f(x))$ are well–defined, by Proposition 7.15.

The trouble with Example 7.17 is only that the compositions of these formal power series do not converge according to Definition 7.10. To circumvent this problem, we can define composition of rational functions instead.

Definition 7.18. Let $f(x) = p(x)/q(x)$ and $g(u)$ be rational functions. The *composition of $g(u)$ into $f(x)$* is defined to be the rational function

$$f(g(u)) := \frac{p(g(u))}{q(g(u))}.$$

Notice that since $p(x)$ and $q(x)$ are polynomials, the expressions $p(g(u))$ and $q(g(u))$ are well–defined rational functions.

Using this definition, all the operations in Example 7.17 are well-defined, resolving the difficulty.

---

7. Exercises.

---

1. Let R be a commutative ring. For $a \in R$ consider the function $\mu_a : R \to R$ defined by $\mu_a(r) := ar$ for all $r \in R$.

(a) Show that if R is an integral domain and $a \neq 0$, then $\mu_a : R \to R$ is an injection.

(b) Show that if R is a finite integral domain then R is a field. (The ring $\mathbb{Z}$ of integers is an integral domain which is not a field. Thus, finiteness of R is essential for this problem. See Exercise 1.4.)

---

2. Let R be a commutative ring.

(a) Show that if R is an integral domain, then $R[x]$ is an integral domain.

(b) Show that neither of $R[[x]]$ nor $R(x)$ contains the other.

(c) Show that if R is a field then $R(x)$ is a proper subset of $R((x))$.

(d) Find an element of $\mathbb{Z}(x)$ which is not in $\mathbb{Z}((x))$.

(e) Show that $R[[x]][y]$ is a proper subset of $R[y][[x]]$.

---

3. Recall the notation of Example 7.8.

(a) Show that in $R[y][[x]]$, $(1-x)^{-y} = \sum_{n=0}^{\infty} \binom{y+n-1}{n} x^n$.

(b) Show that in $R[y,z][[x]]$,

$$\frac{1}{(1-x)^{y+z}} = \frac{1}{(1-x)^y} \cdot \frac{1}{(1-x)^z}.$$

---

4. Let $f(x)$ and $g(x)$ be in $R((x))$. Show that

$$\frac{d}{dx}\left(f(x)g(x)\right) = f'(x)g(x) + f(x)g'(x).$$

---

5. Prove Proposition 7.14.

---

6. Define a sequence of formal power series in $\mathbb{Z}[[x]]$ by $f_0(x) := 1$, $f_1(x) := 1$, and $f_{k+1}(x) := f_k(x) + x^k f_{k-1}(x)$ for all $k \geq 1$. Prove that the limit $\lim_{k\to\infty} f_k(x)$ exists.

---

7. Define a sequence of formal power series in $\mathbb{Z}[[x]]$ by $g_0(x) := 1$, and $g_{k+1}(x) := (1 - x g_k(x))^{-1}$ for all $k \in \mathbb{N}$.

(a)* Prove that the limit $g(x) = \lim_{k\to\infty} g_k(x)$ exists.

(b) Show that $g(x)$ satisfies the equation $g(x) = (1 - xg(x))^{-1}$, and thus is the generating function for SDLPs in Theorem 6.9.

---

8. Consider the Chain Rule: for $f(x)$ and $g(x)$ in $R((x))$ such that $f(g(x))$ is defined,

$$\frac{d}{dx} f(g(x)) = f'(g(x))g'(x).$$

(a) Prove this for $f(x)$ and $g(x)$ in $R[[x]]$.

(b) Prove this for $f(x)$ and $g(x)$ in $R((x))$.

---

9. Does the following limit exist in $R[[x]]$? Explain.

$$\lim_{k\to\infty}\left(1+\frac{x}{k}\right)^k.$$

---

10. Does the following limit exist in $R((x))$? Explain.

$$\lim_{k\to\infty}\frac{x^{-k}}{1-x^k}$$

---

11. Show that, as a sequence of formal power series in $\mathbb{Z}[[q]]$,

$$\lim_{a\to\infty}\left[\begin{array}{c} a+b \\ b \end{array}\right]_q = \frac{1}{(1-q)(1-q^2)\cdots(1-q^b)}.$$

12(a) Show that if $f(x)\in R[[x]]$ is such that $[x^0]f(x)=1$, then $\log(f(x))$ converges.

12(b) With $f(x)$ as in part (a), show that

$$\frac{d}{dx}\log(f(x)) = f^{-1}(x)\frac{d}{dx}f(x).$$

12(c) Show that $\log(\exp(x))=x$.

12(d) Show that

$$\exp\left(\log\left(\frac{1}{1-x}\right)\right)=\frac{1}{1-x}.$$

---

7. Endnotes.

What we have in this section is just a tiny glimpse into the subject of commutative algebra. Most of the motivation for this subject is quite separate from the concerns of enumerative combinatorics – it aims towards algebraic geometry, which describes sets of solutions to a collection of polynomial equations, among other things. If you want to read up on this I recommend the following books, with the caveat that they are really intended for graduate students.

• M.F. Atiyah and I.G. Macdonald, "Introduction to Commutative Algebra," Addison–Wesley, Reading MA, 1969.

• D. Eisenbud, "Commutative Algebra – with a view toward Algebraic Geometry," *Graduate Texts in Mathematics,* **150**, Springer–Verlag, New York, 1995.

• J.M. Ruiz, "The Basic Theory of Power Series," Braunschweig Vieweg, 1993.

The last of these also develops the foundations of geometry for categories other than the algebraic one. It is all very fascinating stuff, but quite complicated, and not really relevant to our present purpose. More down–to–earth expositions of what we need of formal power series are given in Chapter 2 of

• H.S. Wilf, "Generatingfunctionology," Academic Press, New York, 1994

and in Chapter 3 of

• C.D. Godsil, "Algebraic Combinatorics," Chapman and Hall, New York, 1993.

8. The Lagrange Implicit Function Theorem.

The topic of this section – the Lagrange Implicit Function Theorem – is **the most widely applicable technique** for the enumeration of recursively defined structures. As we saw in Section 6, recursive structure leads to a functional equation for the relevant generating function. When this equation is quadratic it can be solved by the Quadratic Formula, as in Section 6. In most cases, however, the equation is **not** quadratic and simple high–school tricks do not suffice. Instead, Lagrange's Theorem is perfectly suited to such tasks.

Theorem 8.1 (LIFT)**.** *Let $\mathbb{K}$ be a commutative ring which contains the rational numbers $\mathbb{Q}$. Let $F(u)$ and $G(u)$ be formal power series in $\mathbb{K}[[u]]$ such that $[u^0]G(u)$ is invertible in $\mathbb{K}$.*
(a) *There is a unique (nonzero) formal power series $R(x)$ in $\mathbb{K}[[x]]$ such that*

$$R(x) = x\,G(R(x)).$$

(b) *The constant term of $R(x)$ is 0, and for all $n \geq 1$,*

$$[x^n]F(R(x)) = \frac{1}{n}[u^{n-1}]F'(u)G(u)^n.$$

Before proceeding to the proof, let's apply this theorem to Example 6.15.

Example 8.2. As in Example 6.15, let $\mathcal{W}$ be the set of ternary rooted trees, and let $W(x)$ be the generating function for $\mathcal{W}$ with respect to number of nodes. We have derived the functional equation

$$W = x(1+W)^3$$

for this generating function. This fits perfectly into the hypothesis of LIFT, using $\mathbb{K} = \mathbb{Q}$, $F(u) = u$ and $G(u) = (1+u)^3$. Thus, we calculate that the number of ternary rooted trees with $n \geq 1$ nodes is

$$[x^n]W(x) = \frac{1}{n}[u^{n-1}](1+u)^{3n} = \frac{1}{n}\binom{3n}{n-1}.$$

That's a piece of cake!

In order to prove Lagrange's Theorem we need to develop a few more facts about formal Laurent series. These have to do with the *formal residue operator*, which is merely the operator $[x^{-1}]$ that extracts the coefficient of x^{-1} from a formal Laurent series. (The terminology is by analogy with the case $\mathbb{K} = \mathbb{C}$ of complex numbers and the Cauchy Residue Theorem.) We require three facts about it, and then we can prove LIFT.

Lemma 8.3. *Let $F(x)$ be a formal Laurent series. Then*

$$[x^{-1}]\frac{d}{dx}F(x) = 0.$$

Proof. If $F(x) = \sum_{n=I(F)}^{\infty} a_n x^n$ then

$$[x^{-1}]\frac{d}{dx}F(x) = [x^{-1}]\sum_{n=I(F)}^{\infty} n a_n x^{n-1} = 0a_0 = 0.$$

□

Lemma 8.4. *Let $F(x)$ and $G(x)$ be formal Laurent series. Then*

$$[x^{-1}]F'(x)G(x) = -[x^{-1}]F(x)G'(x).$$

Proof. This follows by applying Lemma 8.3 to $F(x)G(x)$, since

$$\frac{d}{dx}(F(x)G(x)) = F'(x)G(x) + F(x)G'(x),$$

by Exercise 7.3. □

Lemma 8.5 (Change of Variables)**.** *Let $F(u)$ and $B(x)$ be formal Laurent series. Assume that $I(B) > 0$, and that $[x^{I(B)}]B(x)$ is invertible in $\mathbb{K}$. Then*

$$[x^{-1}]F(B(x))B'(x) = I(B)[u^{-1}]F(u).$$

Proof. First, we consider the case in which $F(u) = u^k$ for some integer k. If $k \neq -1$ then we may write the LHS of the formula as

$$[x^{-1}]B(x)^k B'(x) = \frac{1}{k+1}[x^{-1}]\frac{d}{dx}B(x)^{k+1} = 0,$$

by Lemma 8.3. Also, if $k \neq -1$ then $[u^{-1}]u^k = 0$ on the RHS, so the formula holds in this case. In the remaining case, $k = -1$, we have $I(B)[u^{-1}]u^{-1} = I(B)$ on the RHS. On the LHS we have $[x^{-1}]B(x)^{-1}B'(x)$. To compute this, write

$$B(x) = cx^{I(B)}H(x),$$

in which $c \in \mathbb{K}$ is invertible and $H(x)$ is a formal power series with $[x^0]H(x) = 1$. By Proposition 7.5, $H(x)^{-1}$ exists and is a formal power series – also

$$B(x)^{-1} = c^{-1}x^{-I(B)}H(x)^{-1}$$

and

$$B'(x) = cx^{I(B)}H'(x) + cI(B)x^{I(B)-1}H(x).$$

Therefore,

$$\begin{aligned}[x^{-1}]B(x)^{-1}B'(x) &= [x^{-1}](c^{-1}x^{-I(B)}H(x)^{-1})(cx^{I(B)}H'(x) + cI(B)x^{I(B)-1}H(x)) \\ &= [x^{-1}](H(x)^{-1}H'(x) + I(B)x^{-1}) = I(B),\end{aligned}$$

since $H(x)^{-1}H'(x)$ is a formal power series. This establishes the formula whenever $F(u) = u^k$ for some integer $k \in \mathbb{Z}$.

Now consider any formal Laurent series $F(u) = \sum_{k=I(F)}^{\infty} a_k u^k$. We have

$$F(B(x))B'(x) = \sum_{k=I(F)}^{\infty} a_k B(x)^k B'(x),$$

and so, by using the cases we have already proven, we see that

$$\begin{aligned}
[x^{-1}]F(B(x))B'(x) &= \sum_{k=I(F)}^{\infty} a_k [x^{-1}]B(x)^k B'(x) \\
&= \sum_{k=I(F)}^{\infty} a_k I(B)[u^{-1}]u^k \\
&= I(B)[u^{-1}] \sum_{k=I(F)}^{\infty} a_k u^k = I(B)[u^{-1}]F(u).
\end{aligned}$$

This proves the lemma. □

Now we can prove the Lagrange Implicit Function Theorem. I admit that this argument can be verified line–by–line, but that it does not convey an overall sense of understanding. For that deeper understanding, I present a second – combinatorial – proof of LIFT in Section 13. This second proof is usually postponed until C&O 430/630.

Proof of LIFT. For part (a), let $R(x) = \sum_{n=0}^{\infty} r_n x^n$, and let $G(u) = \sum_{k=0}^{\infty} g_k u^k$ with $g_0 \neq 0$ in $\mathbb{K}$. (We do not need g_0 to be invertible for this part of the proof.) Consider the equation $R(x) = xG(R(x))$. We will show that it has a unique solution $R(x)$ (which is nonzero) by showing that for each $n \in \mathbb{N}$, r_n is determined by the previous coefficients $r_0, r_1, \ldots, r_{n-1}$ and by the coefficients $g_0, g_1, \ldots, g_{n-1}$ of $G(x)$, and that a suitable value for r_n always exists. First of all, notice that

$$r_0 = [x^0]R(x) = [x^0]xG(R(x)) = 0,$$

so we may write $R(x) = \sum_{n=1}^{\infty} r_n x^n$ instead. Next, expand both sides of the equation $R(x) = xG(R(x))$ and equate like powers of x:

$$\begin{aligned}
\sum_{n=1}^{\infty} r_n x^n &= x \sum_{k=0}^{\infty} g_k \left(\sum_{n=1}^{\infty} r_n x^n \right)^k \\
&= \sum_{n=1}^{\infty} x^n \left(\sum_{k=0}^{\infty} g_k \sum_{n_1+n_2+\cdots+n_k=n-1} r_{n_1} r_{n_2} \cdots r_{n_k} \right).
\end{aligned}$$

The inner sum on the RHS is over all ordered k–tuples of positive integers which sum up to $n-1$. For a given value of n, this implies that $k \leq n-1$, and so for every $n \geq 1$,

$$r_n = \sum_{k=0}^{n-1} g_k \sum_{n_1+n_2+\cdots+n_k=n-1} r_{n_1} r_{n_2} \cdots r_{n_k}.$$

Notice that since $r_1 = g_0 \neq 0$, the index of R is $I(R) = 1$. The proof of part (a) is completed by showing that r_n is a polynomial function of $g_0, \ldots, g_{n-1}$. This is accomplished by an easy induction on $n \in \mathbb{N}$, which we leave to the reader.

For this proof of part (b) we require g_0 to be invertible in $\mathbb{K}$, so that $G(u)^{-1}$ exists in $\mathbb{K}[[u]]$. Consider the formal power series $P(u) := uG(u)^{-1}$. Let $R(x)$ be defined as in part (a), and make the change of variables $u := R(x)$. Then, since $R(x) = xG(R(x))$, we get

$$x = R(x)G(R(x))^{-1} = uG(u)^{-1} = P(u),$$

so that $x = P(u)$ is the change of variables inverse to $u = R(x)$. By Proposition 7.16 both of the compositions $x = P(R(x))$ and $u = R(P(u))$ are well–defined.

Now, for $n > 0$ we may calculate, using Lemma 8.4, that

$$\begin{aligned} [x^n]F(R(x)) &= [x^{-1}]x^{-1-n}F(R(x)) = -\frac{1}{n}[x^{-1}]\left(\frac{d}{dx}x^{-n}\right)F(R(x)) \\ &= \frac{1}{n}[x^{-1}]x^{-n}F'(R(x))R'(x) = \frac{1}{n}[x^{-1}]H(R(x))R'(x), \end{aligned}$$

in which we have put $H(u) := P(u)^{-n}F'(u)$, so that $H(R(x)) = x^{-n}F'(R(x))$. Continuing, by Lemma 8.5 we have

$$\begin{aligned} [x^{-1}]H(R(x))R'(x) &= I(R)[u^{-1}]H(u) = [u^{-1}]P(u)^{-n}F'(u) \\ &= [u^{-1}]u^{-n}G(u)^n F'(u) \\ &= [u^{n-1}]F'(u)G(u)^n. \end{aligned}$$

This completes the proof. □

I warned you... it makes sense line–by–line... but where the heck did all that algebra come from?! All I can say is that in Section 13 we find a way to interpret this formula combinatorially and prove it in a conceptual manner (and under slightly weaker hypotheses). In the meantime, here is an illustration of the method in practice.

Example 8.6. What is the expected number of terminals among all ternary rooted trees with n nodes? Let $\tau(T)$ denote the number of terminals of $T \in \mathcal{W}$, and consider the bivariate generating function

$$W(x,y) := \sum_{T \in \mathcal{W}} x^{n(T)} y^{\tau(T)}.$$

Analogously with Question 6.2, the average we seek is A_n/T_n in which

$$T_n = [x^n]W(x,1) = \frac{1}{n}\binom{3n}{n-1}$$

from Example 8.2, and

$$A_n = [x^n]\left.\frac{\partial}{\partial y}W(x,y)\right|_{y=1}.$$

To compute A_n we begin by deriving a functional equation for $W(x,y)$ from the recursive structure of ternary rooted trees:

$$\begin{aligned}
\mathcal{W} &\rightleftharpoons \{\odot\} \times (\{\varnothing\} \cup \mathcal{W})^3 \\
T &\leftrightarrow (\odot, L, M, R) \\
n(T) &= 1 + n(L) + n(M) + n(R) \\
\tau(T) &= \begin{cases} 1 & \text{if } L = M = R = \varnothing, \\ \tau(L) + \tau(M) + \tau(R) & \text{otherwise.} \end{cases}
\end{aligned}$$

This yields the functional equation

$$W = x(y + 3W + 3W^2 + W^3)$$

for the generating function $W(x,y)$. Now LIFT applies with $\mathbb{K} = \mathbb{Q}(y)$, $F(u) = u$, and $G(u) = y + 3u + 3u^2 + u^3$. The following calculation is a bit sneaky, so read carefully and think about why each step is valid.

$$\begin{aligned}
A_n &= [x^n]\left.\frac{\partial}{\partial y}W(x,y)\right|_{y=1} = \left.\frac{\partial}{\partial y}[x^n]W(x,y)\right|_{y=1} \\
&= \left.\frac{\partial}{\partial y}\frac{1}{n}[u^{n-1}](y + 3u + 3u^2 + u^3)^n\right|_{y=1} \\
&= \left.\frac{1}{n}[u^{n-1}]\,\frac{\partial}{\partial y}(y + 3u + 3u^2 + u^3)^n\right|_{y=1} \\
&= \left.\frac{1}{n}[u^{n-1}]\; n(y + 3u + 3u^2 + u^3)^{n-1}\right|_{y=1} \\
&= [u^{n-1}](1+u)^{3n-3} = \binom{3n-3}{n-1}.
\end{aligned}$$

Therefore, the average number of terminals among all ternary rooted trees with n nodes is

$$\frac{A_n}{T_n} = \frac{\binom{3n-3}{n-1}}{\frac{1}{n}\binom{3n}{n-1}} = \frac{(2n+1)(2n)(2n-1)}{3(3n-1)(3n-2)},$$

after some simplification. As $n \to \infty$ this is asyptotic to $8n/27$, so that in a large random ternary rooted tree one expects about $8/27 = 0.\overline{296}$ of the nodes to be terminals.

8. Exercises.

1. Fix a positive integer c. For each $n \in \mathbb{N}$, determine the number of plane planted trees with n nodes in which the number of children of each node is divisible by c. (I remind you that 0 is divisible by every such c.)

2. For a plane planted tree T, let $h(T)$ denote the number of nodes of T which have an even number of children. For each $n \in \mathbb{N}$, determine the average value of $h(T)$ among all the PPTs with n nodes.

3. Fix an integer $k \geq 2$. A *k–ary rooted tree* T has a root node $\odot$, and each node may have at most one child of each of k "types". (The case $k = 2$ gives BRTs, and the case $k = 3$ gives TRTs.)

(a) Show that the number of k–ary rooted trees with n nodes is $\frac{1}{n}\binom{kn}{n-1}$.

(b) Show that, as $n \to \infty$, the expected number of terminals among all k–ary rooted trees with n nodes is asymptotically $(1 - 1/k)^k n$.

4. For a plane planted tree (PPT) T, let $f(T)$ be the number of nodes of T with at least three children. Show that for $n \geq 4$, the average value of $f(T)$ among all the PPTs with n nodes is $(n^2 - 3n)/(8n - 12)$.

5. If an SDLP P touches the diagonal $x = y$ at points

$$(0, 0) = (k_0, k_0), (k_1, k_1), \ldots, (k_r, k_r) = (n, n),$$

then the sub–path of P between the points (k_{i-1}, k_{i-1}) and (k_i, k_i) is called the i–th *block* of P. Show that the expected number of blocks among all SDLPs to (n, n) is $3n/(n + 2)$.

6. For a plane planted tree (PPT) T, let $d(T)$ be the degree of the root nodes of T. Show that for $n \geq 1$, the average value of $d(T)$ among all the PPTs with n nodes is $(3n - 3)/(n + 1)$.

7. For a plane planted tree T, a *middle child* is a non–root node which is neither the leftmost nor the rightmost child of its parent. Let $p(T)$ denote the number of middle children in T. For each $n \in \mathbb{N}$, determine the average value of $p(T)$ among all the PPTs with n nodes.

8. (a) Let α and x be indeterminates. Find a formal power series $f(y)$ such that $\exp(\alpha x) = f(x\exp(-x))$.
(b) Let β be another indeterminate. Prove that

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

(For $n \geq 1$, the polynomial $z(n+z)^{n-1}$ is known as an *Abel polynomial*.)

8. Endnotes.

Incidentally, Lagrange was not a combinatorialist. In fact, combinatorics as such did not even exist during his lifetime (with the exception of a few things which Euler investigated). Lagrange discovered this theorem in order to solve functional equations which he derived in calculations of the orbital motion of the moon. These calculations were by far the most accurate ever done up to that time, and for many decades thereafter. All without electricity, too, of course!

For some entertaining reading, and somewhat fictionalized biographies of famous mathematicians, I recommend the following dated and sexistly titled classic:

- E.T. Bell, "Men of Mathematics," Simon & Schuster, New York, 1986.

The proof of the Lagrange Implicit Function Theorem presented here is adapted from Goulden and Jackson's monumental book:

- I.P. Goulden and D.M. Jackson, "Combinatorial Enumeration," John Wiley & Sons, New York, 1983.

9. Integer Partitions.

Integer partitions are fundamental combinatorial objects which occur naturally in a variety of subjects including number theory and abstract algebra. They also provide a convenient collection of examples to which we may apply the concepts we have developed so far.

Definition 9.1. An *integer partition* is a finite sequence

$$\lambda := (\lambda_1, \lambda_2, \ldots, \lambda_k)$$

of positive integers such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$. The entries λ_i of λ are called the *parts* of the partition. The set of all partitions is denoted by $\mathcal{Y}$ (after Alfred Young). For a partition $\lambda \in \mathcal{Y}$, we use

$$n(\lambda) := \lambda_1 + \lambda_2 + \cdots + \lambda_k$$

to denote the *size* of λ, and

$$k(\lambda) := k$$

to denote the *length* of λ. Notice that there is a unique partition ε with $n(\varepsilon) = 0$ and $k(\varepsilon) = 0$, namely the *empty* or *null* partition.

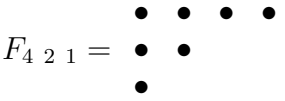
Integer partitions are traditionally written without the parentheses or commas. Thus, we write 6 4 3 3 1 instead of $(6, 4, 3, 3, 1)$.

Definition 9.2. Given an integer partition λ, the *Ferrers diagram* of λ is the set of ordered pairs

$$F_\lambda := \{(i, j) : \ 1 \leq i \leq k(\lambda) \text{ and } 1 \leq j \leq \lambda_i\}.$$

This set is visualized using matrix indexing, so for instance

$F_{4\ 2\ 1} =$



Often, we will draw pictures with little boxes instead of little dots, but the meaning is the same. The number of dots (or boxes) on the main diagonal of F_λ is denoted by $d(\lambda)$. This can also be defined by the formula

$$d(\lambda) := \#\{i : \ \lambda_i \geq i\}.$$

For each $n \in \mathbb{N}$, let $p(n)$ denote the number of partitions of size n. The entries of the following table are easily verified.

n	0	1	2	3	4	5	6	$\cdots$
$p(n)$	1	1	2	3	5	7	11	$\cdots$

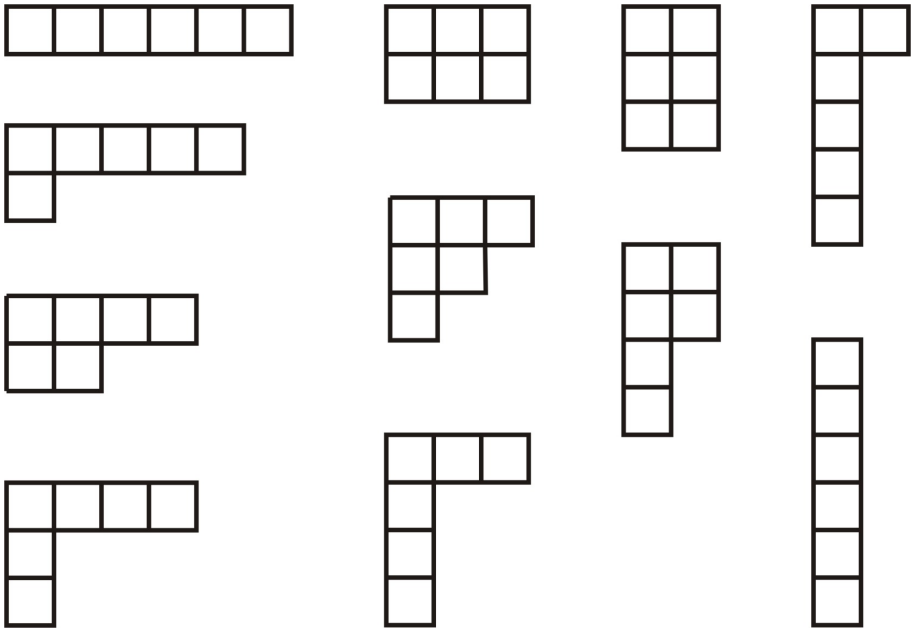


FIGURE 9.1. the partitions of size six.

The Ferrers diagrams of all partitions of size six are shown in Figure 9.1.

There is an alternative means of specifying a partition which proves to be extremely useful. This is in terms of "multiplicity vectors".

Definition 9.3. A *multiplicity vector* is an infinite sequence of nonnegative integers, only finitely many of which are nonzero. That is,

$$\rho = \langle r_1, r_2, r_3, \ldots, r_j, \ldots \rangle$$

in which $r_j \in \mathbb{N}$ for all $j \geq 1$ and $r_j \neq 0$ for only finitely many values of j. The set of all multiplicity vectors is denoted by $\mathcal{M}$. The all–zero vector $\mathbf{0}$ is a multiplicity vector, for example. We define $\ell(\mathbf{0}) := 0$, and if $\mathbf{0} \neq \rho \in \mathcal{M}$ then

$$\ell(\rho) := \max\{j : \ r_j \neq 0\}.$$

With this convention, we also use the notation

$$\rho = \langle r_1, r_2, r_3, \ldots, r_{\ell(\rho)} \rangle,$$

so that $\mathbf{0} = \langle\rangle$, for example.

Definition 9.4. Let λ be an integer partition. For each $j \geq 1$, let

$$m_j(\lambda) := \#\{i : \ \lambda_i = j\}$$

be the number of parts of λ that are equal to j, and let

$$\mathbf{m}(\lambda) := \langle m_1(\lambda), m_2(\lambda), \ldots \rangle$$

be the *multiplicty vector* of λ.

Proposition 9.5. *The function $\lambda \mapsto \mathbf{m}(\lambda)$ defines a bijection $\mathcal{Y} \rightleftharpoons \mathcal{M}$. Furthermore, if $\lambda = (\lambda_1, \ldots, \lambda_k) \in \mathcal{Y}$ corresponds to $\rho = \langle r_1, \ldots, r_\ell \rangle \in \mathcal{M}$ in this bijection, then*
(a) $k(\lambda) = r_1 + r_2 + \cdots + r_\ell$,
(b) $n(\lambda) = r_1 + 2r_2 + \cdots + \ell r_\ell$,
(c) $\lambda_1 = \ell(\rho)$,
(d) $d(\lambda) = \max\{j : \ j \leq r_j + r_{j+1} + \cdots + r_\ell\}$.

Proof. The function from $\mathcal{Y}$ to $\mathcal{M}$ is given by $\lambda \mapsto \mathbf{m}(\lambda)$. To describe the inverse function from $\mathcal{M}$ to $\mathcal{Y}$, let $\rho = \langle r_1, \ldots, r_\ell \rangle \in \mathcal{M}$. From ρ we construct the partition $(\cdots \ 3^{r_3} \ 2^{r_2} \ 1^{r_1})$, in which the notation indicates a sequence of r_1 1s, preceded by a sequence of r_2 2s, preceded by a sequence of r_3 3s, and so on. Since only finitely many of the r_j are nonzero, this produces a finite sequence of positive integers in weakly decreasing order – that is, an integer partition. It is quite easy to see that these functions are mutually inverse bijections. Parts (a), (b), and (c) are also easily verified. Part (d) is somewhat trickier, but is left as an exercise. □

Using this correspondence $\mathcal{Y} \rightleftharpoons \mathcal{M}$ between integer partitions and multiplicity vectors, we are able to deduce a wide variety of generating function identities involving certain subsets of $\mathcal{Y}$ and weight functions depending on the parameters $n(\lambda)$, $k(\lambda)$, λ_1, and $d(\lambda)$. The prototype for all these results is the following generating function for the set of all partitions with respect to size.

Theorem 9.6.

$$\Phi_{\mathcal{Y}}^n(x) = \prod_{j=1}^{\infty} \frac{1}{1-x^j}.$$

Proof. (Notice that the infinite product of formal power series on the RHS is convergent, by Proposition 7.14.) We calculate that

$$\begin{aligned}
\Phi_{\mathcal{Y}}^n(x) &:= \sum_{\lambda \in \mathcal{Y}} x^{n(\lambda)} = \sum_{\rho \in \mathcal{M}} x^{r_1 + 2r_2 + 3r_3 + \cdots} \\
&= \left(\sum_{r_1 \in \mathbb{N}} x^{r_1} \right) \left(\sum_{r_2 \in \mathbb{N}} x^{2r_2} \right) \left(\sum_{r_3 \in \mathbb{N}} x^{3r_3} \right) \cdots \\
&= \left(\frac{1}{1-x} \right) \left(\frac{1}{1-x^2} \right) \left(\frac{1}{1-x^3} \right) \cdots = \prod_{j=1}^{\infty} \frac{1}{1-x^j}.
\end{aligned}$$

The first equality is by definition. The second equality follows from the bijection $\mathcal{Y} \rightleftharpoons \mathcal{M}$ of Proposition 9.5. The fourth and fifth equalities are elementary – so the main issue is to justify the third equality, since it involves an infinite product. You might be inclined to believe that such manipulations are valid, but for completeness I will include some explanation just this once.

For each $L \in \mathbb{N}$, let $\mathcal{M}_L \subset \mathcal{M}$ be the set of those multiplicity vectors $\rho \in \mathcal{M}$ such that $\ell(\rho) \leq L$. Then $\mathcal{M} = \bigcup_{L=0}^{\infty} \mathcal{M}_L$, and it follows that

$$\begin{aligned}
\Phi_{\mathcal{M}}(x) &:= \sum_{\rho \in \mathcal{M}} x^{r_1+2r_2+\cdots} = \lim_{L\to\infty} \sum_{\rho \in \mathcal{M}_L} x^{r_1+2r_2+\cdots+Lr_L} \\
&= \lim_{L\to\infty} \left(\sum_{r_1 \in \mathbb{N}} x^{r_1} \right) \left(\sum_{r_2 \in \mathbb{N}} x^{2r_2} \right) \cdots \left(\sum_{r_L \in \mathbb{N}} x^{Lr_L} \right) \\
&= \lim_{L\to\infty} \prod_{j=1}^{L} \frac{1}{1-x^j} = \prod_{j=1}^{\infty} \frac{1}{1-x^j}.
\end{aligned}$$

For any fixed value of L, the algebraic operations applied above involve only finitely many sums and products of formal power series, and so are justified. Existence of the limit is assured by Proposition 7.14, as was already mentioned. □

Theorem 9.6 can be generalized extensively. For example, we have the following.

Theorem 9.7.

$$\Phi_{\mathcal{Y}}^{(n,k)}(x,y) := \sum_{\lambda \in \mathcal{Y}} x^{n(\lambda)} y^{k(\lambda)} = \prod_{j=1}^{\infty} \frac{1}{1-x^j y}.$$

In fact, we can make many variations on this theme by considering various sets of partitions which are defined by putting some restriction on the corresponding set of multiplicity vectors. Theorem 9.7 is just the first special case of the following general result.

Theorem 9.8. *For each integer $j \geq 1$, let $M_j \subseteq \mathbb{N}$ be a set of natural numbers. Let $\mathcal{Z} \subseteq \mathcal{Y}$ be the set of partitions $\lambda \in \mathcal{Y}$ such that $m_j(\lambda) \in M_j$ for all $j \geq 1$. Then*

$$\Phi_{\mathcal{Z}}^{(n,k)}(x,y) := \sum_{\lambda \in \mathcal{Z}} x^{n(\lambda)} y^{k(\lambda)} = \prod_{j=1}^{\infty} \left(\sum_{m \in M_j} x^{jm} y^m \right).$$

Proof. Let $\mathcal{W} \subseteq \mathcal{M}$ be the set of multiplicity vectors $\rho \in \mathcal{M}$ such that $r_j \in M_j$ for all $j \geq 1$. That is, $\mathcal{Z} \rightleftharpoons \mathcal{W}$ with the correspondence of Proposition 9.5. We calculate

that

$$\begin{aligned}\Phi_{\mathcal{Z}}^{(n,k)}(x,y) &:= \sum_{\lambda\in\mathcal{Z}} x^{n(\lambda)}y^{k(\lambda)} = \sum_{\rho\in\mathcal{W}} \left(x^{r_1+2r_2+\cdots}\right)\left(y^{r_1+r_2+\cdots}\right)\\ &= \left(\sum_{r_1\in M_1} x^{r_1}y^{r_1}\right)\left(\sum_{r_2\in M_2} x^{2r_2}y^{r_2}\right)\left(\sum_{r_3\in M_3} x^{3r_3}y^{r_3}\right)\cdots\\ &= \prod_{j=1}^{\infty}\left(\sum_{m\in M_j} x^{jm}y^{m}\right).\end{aligned}$$

(The operations involving infinite products may be justified by applying limits as in the proof of Theorem 9.6.) □

Notice that, in Theorem 9.8, if there are infinitely many $j \geq 1$ for which $0 \notin M_j$ then $\mathcal{Z} = \varnothing$. The formula of Theorem 9.8 remains valid, however, since then the infinite product on the RHS converges to zero.

Example 9.9. Let $\mathcal{O}$ denote the set of all partitions in which each part is odd. Then, in Theorem 9.8, we have $M_{2i} = \{0\}$ and $M_{2i-1} = \mathbb{N}$ for all $i \geq 1$. Therefore

$$\Phi_{\mathcal{O}}^{(n,k)}(x,y) := \sum_{\lambda\in\mathcal{O}} x^{n(\lambda)}y^{k(\lambda)} = \prod_{i=1}^{\infty}\frac{1}{1-x^{2i-1}y}.$$

Example 9.10. Let $\mathcal{D}$ denote the set of all partitions in which each part occurs at most once. That is, $\mathcal{D}$ is the set of partitions with *distinct parts*. Then, in Theorem 9.8, we have $M_j = \{0,1\}$ for all $j \geq 1$. Therefore

$$\Phi_{\mathcal{D}}^{(n,k)}(x,y) := \sum_{\lambda\in\mathcal{D}} x^{n(\lambda)}y^{k(\lambda)} = \prod_{j=1}^{\infty}(1+x^{j}y).$$

Example 9.11. For any $L \in \mathbb{N}$, let $\mathcal{R}_L$ denote the set of all partitions with $\lambda_1 \leq L$. Then, in Theorem 9.8, we have $M_j = \mathbb{N}$ for all $1 \leq j \leq L$ and $M_j = \{0\}$ for all $j > L$. Therefore

$$\Phi_{\mathcal{R}_L}^{(n,k)}(x,y) := \sum_{\lambda\in\mathcal{R}_L} x^{n(\lambda)}y^{k(\lambda)} = \prod_{j=1}^{L}\frac{1}{1-x^{j}y}.$$

Example 9.12. In Theorem 9.8, we do not have to keep track of the number $k(\lambda)$. That is, we can obtain the generating functions with respect to the single parameter $n(\lambda)$ by setting $y = 1$ in Theorem 9.8. For example, we have

$$\Phi_{\mathcal{O}}^{n}(x) = \prod_{i=1}^{\infty}\frac{1}{1-x^{2i-1}} \quad \text{and} \quad \Phi_{\mathcal{D}}^{n}(x) = \prod_{j=1}^{\infty}(1+x^{j}).$$

Now, notice the following:

$$\begin{aligned}\Phi_{\mathcal{D}}^n(x) &= \prod_{j=1}^{\infty}(1+x^j) = \prod_{j=1}^{\infty}(1+x^j)\left(\frac{1-x^j}{1-x^j}\right) \\ &= \prod_{j=1}^{\infty}\frac{1-x^{2j}}{1-x^j} = \frac{\prod_{i=1}^{\infty}(1-x^{2i})}{\prod_{i=1}^{\infty}(1-x^i)} \\ &= \prod_{i=1}^{\infty}\frac{1}{1-x^{2i-1}} = \Phi_{\mathcal{O}}^n(x).\end{aligned}$$

Since these two formal power series are equal, for every $n \in \mathbb{N}$ we have $[x^n]\Phi_{\mathcal{D}}(x) = [x^n]\Phi_{\mathcal{O}}(x)$. That is, for every $n \in \mathbb{N}$, the number of partitions of size n with distinct parts equals the number of partitions of size n with only odd parts. We have shown that these finite sets have the same cardinality without constructing a bijection between them! Neither have we obtained an explicit formula for the cardinalities of these sets. That's kind of amazing when you think about it.

Definition 9.13. Given a partition $\lambda \in \mathcal{Y}$, the *conjugate partition* of λ is obtained by taking the transpose of the Ferrers diagram F_λ and reading off the number of boxes in each row of the transposed diagram. The conjugate of λ is denoted by $\widetilde{\lambda}$. Notice that the conjugate of $\widetilde{\lambda}$ is λ itself. (See Figure 9.2.) For any partition $\lambda \in \mathcal{Y}$ we have

$$\begin{aligned}n(\lambda) &= n(\widetilde{\lambda}) \\ k(\lambda) &= \widetilde{\lambda}_1 \\ \lambda_1 &= k(\widetilde{\lambda}) \\ d(\lambda) &= d(\widetilde{\lambda})\end{aligned}$$

(A numerical formula for the conjugate partition is given in Exercise 9.2.)

Example 9.14. For any $L \in \mathbb{N}$, let $\mathcal{C}_L$ denote the set of all partitions with $k(\lambda) \leq L$. That is, $\mathcal{C}_L$ consists of those partitions with at most L parts. In this case, Theorem 9.8 does not apply, since the set $\mathcal{C}_L$ is not defined in terms of some simple condition on the corresponding set of multiplicity vectors. However, notice that a partition λ is in $\mathcal{C}_L$ if and only if $\widetilde{\lambda}$ is in $\mathcal{R}_L$. That is, we have a bijection $\mathcal{C}_L \rightleftharpoons \mathcal{R}_L$ via the correspondence $\lambda \leftrightarrow \widetilde{\lambda}$. Therefore,

$$\Phi_{\mathcal{C}_L}^{(n,\lambda_1)}(x,y) = \Phi_{\mathcal{R}_L}^{(n,k)}(x,y) = \prod_{j=1}^{L}\frac{1}{1-x^jy}.$$

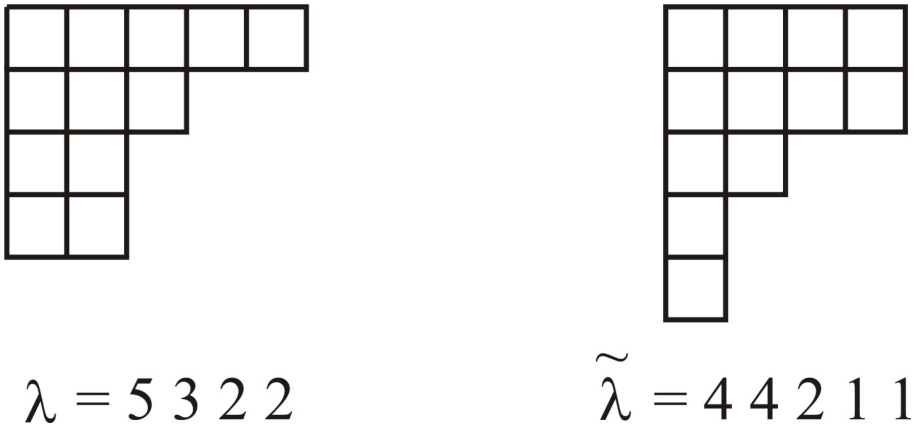


FIGURE 9.2. a partition and its conjugate.

Example 9.15. A partition λ is called *self–conjugate* provided that $\widetilde{\lambda} = \lambda$. Let $\mathcal{SC}$ denote the set of self–conjugate partitions. We claim that

$$\Phi_{\mathcal{SC}}^{(n,d)}(x,y) = \prod_{i=1}^{\infty} \left(1 + x^{2i-1}y\right).$$

To prove this, let $\mathcal{OD}$ be the set of partitions which have only odd parts and in which each part occurs at most once; that is, partitions with **odd and distinct** parts. Theorem 9.8 implies that

$$\Phi_{\mathcal{OD}}^{(n,k)} = \prod_{i=1}^{\infty} \left(1 + x^{2i-1}y\right).$$

To complete the proof of the formula for $\Phi_{\mathcal{SC}}^{(n,d)}(x,y)$, it suffices to find a bijection with the following properties:

$$\begin{aligned} \mathcal{SC} &\rightleftharpoons \mathcal{OD} \\ \lambda &\leftrightarrow \mu \\ n(\lambda) &= n(\mu) \\ d(\lambda) &= k(\mu) \end{aligned}$$

We will define a pair of functions $\psi : \mathcal{SC} \to \mathcal{OD}$ and $\phi : \mathcal{OD} \to \mathcal{SC}$ which are mutually inverse bijections. The effect of these functions is illustrated in Figure 9.3, but verification of the details is left as an exercise. (Recall that ε denotes the empty partition.)

```
FUNCTION: ψ from SC to OD;
INPUT: λ;
if λ = ε then let μ := ε;
```

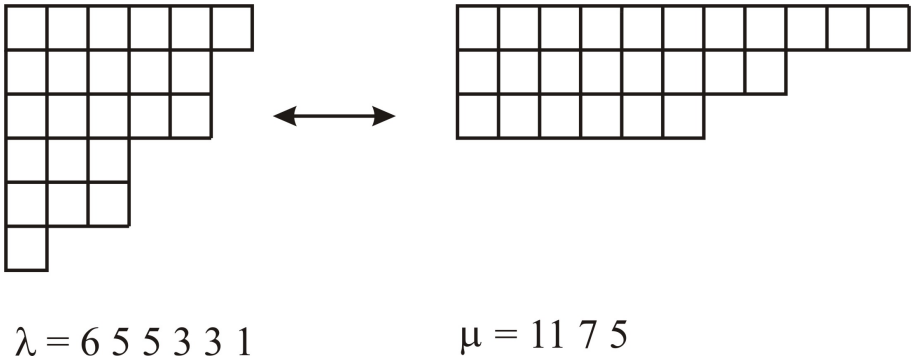


FIGURE 9.3. the bijection of Example 9.15.

```
if λ ≠ ε then remove the first row and column of F_λ
  to obtain the Ferrers diagram of a partition θ;
  let μ := (2λ_1 − 1, ψ(θ));
end if;
OUTPUT: μ.

FUNCTION: φ from 𝒪𝒟 to 𝒮𝒞;
INPUT: μ = (μ_1, ..., μ_k);
if μ = ε then let λ := ε;
if μ ≠ ε then
  let a := (μ_1 − 1)/2;
  let θ := φ(μ_2, ..., μ_k);
  join a column of a boxes to the left of F_θ;
  join a row of a + 1 boxes to the top of the result
    of the previous line;
  let λ be the partition which has the resulting
    Ferrers diagram;
end if;
OUTPUT: λ.
```

Finally, we conclude this section with two examples of what are called *Euler Identities*. Another example is given in the exercises. They can all be proved by combinatorial manipulations with Ferrers diagrams. Some more complicated – and much more important! – identities are discussed in the next section.

Theorem 9.16.

$$\prod_{j=1}^{\infty} \frac{1}{1-x^j y} = \sum_{d=0}^{\infty} \frac{x^{d^2} y^d}{(1-xy)(1-x^2y)\cdots(1-x^d y)(1-x)(1-x^2)\cdots(1-x^d)}.$$

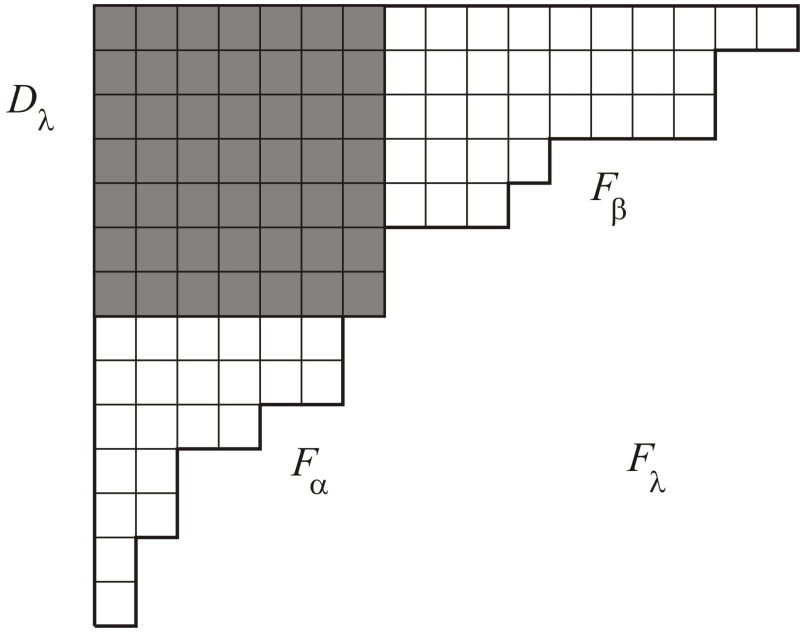


FIGURE 9.4. illustration of Theorem 9.16.

Proof. By Theorem 9.7, the LHS is the generating function $\Phi_{\mathcal{Y}}^{(n,k)}(x,y)$ for the set of all partitions with respect to both size and length. It remains to interpret the RHS in the same way.

Consider any partition $\lambda \in \mathcal{Y}$, and the corresponding Ferrers diagram F_λ. The number of boxes on the main diagonal of F_λ is, by definition, $d(\lambda)$. Thus, we can fit a square D of size $d(\lambda)$–by–$d(\lambda)$ inside of F_λ so that the upper–left corners of D and of F_λ coincide. Furthermore, no larger square will fit inside of F_λ in this way. This square is called the *Durfee square* of λ, and is denoted by D_λ. Each box of F_λ is either in D_λ, to the right of D_λ, or directly below D_λ. The boxes of F_λ which are below D_λ are in the shape of a Ferrers diagram of some partition – let's call this partition α. The boxes of F_λ which are to the right of D_λ are in the shape of a Ferrers diagram of some partition – let's call this partition β. So from λ we construct the triple $(d, \alpha, \beta) \in \mathbb{N} \times \mathcal{Y} \times \mathcal{Y}$ in which $d := d(\lambda)$ and α and β are defined as above. Figure 9.4 illustrates this construction.

Not every triple in $\mathbb{N} \times \mathcal{Y} \times \mathcal{Y}$ is constructed by this procedure. After a bit of thought, one sees that (d, α, β) occurs if and only if F_α has at most d rows and F_β has at most d columns. That is, with the notations of Examples 9.11 and 9.14, if and only if $\alpha \in \mathcal{R}_d$ and $\beta \in \mathcal{C}_d$. That is, the set of triples (d, α, β) which occur is

$$\bigcup_{d=0}^{\infty} (\{d\} \times \mathcal{R}_d \times \mathcal{C}_d) .$$

In other words, we have defined a bijection with the following properties:

$$\begin{aligned} \mathcal{Y} &\rightleftharpoons \bigcup_{d=0}^{\infty} (\{d\} \times \mathcal{R}_d \times \mathcal{C}_d) \\ \lambda &\leftrightarrow (d, \alpha, \beta) \\ n(\lambda) &= d^2 + n(\alpha) + n(\beta) \\ k(\lambda) &= d + k(\alpha) \end{aligned}$$

The last two statements are easily verified from the definition of the correspondence. It follows that

$$\begin{aligned} \prod_{j=1}^{\infty} \frac{1}{1-x^j y} &= \Phi_{\mathcal{Y}}^{(n,k)}(x,y) \\ &= \sum_{d=0}^{\infty} \sum_{\alpha \in \mathcal{R}_d} \sum_{\beta \in \mathcal{C}_d} \left(x^{d^2+n(\alpha)+n(\beta)} \right) \left(y^{d+k(\alpha)} \right) \\ &= \sum_{d=0}^{\infty} x^{d^2} y^d \left(\sum_{\alpha \in \mathcal{R}_d} x^{n(\alpha)} y^{k(\alpha)} \right) \left(\sum_{\beta \in \mathcal{C}_d} x^{n(\alpha)} \right) \\ &= \sum_{d=0}^{\infty} \frac{x^{d^2} y^d}{(1-xy)(1-x^2y)\cdots(1-x^dy)(1-x)(1-x^2)\cdots(1-x^d)}, \end{aligned}$$

by Examples 9.11 and 9.14. This completes the proof. □

Theorem 9.17.

$$\prod_{j=1}^{\infty} (1+x^j y) = \sum_{k=0}^{\infty} \frac{x^{k(k+1)/2} y^k}{(1-x)(1-x^2)\cdots(1-x^k)}.$$

Proof. By Example 9.10, the LHS is the generating function $\Phi_{\mathcal{D}}^{(n,k)}(x,y)$ for the set $\mathcal{D}$ of partitions with distinct parts, with respect to both size and length. It remains to interpret the RHS in the same way.

Consider a partition $\lambda \in \mathcal{D}$ of length $k(\lambda) = k$. Then $\lambda_1 > \lambda_2 > \cdots > \lambda_k \geq 1$, from which it follows that $\lambda_i \geq k+1-i$ for each $1 \leq i \leq k$. In other words, if θ denotes the partition of length k for which $\theta_i := k+1-i$ for all $1 \leq i \leq k$, then the "k–th staircase shape" F_θ is contained in F_λ, and no larger staircase shape is contained in F_λ. (See Figure 9.5.)

Now, for each $1 \leq i \leq k$, let $\mu_i := \lambda_i - (k+1-i)$. Since λ has distinct parts, it follows that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_k \geq 0$. Some of the parts μ_i may be zero, but if we remove these then we obtain a partition μ with at most k parts. Conversely, if we are given a partition μ with at most k parts, then by defining $\lambda_i := \mu_i + (k+1-i)$ for all $1 \leq i \leq k$ we obtain a partition λ with exactly k parts, no two of which are

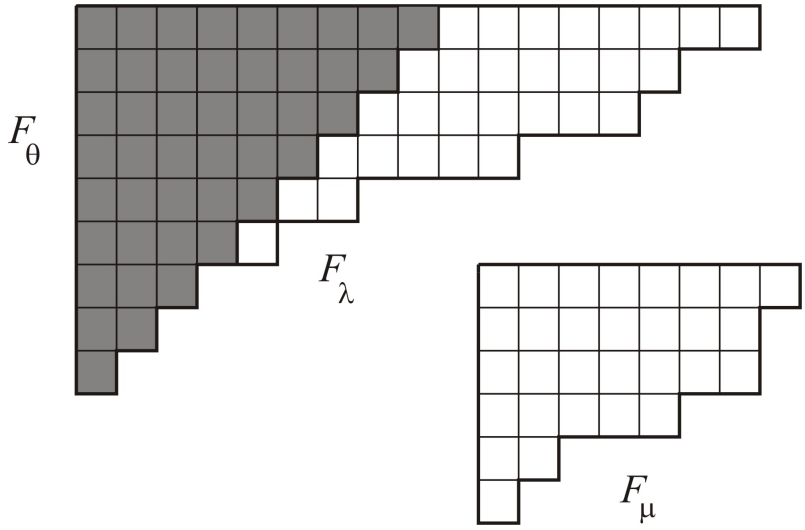


FIGURE 9.5. illustration of Theorem 9.17.

equal. (Here we use the convention that $\mu_i = 0$ if $i > k(\mu)$.) In short, we have constructed a bijection with the following properties:

$$\begin{aligned} \mathcal{D} &\rightleftharpoons \bigcup_{k=0}^{\infty} (\{k\} \times \mathcal{C}_k) \\ \lambda &\leftrightarrow (k, \mu) \\ n(\lambda) &= (1 + 2 + \cdots + k) + n(\mu) \\ k(\lambda) &= k \end{aligned}$$

Since $1 + 2 + \cdots + k = k(k+1)/2$ we calculate that

$$\begin{aligned} \prod_{j=1}^{\infty}(1 + x^j y) &= \Phi_{\mathcal{D}}^{(n,k)}(x, y) = \sum_{k=0}^{\infty} \sum_{\mu \in \mathcal{C}_k} x^{k(k+1)/2 + n(\mu)} y^k \\ &= \sum_{k=0}^{\infty} x^{k(k+1)/2} y^k \left(\sum_{\mu \in \mathcal{C}_k} x^{n(\mu)} \right) \\ &= \sum_{k=0}^{\infty} \frac{x^{k(k+1)/2} y^k}{(1-x)(1-x^2)\cdots(1-x^k)}, \end{aligned}$$

by Example 9.14. This completes the proof. □

9. Exercises.

1. Write down the generating functions for each of the following sets of partitions, with a partition λ contributing $x^{n(\lambda)}y^{k(\lambda)}$ to the generating function. You need not explain your reasoning.

(a) Partitions in which only even parts occur.

(b) Partitions in which odd parts occur at most once.

(c) Partitions such that each even part is divisible by four, and each odd part occurs an even number of times.

(d) Partitions in which the multiplicity of a part j is either 0 or is of the same parity (odd or even) as j.

2(a) Prove Proposition 9.5(d).

(b) Let λ be an integer partition, and let $\widetilde{\lambda}$ be its conjugate partition. Show that for all $1 \leq k \leq \lambda_1$,

$$\widetilde{\lambda}_k = \#\{i : \ \lambda_i \geq k\}.$$

3(a) Let $\mathcal{A}$ be the set of partitions in which each part occurs at most three times. Obtain a formula for the generating function of this set with respect to the size of the partitions.

(b) Let $\mathcal{B}$ be the set of partitions in which each even part occurs at most once. Obtain a formula for the generating function of this set with respect to the size of the partitions.

(c) Prove that for each $n \in \mathbb{N}$, the number of partitions of size n in $\mathcal{A}$ equals the number of partitions of size n in $\mathcal{B}$.

4(a) Let $\mathcal{A}$ be the set of partitions in which each part may occur 0, 1, 4, or 5 times. Obtain a formula for the generating function of this set with respect to the size of the partitions.

(b) Let $\mathcal{B}$ be the set of partitions which have no parts congruent to 2 (mod 4), and in which parts divisible by four occur at most once each. Obtain a formula for the generating function of this set with respect to the size of the partitions.

(c) Prove that for all $n \in \mathbb{N}$, the number of partitions in $\mathcal{A}$ of size n equals the number of partitions in $\mathcal{B}$ of size n.

5(a) Let $\mathcal{A}$ be the set of partitions in which no part occurs exactly once. Obtain a formula for the generating function of this set with respect to the size of the partitions.

(b) Let $\mathcal{B}$ be the set of partitions in which every odd part is divisible by 3. Obtain a formula for the generating function of this set with respect to the size of the partitions.

(c) Prove that for all $n \in \mathbb{N}$, the number of partitions in $\mathcal{A}$ of size n equals the number of partitions in $\mathcal{B}$ of size n.

---

6. Let $\mathrm{pe}(n)$ be the number of partitions of size n with an even number of parts, and let $\mathrm{po}(n)$ be the number of partitions of size n with an odd number of parts. Let $\mathrm{od}(n)$ be the number of partitions of size n which have odd and distinct parts. Show that for all $n \in \mathbb{N}$:

$$\mathrm{pe}(n) - \mathrm{po}(n) = (-1)^n \mathrm{od}(n).$$

---

7. For $n \in \mathbb{N}$, let $a(n)$ be the number of partitions of size n with an even number of even parts, let $b(n)$ be the number of partitions of size n with an odd number of even parts, and let $\mathrm{od}(n)$ be the number of partitions of size n with odd and distinct parts. Show that for all $n \in \mathbb{N}$,

$$a(n) = b(n) + \mathrm{od}(n).$$

8(a) For a partition λ, let $m_1(\lambda)$ be the number of times 1 occurs as a part of λ. Obtain a formula for the generating function

$$A(x, y) = \sum_{\lambda \in \mathcal{Y}} x^{n(\lambda)} y^{m_1(\lambda)}.$$

(b) For a partition λ, let $b(\lambda)$ be the number of different sizes of parts which occur in λ. (For example, $b(7\ 6\ 4\ 4\ 4\ 2\ 2) = 4$.) Obtain a formula for the generating function

$$B(x, y) = \sum_{\lambda \in \mathcal{Y}} x^{n(\lambda)} y^{b(\lambda)}.$$

(c) Show that for all $n \in \mathbb{N}$, the sum of $m_1(\lambda)$ over all partitions of size n is equal to the sum of $b(\lambda)$ over all partitions of size n.

---

9. Show that

$$\prod_{i=1}^{\infty}(1+x^{2i-1}y)=\sum_{d=0}^{\infty}\frac{x^{d^2}y^d}{(1-x^2)(1-x^4)\cdots(1-x^{2d})}.$$

10.* In Example 9.12 we saw that $\Phi_{\mathcal{D}}^n(x)=\Phi_{\mathcal{O}}^n(x)$. By Proposition 4.7 there is thus a weight–preserving bijection between the sets $\mathcal{D}$ and $\mathcal{O}$. Give an explicit description of such a bijection.

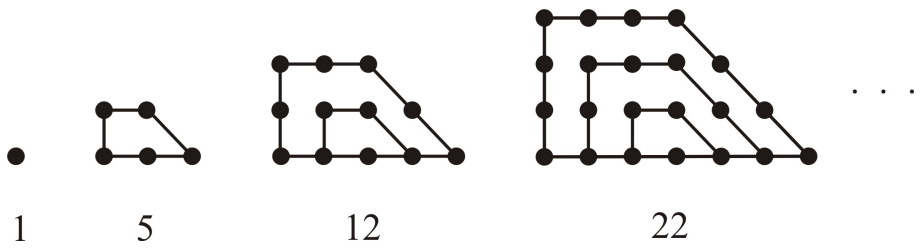


FIGURE 10.1. the pentagonal numbers.

10. More about Integer Partitions.

In this section we'll see a few of the more interesting facts from the theory of partitions – Euler's Pentagonal Number Theorem, Jacobi's Triple Product Formula, and Ramanujan's First Partition Congruence. The theory goes much further – a few references and indications of this are given in the Endnotes.

Euler's Pentagonal Number Theorem

The sequence of *pentagonal numbers* begins as shown in Figure 10.1, by analogy with the more familiar sequences of *triangular numbers* $h(h+1)/2$ and *square numbers* h^2, for $h \in \mathbb{N}$. As indicated in Figure 10.1, the h–th pentagonal number is $h^2 + \binom{h}{2} = h(3h-1)/2$ for all $h \in \mathbb{N}$. This accounts for the name of Theorem 10.1.

Theorem 10.1 (Euler's Pentagonal Number Theorem)**.**

$$\prod_{j=1}^{\infty}(1-x^j) = \sum_{h=-\infty}^{\infty}(-1)^h x^{h(3h-1)/2}$$

Notice that $(-h)(3(-h)-1)/2 = h(3h+1)/2$, so that the formula of the theorem my be rewritten as

$$\begin{aligned}\prod_{j=1}^{\infty}(1-x^j) &= 1+\sum_{h=1}^{\infty}(-1)^h\left(x^{h(3h-1)/2}+x^{h(3h+1)/2}\right)\\ &= 1-x-x^2+x^5+x^7-x^{12}-x^{15}+x^{22}+x^{26}-x^{35}-x^{40}+\cdots\end{aligned}$$

The product on the LHS of the formula is the reciprocal of the generating function $\Phi_y^n(x)$ for all partitions with respect to size. That fact leads to several applications of this formula, as we shall see.

Proof of EPNT. Recall from Example 9.10 that

$$\Phi_{\mathcal{D}}^{(n,k)}(x,y) = \prod_{j=1}^{\infty}(1+x^j y),$$

in which $\mathcal{D}$ denotes the set of partitions with distinct parts. Upon making the substitution $y=-1$ in this we obtain the LHS of the EPNT, yielding the combinatorial interpretation

$$\prod_{j=1}^{\infty}(1-x^j) = \sum_{\lambda\in\mathcal{D}}(-1)^{k(\lambda)}x^{n(\lambda)}$$

for the product. (Note that the substitution $y=-1$ is valid since, by Proposition 7.14, the infinite product is still well–defined.) In this formula, a partition with an even number of parts contributes $x^{n(\lambda)}$ to the series, while a partition with an odd number of parts contributes $-x^{n(\lambda)}$. For each $n\in\mathbb{N}$, let $\overline{\text{pe}}(n)$ denote the number of partitions $\lambda\in\mathcal{D}$ of size n which have an even number of parts, and let $\overline{\text{po}}(n)$ denote the number of partitions $\lambda\in\mathcal{D}$ of size n which have an odd number of parts. Collecting like powers of x in the previous equation, we obtain

$$\prod_{j=1}^{\infty}(1-x^j) = \sum_{n=0}^{\infty}\left(\overline{\text{pe}}(n)-\overline{\text{po}}(n)\right)x^n.$$

To finish the proof, we must show that

$$\overline{\text{pe}}(n)-\overline{\text{po}}(n) = \begin{cases} (-1)^h & \text{if } n=h(3h-1)/2 \text{ for some } h\in\mathbb{Z}, \\ 0 & \text{otherwise.}\end{cases}$$

The combinatorial argument we will use to prove this is due to Franklin in 1881. (Euler's original proof in the 1750s was purely algebraic.)

We will construct a function $\phi:\mathcal{D}\to\mathcal{D}$ with the following properties.
(i) For every $\lambda\in\mathcal{D}$, $n(\phi(\lambda))=n(\lambda)$.
(ii) For every $\lambda\in\mathcal{D}$, $\phi(\phi(\lambda))=\lambda$.
(iii) For every $\lambda\in\mathcal{D}$, either $\phi(\lambda)=\lambda$ or $|k(\phi(\lambda))-k(\lambda)|=1$.
(iv) For every $n\in\mathbb{N}$, there is at most one partition $\lambda\in\mathcal{D}$ of size n for which $\phi(\lambda)=\lambda$.

Before showing that such a function exists, let's see how it suffices to complete the proof of the theorem. By condition (i), for any $\lambda\in\mathcal{D}$, both λ and $\phi(\lambda)$ have the same size. By condition (ii), $\phi:\mathcal{D}\to\mathcal{D}$ is its own inverse function, and so ϕ is a bijection. In fact, ϕ is a permutation on the set $\mathcal{D}$ in which each cycle has exactly one or two elements; such a permutation is called an *involution*. By condition (iii) each cycle of ϕ with two elements consists of one partition with an even number

FIGURE 10.2. two Ferrers diagrams with Xs and Os.

of parts and one partition with an odd number of parts. By condition (i) these two partitions have the same size, say $n \in \mathbb{N}$. The total contribution of these two partitions to the difference $\overline{\mathrm{pe}}(n) - \overline{\mathrm{po}}(n)$ is therefore $1-1=0$, so these cycles of ϕ of length two end up contributing nothing to the difference in question. Therefore, to determine the value of $\overline{\mathrm{pe}}(n) - \overline{\mathrm{po}}(n)$ we need only consider the contribution of those partitions $\lambda \in \mathcal{D}$ of size n which are such that $\phi(\lambda) = \lambda$. By condition (iv), there is at most one such partition of size n, for each $n \in \mathbb{N}$. Examining the definition of ϕ in more detail, we will be able to determine whether this partition contributes $+1$ or -1 and to establish the desired formula, thereby completing the proof of the theorem.

Now we construct the function $\phi : \mathcal{D} \to \mathcal{D}$ and verify that it has the required properties. Given a partition $\lambda = (\lambda_1 > \lambda_2 > \cdots > \lambda_k) \in \mathcal{D}$, consider its Ferrers diagram F_λ. Put an X in each box in the last row of F_λ, and let $\mathrm{ex}(\lambda)$ be the number of Xs in F_λ. Also, put an O in the rightmost box of the first row of F_λ, and in every box of the reverse diagonal containing that box. That is, put an O in the last box of row i if and only if $\lambda_i = \lambda_1 + 1 - i$. Let $\mathrm{oh}(\lambda)$ denote the number of Os in F_λ. (Figure 10.2 illustrates two examples of these definitions.) Notice that $\mathrm{ex}(\varepsilon) = \mathrm{oh}(\varepsilon) = 0$ and that if $\lambda \neq \varepsilon$ then $1 \leq \mathrm{ex}(\lambda)$ and $1 \leq \mathrm{oh}(\lambda) \leq k(\lambda)$.

The partitions $\lambda \in \mathcal{D}$ fall into two cases and several subcases, depending on how the boxes of F_λ are marked by the Xs and Os. The definition of ϕ depends crucially on this case analysis.

Case I: $\mathrm{ex}(\lambda) > \mathrm{oh}(\lambda)$.
I(a): $\mathrm{ex}(\lambda) \geq \mathrm{oh}(\lambda) + 2$.
I(b): $\mathrm{ex}(\lambda) = \mathrm{oh}(\lambda) + 1$ and no boxes of F_λ are marked with both an X and an O.
I(c): $\mathrm{ex}(\lambda) = \mathrm{oh}(\lambda) + 1$ and some box of F_λ is marked with both an X and an O.

Case II: $\mathrm{ex}(\lambda) \leq \mathrm{oh}(\lambda)$.
II(a): $\mathrm{ex}(\lambda) < \mathrm{oh}(\lambda)$.
II(b): $\mathrm{ex}(\lambda) = \mathrm{oh}(\lambda) > 0$ and no boxes of F_λ are marked with both an X and an O.
II(c): $\mathrm{ex}(\lambda) = \mathrm{oh}(\lambda)$ and either $\lambda = \varepsilon$ or some box of F_λ is marked with both an X

and an O.

Figure 10.3 shows one partition λ from each of these subcases, on the left, as well as the corresponding partition $\phi(\lambda)$ on the right. Notice that a nonempty F_λ has a box marked with both an X and an O if and only if $\mathrm{oh}(\lambda) = k(\lambda)$, and that there is at most one such box.

It remains to define the function $\phi : \mathcal{D} \to \mathcal{D}$. To do this, consider a partition $\lambda \in \mathcal{D}$. Consider which of the subcases includes the partition λ.

For subcases I(a) or I(b): Remove the boxes marked O from F_λ and adjoin a new row of $\mathrm{oh}(\lambda)$ boxes to the bottom of F_λ. This produces the Ferrers diagram of the partition $\phi(\lambda)$.

For subcases II(a) or II(b): Remove the boxes marked X from F_λ and adjoin one new box to the right of each of the top $\mathrm{ex}(\lambda)$ rows of F_λ. This produces the Ferrers diagram of the partition $\phi(\lambda)$.

For subcases I(c) or II(c): Let $\phi(\lambda) := \lambda$.

Of course, there are now **many** details to check. Before verifying conditions (i) to (iv) for this construction, we should verify that, given $\lambda \in \mathcal{D}$, the output $\phi(\lambda)$ is also a partition in $\mathcal{D}$. In cases I(c) and II(c) this is obvious. In cases I(a) or I(b), let

$$\lambda = (\lambda_1 > \cdots > \lambda_k)$$

and let

$$b := \mathrm{oh}(\lambda).$$

Then

$$\phi(\lambda) := (\lambda_1 - 1 > \cdots > \lambda_b - 1 > \lambda_{b+1} > \cdots > \lambda_k > b).$$

The strict inequalities $\lambda_1 - 1 > \cdots > \lambda_b - 1$ follow from the corresponding inequalities for λ. The strict inequality $\lambda_b - 1 > \lambda_{b+1}$ is a consequence of the definition of $b := \mathrm{oh}(\lambda)$. The strict inequality $\lambda_k > b$ is a restatement of the inequality $\mathrm{ex}(\lambda) > \mathrm{oh}(\lambda)$. In case I(a) we may have $b = k$, in which case the strict inequality $\lambda_k - 1 > b$ does hold. In case I(b) we can not have $b = k$, since there is no box of F_λ marked both X and O. Therefore, in cases I(a) or I(b), the resulting $\phi(\lambda)$ really is a partition with distinct parts. Verification of this fact in subcases II(a) or II(b) is left as an exercise.

Next, we proceed to check conditions (i) to (iv) for ϕ. Condition (i) is easy to see, since in each case the number of boxes which are removed from F_λ is equal to the number of boxes which are reattached. Therefore $n(\phi(\lambda)) = n(\lambda)$, as desired. Condition (iii) is also clear, since in cases I(a) or I(b) we have $k(\phi(\lambda)) = k(\lambda) + 1$, while in cases II(a) or II(b) we have $k(\phi(\lambda)) = k(\lambda) - 1$.

Next consider condition (ii). If $\lambda \in \mathcal{D}$ is in case I(c) or II(c) then $\phi(\phi(\lambda)) = \lambda$ certainly holds. If λ is in case II(a) or II(b) then $\lambda_k = \mathrm{ex}(\lambda) \leq \mathrm{oh}(\lambda) \leq k(\lambda)$, and one of the two inequalities must be strict. Considering the operation of ϕ, we see

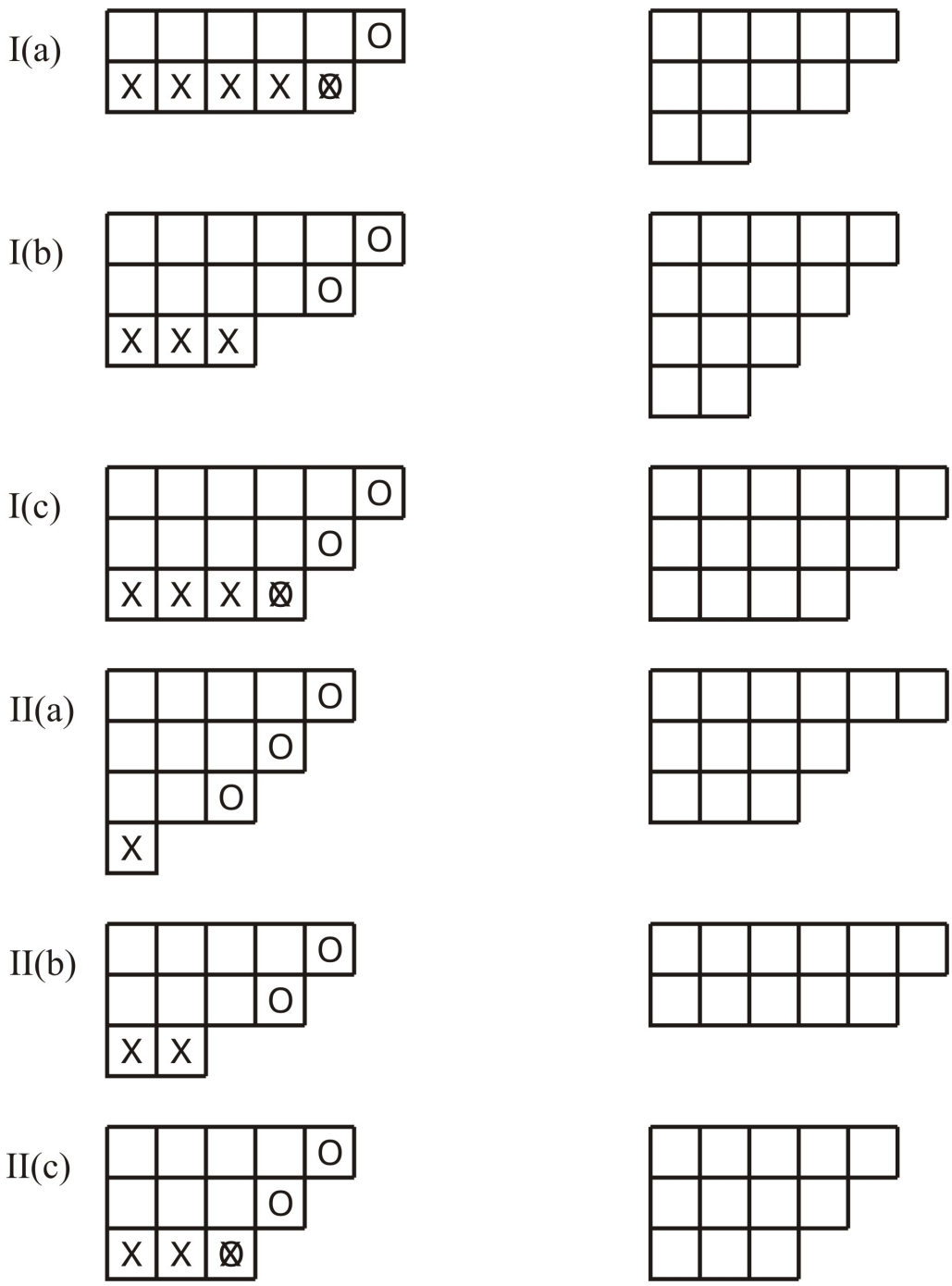


FIGURE 10.3. the six subcases to define ϕ.

that

$$\mathrm{oh}(\phi(\lambda)) = \mathrm{ex}(\lambda) = \lambda_k$$

and that

$$\mathrm{ex}(\phi(\lambda)) = \lambda_{k-1} + \delta,$$

in which

$$\delta := \begin{cases} 1 & \text{if } \lambda_k = k-1, \\ 0 & \text{if } \lambda_k \leq k-2. \end{cases}$$

Since

$$\mathrm{oh}(\phi(\lambda)) = \lambda_k < \lambda_{k-1} \leq \mathrm{ex}(\phi(\lambda)),$$

the partition $\phi(\lambda)$ is in Case I. Also notice that $\delta = 1$ if and only if the Ferrers diagram of $\phi(\lambda)$ has a box marked both X and O. Thus, if there is such a box then $\mathrm{oh}(\phi(\lambda)) + 2 \leq \mathrm{ex}(\phi(\lambda))$, so that $\phi(\lambda)$ is not in subcase I(c). We have shown that if λ is in subcase II(a) or II(b) then $\phi(\lambda)$ is in subcase I(a) or I(b). We leave as an exercise the similar verification that if λ is in subcase I(a) or I(b) then $\phi(\lambda)$ is in subcase II(a) or II(b). From these claims and the definition of ϕ, it readily follows that $\phi(\phi(\lambda)) = \lambda$ for all $\lambda \in \mathcal{D}$.

It remains to check condition (iv) for ϕ, and more precisely, to determine the (signed) generating function for those partitions $\lambda \in \mathcal{D}$ such that $\phi(\lambda) = \lambda$. First consider a partition $\lambda \in \mathcal{D}$ in subcase II(c). We have $\mathrm{ex}(\lambda) = \mathrm{oh}(\lambda) = k(\lambda)$, so for each $h \geq 0$ there is exactly one choice for λ: the partitions of the form

$$\lambda = (2h-1 > 2h-2 > \cdots > h+1 > h)$$

satisfy subcase II(c), and these are all. The size of this partition is $n(\lambda) = h^2 + \binom{h}{2} = h(3h-1)/2$ and its length is $k(\lambda) = h$. (Notice that the case $h = 0$ corresponds to the empty partition ε.) Next, consider a partition $\lambda \in \mathcal{D}$ in subcase I(c). We have $\mathrm{ex}(\lambda) = \mathrm{oh}(\lambda) + 1 = k(\lambda) + 1$, so for each $h \geq 1$ there is exactly one choice for λ: the partitions of the form

$$\lambda = (2h > 2h-1 > \cdots > h+2 > h+1)$$

satisfy subcase I(c), and these are all. The size of this partition is $n(\lambda) = h^2 + \binom{h+1}{2} = h(3h+1)/2$ and its length is $k(\lambda) = h$. The contribution of all partitions in $\mathcal{D}$ in subcases I(c) and II(c) is therefore

$$\begin{aligned} \sum_{\lambda \in \mathcal{D}:\ \phi(\lambda) = \lambda} (-1)^{k(\lambda)} x^{n(\lambda)} &= \sum_{h=0}^{\infty} (-1)^h x^{h(3h-1)/2} + \sum_{h=1}^{\infty} (-1)^h x^{h(3h+1)/2} \\ &= \sum_{h=-\infty}^{\infty} (-1)^h x^{h(3h-1)/2}. \end{aligned}$$

As explained above, this suffices to complete the proof of Euler's Pentagonal Number Theorem. $\square$

Jacobi's Triple Product Formula

Theorem 10.2 (Jacobi's Triple Product Formula)**.**

$$\sum_{h=-\infty}^{\infty} x^{h^2} y^h = \prod_{i=1}^{\infty} (1 + x^{2i-1}y)(1 + x^{2i-1}y^{-1})(1 - x^{2i})$$

Proof. First of all, notice that the infinite product on the RHS is well–defined when considered as a formal power series in the indeterminate x with coefficients in the ring $\mathbb{Z}((y))$.

To prove this rather mysterious–looking identity we rewrite it in the form

$$\left(\sum_{h=-\infty}^{\infty} x^{h^2} y^h \right) \prod_{i=1}^{\infty} \frac{1}{1 - x^{2i}} = \prod_{i=1}^{\infty} (1 + x^{2i-1}y)(1 + x^{2i-1}y^{-1})$$

and interpret both sides combinatorially as generating functions for some sets of objects. Then we establish the identity by constructing a suitable bijection between these two sets.

The infinite product on the LHS is the generating function $\Phi^n_{\mathcal{E}}(x)$ for the set $\mathcal{E}$ of partitions with only even parts, with respect to size. Therefore, the LHS is the generating function for the set $\mathbb{Z} \times \mathcal{E}$, in which the pair $(h, \lambda) \in \mathbb{Z} \times \mathcal{E}$ contributes the monomial $x^{h^2+n(\lambda)} y^h$. The RHS can be written as

$$\prod_{i=1}^{\infty} (1 + x^{2i-1}y) \cdot \prod_{i=1}^{\infty} (1 + x^{2i-1}y^{-1}).$$

The first of these infinite products is the generating function $\Phi^{(n,k)}_{\mathcal{OD}}(x, y)$ for the set $\mathcal{OD}$ of partitions with odd and distinct parts, with respect to both size and length. By Example 9.15, this is also the generating function $\Phi^{(n,d)}_{\mathcal{SC}}(x, y)$ for the set $\mathcal{SC}$ of self–conjugate partitions with respect to both size and diagonal. The second of these factors is obtained from the first by making the substitution of y^{-1} in place of y, so it is $\Phi^{(n,d)}_{\mathcal{SC}}(x, y^{-1})$. Thus, the RHS of the identity we wish to prove is the generating function for the set $\mathcal{SC} \times \mathcal{SC}$ in which the pair (α, β) contributes the monomial $x^{n(\alpha)+n(\beta)} y^{d(\alpha)-d(\beta)}$.

Having interpreted the identity as in the previous paragraph, to complete the proof it suffices to construct a bijection with the following properties:

$$\begin{aligned} \mathbb{Z} \times \mathcal{E} &\rightleftharpoons \mathcal{SC} \times \mathcal{SC} \\ (h, \lambda) &\leftrightarrow (\alpha, \beta) \\ h^2 + n(\lambda) &= n(\alpha) + n(\beta) \\ h &= d(\alpha) - d(\beta) \end{aligned}$$

This we proceed to do.

For one direction, begin with a pair $(h, \lambda) \in \mathbb{Z} \times \mathcal{E}$. Since every part of λ is even, there is a unique partition μ of the same length as λ such that $\lambda_i = 2\mu_i$ for each $1 \leq i \leq k(\lambda)$. To construct (α, β), begin with an array of dots in the shape of the Ferrers diagram F_μ of μ. Attach a $|h|$–by–$|h|$ square array D of dots to the top of F_μ, with its left edge aligned with the left edge of F_μ. Then draw an array of open circles in the shape of the Ferrers diagram $F_{\widetilde{\mu}}$ of $\widetilde{\mu}$ just to the right of D, with its top edge aligned with the top edge of D. Some open circles in the shape $F_{\widetilde{\mu}}$ might have to be drawn around dots in the shape F_μ – that's okay, just go ahead and do that.

For example, beginning with the pair $h = -2$ and $\lambda = 10\ 6\ 2\ 2$, we obtain partitions $\mu = 5\ 3\ 1\ 1$ and $\widetilde{\mu} = 4\ 2\ 2\ 1\ 1$, and draw the picture

$$\begin{array}{cccccc}
\bullet & \bullet & \circ & \circ & \circ & \circ \\
\bullet & \bullet & \circ & \circ & & \\
\bullet & \bullet & \odot & \odot & \bullet & \\
\bullet & \bullet & \odot & & & \\
\bullet & & \circ & & & \\
\bullet & & & & &
\end{array}$$

Now consider the two following shapes. F_θ consists of those positions which are in D, those positions which are on or below the main diagonal and contain dots $\bullet$, and those positions which are strictly above the main diagonal and contain circles $\circ$. F_ξ consists of those positions which are on or below the main diagonal and contain circles $\circ$, and those positions which are strictly above the main diagonal but not in D and contain dots $\bullet$. (The symbol $\odot$ counts both as a dot $\bullet$ and as a circle $\circ$.). We claim that F_θ and F_ξ are Ferrers diagrams of self–conjugate partitions θ and ξ – verification of this is left as an exercise. Finally, if $h \geq 0$ then let $\alpha := \theta$ and $\beta := \xi$, while if $h < 0$ then let $\alpha := \xi$ and $\beta := \theta$. This defines the function $(h, \lambda) \mapsto (\alpha, \beta)$ from $\mathbb{Z} \times \mathcal{E}$ to $\mathcal{SC} \times \mathcal{SC}$.

Continuing the above example, we see that $\theta = 6\ 4\ 4\ 3\ 1\ 1$ and $\xi = 3\ 1\ 1$. Since $h = -2 < 0$ we conclude that $\alpha = 3\ 1\ 1$ and $\beta = 6\ 4\ 4\ 3\ 1\ 1$.

For the converse direction, begin with a pair (α, β) of self–conjugate partitions. Draw an array of $+$ signs in the shape of the Ferrers diagram F_α of α. Superimpose upon this an array of $-$ signs in the shape of the Ferrers diagram F_β of β, in such a way that the southeasternmost symbols of the main diagonals of these shapes coincide with one another. For example, from $\alpha = 3\ 1\ 1$ and $\beta = 6\ 4\ 4\ 3\ 1\ 1$ we draw the picture

$$\begin{array}{cccccc}
- & - & - & - & - & - \\
- & - & - & - & & \\
- & - & \pm & \pm & + & \\
- & - & \pm & & & \\
- & & + & & & \\
- & & & & &
\end{array}$$

(If $\alpha = \varepsilon$ then just draw F_β full of $-$ signs. If $\beta = \varepsilon$ then just draw F_α full of $+$ signs.)

Now define h to be the sum of the main diagonal, with each $+$ sign counting as $+1$ and each $-$ sign counting as -1 (so each $\pm$ sign counts as zero). For the following we consider 0 to have a $+$ sign. Let F_μ be the shape consisting of all positions above the main diagonal which contain a sign opposite that of h, as well as all positions on or below the main diagonal which are also not in the top $|h|$ rows and contain the same sign as h. We claim that F_μ is the Ferrers diagram of some partition μ, and we let $\lambda \in E$ be the partition obtained by multiplying each part of μ by two. This defines the function $(\alpha, \beta) \mapsto (h, \lambda)$ from $\mathcal{SC} \times \mathcal{SC}$ to $\mathbb{Z} \times \mathcal{E}$. (Continuing with example pictured above, we obtain $h = -2$ and $\mu = 5\ 3\ 1\ 1$, so that $\lambda = 10\ 6\ 2\ 2$.)

To complete the proof we must show that these functions are mutually inverse bijections between $\mathbb{Z} \times \mathcal{E}$ and $\mathcal{SC} \times \mathcal{SC}$ with the desired properties. This is a bit tedious, however, and we relegate the verification to Exercise 10.7. $\square$

The Jacobi Triple Product Formula has very interesting geometric applications. The point is that while the LHS is somewhat strange–looking, the RHS is a nicely behaved infinite product. So we can regard the JTPF as giving a good algebraic formula for the power series defined by

$$\vartheta(x, y) := \sum_{h=-\infty}^{\infty} x^{h^2} y^h,$$

which we'll call the *Jacobi theta function*. (This is slightly nonstandard terminology, but not terribly off–base.)

Example 10.3. Fix a positive integer d and a positive real number r such that r^2 is an integer. How many points of $\mathbb{Z}^d$ lie on the sphere of radius r centered at the origin? Call this number $f(r^2)$. We want the number of solutions to the equation $a_1^2 + \cdots + a_d^2 = r^2$ in which each $a_i \in \mathbb{Z}$. That is,

$$\sum_{n=0}^{\infty} f(n) x^n = \sum_{(a_1, \ldots, a_d) \in \mathbb{Z}^d} x^{a_1^2 + \cdots + a_d^2} = \vartheta(x, 1)^d.$$

Thus, the answer to the question is

$$f(r^2) = [x^{r^2}] \vartheta(x, 1)^d = [x^{r^2}] \prod_{i=1}^{\infty} (1 + x^{2i-1})^{2d} (1 - x^{2i})^d.$$

Of course, actually extracting this coefficient remains something of a challenge. At least this example illustrates the geometric interpretation of theta functions.

Ramanujan's First Partition Congruence

The JTPF has other consequences of a number–theoretic nature. We prove the simplest of them, leaving another for an exercise.

Proposition 10.4.

$$\prod_{i=1}^{\infty}(1-x^i)^3 = \sum_{m=0}^{\infty}(-1)^m(2m+1)x^{m(m+1)/2}.$$

Proof. The JTPF implies that

$$\prod_{i=1}^{\infty}(1-x^{2i+1}y)(1-x^{2i-1}y^{-1})(1-x^{2i}) = \frac{1}{1-xy}\sum_{h=-\infty}^{\infty}x^{h^2}(-y)^h.$$

Upon substituting $y = x^{-1}$ in the LHS we obtain the LHS of the desired formula, with x^2 in place of x. On the RHS this substitution is not obviously valid because of the factor $1/(1-xy)$. Expand the product on the RHS and re-index the summations to obtain

$$\frac{1}{1-xy}\sum_{h=-\infty}^{\infty}x^{h^2}(-y)^h = \sum_{k=0}^{\infty}x^k\sum_{h\in\mathbb{Z}:\ h^2\le k}(-1)^h y^{h+k-h^2}.$$

For each $k \in \mathbb{N}$, let $m(k) := \lfloor k^{1/2} \rfloor$. For each $1 \le h \le m(k)$, consider the sum of the terms of the inner sum indexed by h and by $1-h$:

$$(-1)^h y^{h+k-h^2} + (-1)^{1-h}y^{1-h+k-(1-h)^2} = 0.$$

Since these cancel in pairs, the only remaining term is that for which $h = -m(k)$. Thus,

$$\frac{1}{1-xy}\sum_{h=-\infty}^{\infty}x^{h^2}(-y)^h = \sum_{k=0}^{\infty}x^k(-1)^{-m(k)}y^{-m(k)+k-m(k)^2}.$$

Upon substituting $y = x^{-1}$ in this we obtain

$$\sum_{k=0}^{\infty}(-1)^{m(k)}x^{m(k)^2+m(k)}.$$

For each $m \in \mathbb{N}$ there are $2m+1$ values of $k \in \mathbb{N}$ such that $m(k) = m$. Thus, we obtain for the RHS

$$\sum_{m=0}^{\infty}(-1)^m(2m+1)x^{m(m+1)}.$$

Compared with the result of the substitution on the LHS, this proves the formula. □

Theorem 10.5 (Ramanujan's First Partition Congruence)**.** *For each $n \in \mathbb{N}$, let $p(n)$ denote the number of partitions of size n. If $n \equiv 4 \pmod 5$, then $p(n) \equiv 0 \pmod 5$.*

The proof we give of Theorem 10.5 is more number–theoretic than combinatorial in nature. (There is a combinatorial proof, but it is rather intricate.)

The starting point is to consider the power series

$$F_k(x) := \sum_{n=0}^{\infty} f_k(n)x^n := \prod_{j=1}^{\infty} \left(1 - x^j\right)^k$$

for all $k \in \mathbb{Z}$. (Notice that these are all well–defined, by Proposition 7.14.) Our primary interest is in the partition generating function

$$F_{-1}(x) = \sum_{n=0}^{\infty} f_{-1}(n)x^n = \prod_{j=1}^{\infty} \frac{1}{1-x^j} = \Phi_y^n(x),$$

but the series for other values of $k \in \mathbb{Z}$ are useful. For the remainder of this section we will use the notation $f_{-1}(n)$ instead of $p(n)$ for the number of partitions of size n, since we will be using the letter p to stand for a prime number. In particular, by EPNT we have

$$F_1(x) = \sum_{h=-\infty}^{\infty} (-1)^h x^{h(3h-1)/2}$$

and by Proposition 10.4 we have

$$F_3(x) = \sum_{h=0}^{\infty} (-1)^h (2h+1) x^{h(h+1)/2}.$$

The multiplicative formula $F_k(x)F_\ell(x) = F_{k+\ell}(x)$ for all $k, \ell \in \mathbb{Z}$ is obvious but important. Two other less obvious properties of these series are as follows.

First, considering the coefficients of these series modulo a prime number yields an interesting identity.

Lemma 10.6. *For any prime number p, $F_p(x) \equiv F_1(x^p)$ in $\mathbb{Z}_p[[x]]$.*

Proof. Since p is a prime number and $\binom{p}{i} = p!/(i!(p-i)!)$, if $1 \le i \le p-1$ then p divides $\binom{p}{i}$. From this it follows that

$$(1-x^j)^p = \sum_{i=0}^{p} \binom{p}{i} (-x^j)^p \equiv 1 + (-1)^p x^{jp} = 1 - (x^p)^j$$

in $\mathbb{Z}_p[x]$. (Notice that if $p = 2$ then $1 \equiv -1$ in $\mathbb{Z}_2$.) Taking the product over all $j \ge 1$ shows that $F_p(x) \equiv F_1(x^p)$ in $\mathbb{Z}_p[[x]]$, as claimed. □

Lemma 10.7. *Let p be prime and $0 \le r < p$. Assume that $f_{p-1}(pm+r) \equiv 0 \pmod p$ for all $m \in \mathbb{N}$, and that $f_{-1}(r) \equiv 0 \pmod p$. Then $f_{-1}(pm+r) \equiv 0 \pmod p$ for all $m \in \mathbb{N}$.*

Proof. By Lemma 10.6 we have

$$F_{p-1}(x) = F_p(x)F_{-1}(x) \equiv F_1(x^p)F_{-1}(x)$$

in $\mathbb{Z}_p[[x]]$. Now we compare the coefficients of x^n on each side. Using EPNT to expand $F_1(x^p)$ and calculating the product on the RHS, we find that

$$f_{p-1}(n) \equiv \sum_{h=-\infty}^{\infty} (-1)^h f_{-1}\left(n - \frac{h(3h-1)p}{2}\right)$$

(modulo p). Taking all but the $h = 0$ term of the summation to the other side yields

$$f_{-1}(n) \equiv f_{p-1}(n) - \sum_{0\neq h\in\mathbb{Z}} (-1)^h f_{-1}\left(n - \frac{h(3h-1)p}{2}\right)$$

(modulo p). Now substitute $n = pm + r$ in this, with the result that

$$f_{-1}(pm+r) \equiv f_{p-1}(pm+r) - \sum_{0\neq h\in\mathbb{Z}} (-1)^h f_{-1}\left(\frac{(2m - h(3h-1))p}{2} + r\right)$$

(modulo p). The hypotheses of the lemma now support a proof by induction on $m \in \mathbb{N}$ that $f_{-1}(pm + r) \equiv 0 \pmod{p}$ for all $m \in \mathbb{N}$. □

Proof of Theorem 10.5. We verify the hypotheses of Lemma 10.7 with $p = 5$ and $r = 4$. That $f_{-1}(4) = 5 \equiv 0 \pmod 5$ is elementary and was seen at the beginning of Section 9. To verify that $f_4(5m + 4) \equiv 0 \pmod 5$ for all $m \in \mathbb{N}$, consider the power series $F_4(x) = F_1(x)F_3(x)$. We expand the product on the RHS using EPNT and Proposition 10.4. Comparing the coefficients of x^n on both sides we see that

$$f_4(n) = \sum_{h,k} (-1)^{h+k}(2k+1),$$

in which the sum is over all $h \in \mathbb{Z}$ and $k \in \mathbb{N}$ such that

$$n = \frac{h(3h-1)}{2} + \frac{k(k+1)}{2}.$$

Clearing the denominator, we have

$$2n = 3h^2 - h + k^2 + k.$$

If $n \equiv 4 \pmod 5$ then $2n \equiv 3 \pmod 5$. Let's examine the residue (modulo 5) of $3h^2 - h + k^2 + k$, for all possible values of h and k. First, we have

h	0	1	2	3	4
$3h^2 - h$	0	2	0	4	4

and

k	0	1	2	3	4
$k^2 + k$	0	2	1	2	0

so that the residue (modulo 5) of $3h^2 - h + k^2 + k$ is given by the table

$k \setminus h$	0	1	2	3	4
0	0	2	0	4	4
1	2	4	2	1	1
2	1	3	1	0	0
3	2	4	2	1	1
4	0	2	0	4	4

We see that if $n \equiv 4 \pmod 5$ and (h, k) is a solution to $2n = 3h^2 - h + k^2 + k$, then $h \equiv 1 \pmod 5$ and $k \equiv 2 \pmod 5$. The contribution to the value of $f_4(n)$ corresponding to this pair (h, k) is $(-1)^{h+k}(2k + 1) \equiv 0 \pmod 5$. It follows that $f_4(5m + 4) \equiv 0 \pmod 5$ for all $m \in \mathbb{N}$. The hypotheses of Lemma 10.7 have been verified with $p = 5$ and $r = 4$, so that Theorem 10.5 follows directly. □

---

10. Exercises.

---

1. In the proof of Theorem 10.1, show that if $\lambda \in \mathcal{D}$ is in subcase II(a) or II(b) then $\phi(\lambda)$ is also a partition with distinct parts.

---

2. In the proof of Theorem 10.1, show that if $\lambda \in \mathcal{D}$ is in subcase I(a) or I(b) then $\phi(\lambda)$ is in subcase II(a) or II(b).

---

3. Let $p(n)$ be the number of partitions of size n.

(a) Prove that

$$\sum_{h=-\infty}^{\infty} p\left(n - \frac{h(3h-1)}{2}\right) = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{if } n > 0. \end{cases}$$

(b) Write some computer code based on part (a), and compute $p(n)$ for all $0 \leq n \leq 100$. (In the 1880s, McMahon computed up to $p(200)$ with this method – by hand, of course!)

---

4. Show that the following limit of formal power series in the indeterminate q exists, and determine its value:

$$\lim_{n\to\infty} \sum_{k=0}^{n} q^{k(k+1)/2} \left[\begin{array}{c} n \\ k \end{array} \right]_q (-1)^k.$$

---

5. Does the following limit of formal power series in the indeterminate q exist? Explain. If so, then compute the limit power series.

$$\lim_{n\to\infty} \frac{1}{(1-q)^n} \sum_{k=0}^{n} q^{k(k+1)/2} \left[\begin{array}{c} n \\ k \end{array} \right]_q (-1)^k$$

---

6. In the proof of Theorem 10.2, when defining the function $(h, \lambda) \mapsto (\alpha, \beta)$, show that the partitions θ and ξ really are self–conjugate.

---

7. Complete checking the details of the proof of Theorem 10.2.

---

8. By specializing the indeterminates, derive the Euler Pentagonal Number Theorem from the Jacobi Triple Product Formula.

---

9. For each $m \in \mathbb{N}$ and $n \in \mathbb{Z}$, let $f(m, n)$ be the number of solutions $(a, b, c) \in \mathbb{Z}^3$ to the pair of simultaneous equations

$$\left\{ \begin{array}{rcl} a^2 + 2b^2 + 5c^2 & = & m, \\ a - b + 3c & = & n. \end{array} \right.$$

(a) Explain why $f(m, n)$ is finite, for all $m \in \mathbb{N}$ and $n \in \mathbb{Z}$.

(b) Give an infinite product formula for the generating function

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} f(m, n) x^m y^n.$$

---

10. For each $m \in \mathbb{Z}$ and $n \in \mathbb{Z}$, let $f(m, n)$ be the number of points $(a, b, c) \in \mathbb{Z}^3$ with integer coordinates which lie on the intersection of the elliptic hyperboloid

$$a^2 + 2b^2 - c^2 = m$$

and the circular paraboloid

$$a^2 - 3b + c^2 = n.$$

(a) Explain why $f(m, n)$ is finite, for all $m \in \mathbb{Z}$ and $n \in \mathbb{Z}$.

(b) Give an infinite product formula for the generating function

$$\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(m,n) x^m y^n.$$

---

11. Prove the follwing identity, due to Gauss:

$$\prod_{j=1}^{\infty} (1+x^j)(1-x^{2j}) = \sum_{h=0}^{\infty} x^{h(h+1)/2}.$$

---

12. Prove that if $n \equiv 5 \pmod 7$ then $p(n) \equiv 0 \pmod 7$.

---

10. Endnotes.

This has been just the merest glimpse of the theory of partitions. The best book to turn to for further information is

- G.E. Andrews, "The Theory of Partitions", *Encyclopedia of Mathematics* **2**, Addison–Wesley, Reading MA, 1976.

Chapters 7 and 8 of

- H. Gupta, "Selected Topics in Number Theory", Abacus Press, Kent, 1980.

are also quite good, and there are undoubtedly many other fine books on this popular subject. Most of the material of this section has been adapted from these two sources. The main exception is the combinatorial proof of the JTPF, which I found in Prof. Jackson's course notes for C&O 330. (In conversation he attributed the proof to Hirschorn, but I do not know the exact reference.)

I would like to mention without proof a few more important facts about partitions, in order hopefully to whet your interest.

In 1918, Hardy and Ramanujan obtained an asymptotic formula for $p(n)$, the crudest form of which is

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

This led to an **exact formula** for $p(n)$ by Rademacher in 1937. The formula is an infinite series converging to $p(n)$, which rather mysteriously involves certain 24–th roots of unity.

Theorem 10.5 is the tip of a very big iceberg. Around 1919, Ramanujan examined the values of $p(n)$ for $0 \leq n \leq 200$ tabulated by MacMahon and made the following conjecture:

• if $p \in \{5, 7, 11\}$ and $24n - 1 \equiv 0 \pmod{p^c}$ then $p(n) \equiv 0 \pmod{p^c}$.

He proved Theorem 10.5 and Exercise 10.12, as well as:

• if $n \equiv 24 \pmod{25}$ then $p(n) \equiv 0 \pmod{25}$

and

• if $n \equiv 47 \pmod{49}$ then $p(n) \equiv 0 \pmod{49}$.

Ramanujan's Conjecture is not quite true for powers of 7 (it fails for $n = 243$), but it was proved for $p = 5$ and in a slightly modified form for $p = 7$ by Watson in 1938. The case $p = 11$ was settled affirmatively by Atkin in 1967. Recently, Ken Ono and his collaborators have found a vast generalization of these congruences – see

`http://www.aimath.org/news/partition/`

For a combinatorial proof of Theorem 10.5, see

• F. Garvan, D. Kim, and D. Stanton, *Cranks and t-cores*, Invent. Math. **101** (1990), 1–17.

Finally, in relation to Theorem 10.5, Ramanujan also proved this amazing formula:

$$\sum_{m=0}^{\infty} p(5m+4)x^m = 5\prod_{j=1}^{\infty} \frac{(1-x^{5j})^5}{(1-x^j)^6}.$$

As a real challenge, try to find a combinatorial proof of this fact!

11. Introduction to Exponential Generating Functions.

We have seen several applications of generating functions – more specifically, of ordinary generating functions. Exponential generating functions are of another kind and are useful for solving problems to which ordinary generating functions are not applicable.

Ordinary generating functions arise when we have a (finite or countably infinite) set of objects S and a weight function $\omega : S \to \mathbb{N}^r$. Then the ordinary generating function $\Phi_S^\omega(\mathbf{x})$ is defined and we can proceed with calculations. Exponential generating functions arise in a somewhat more complicated situation. The basic idea is that they are used to enumerate "combinatorial structures on finite sets". In this section I will try to give you some idea of what this means without getting bogged down in an axiomatic development. In the process, we will be able to derive enough of the theory to solve some interesting problems. A more formal treatment of the subject is postponed until Chapter 12, which includes proper foundations of the theory as well as discussion of some subtle issues which we can't even speak about until the language is developed. But all that is for later! Right now, let's concentrate on the general ideas, and save the niggling for when we've already got the big picture.

So – a "combinatorial structure on a finite set" – just what does that mean? Graphs are good examples: a graph $G = (V, E)$ consists of a finite set V together with some additional structure, in this case a set E of two–element subsets of V. Endofunctions are also good examples: an *endofunction* is a finite set V together with some additional structure, in this case a function $\phi : V \to V$ from V to itself. Generally, a "combinatorial structure on a finite set" means a finite set together with some additional information defined in terms of that set.

Of course, we are interested in counting things. So we will not consider just one (combinatorial) structure (on a finite set), but an entire family of related structures, called a *class* of structures. For example, we can consider the class $\mathcal{G}$ of all graphs. This consists of all finite sets V and all graph structures $G = (V, E)$ on these finite sets. It's a pretty big thing! (In fact, it is way too big even to be a set – but that's another story.) The point about the class $\mathcal{G}$ is that to each finite set X it associates the finite set $\mathcal{G}_X$ of all graphs which have X as their set of vertices. Notice that if $X \neq Y$ are two different finite sets then $\mathcal{G}_X \cap \mathcal{G}_Y = \varnothing$, since if $G \in \mathcal{G}_X \cap \mathcal{G}_Y$ then $X = V(G) = Y$. Also notice that if $\#X = n$ then $\#\mathcal{G}_X = 2^{n(n-1)/2}$, so that $\#\mathcal{G}_X$ depends only on $\#X$. These properties are the essential ones that we abstract to define a class of structures.

Definition 11.1 (Classes of Structures)**.** A *class* $\mathcal{A}$ of structures associates to every finite set X another finite set $\mathcal{A}_X$, in such a way that the following two conditions are satisfied:
(i) if $X \neq Y$ are distinct finite sets then $\mathcal{A}_X \cap \mathcal{A}_Y = \varnothing$;
(ii) if X and Y are finite sets with $\#X = \#Y$, then $\#\mathcal{A}_X = \#\mathcal{A}_Y$.

(In Section 12 we will enrich this definition and speak about "natural" classes, but this will suffice for now.) The interpretation of the class $\mathcal{A}$ is that $\mathcal{A}_X$ is the finite set of combinatorial structures in the class $\mathcal{A}$ which are defined in terms of the finite set X of "vertices".

Now we see the kind of enumeration problem that exponential generating functions are designed to solve: given a class $\mathcal{A}$ of structures, determine $\#\mathcal{A}_X$ for all finite sets X. That is, determine how many $\mathcal{A}$–structures are defined on each finite set. Of course, by condition (ii) of Definition 11.1, this amounts to determining $\#\mathcal{A}_{N_n}$ for all $n \in \mathbb{N}$. The notation $\mathcal{A}_{N_n}$ is a bit awkward, so let's use $\mathcal{A}_n := \mathcal{A}_{N_n}$ to mean the same thing.

We could put all the numbers $\#\mathcal{A}_n$ for $n \in \mathbb{N}$ together into a generating function:

$$\sum_{n=0}^{\infty} (\#\mathcal{A}_n) x^n$$

but it turns out that this is not the way to do it. The reason why this is no good is that the combinatorial operations we will use to analyze and manipulate classes of structures are not reflected by algebraic operations on these power series. Instead, the proper way to translate the combinatorics into algebra in this situation is as follows.

Definition 11.2 (Exponential Generating Functions)**.** Let $\mathcal{A}$ be a class of structures. The *exponential generating function of* $\mathcal{A}$ is

$$A(x) := \sum_{n=0}^{\infty} (\#\mathcal{A}_n) \frac{x^n}{n!}.$$

(This definition will be embellished a little later to include more indeterminates, but this is the essential form.)

Let's illustrate this with a few cheap examples for which we already know the answer.

Example 11.3. First, consider the class $\mathcal{S}$ of permutations: to each finite set X it associates the finite set $\mathcal{S}_X$ of all bijections $\sigma : X \to X$ from X to X. Condition (i) is easy, and condition (ii) follows from Example 2.2, so that $\mathcal{S}$ satisfies Definition 11.1. Way back in Theorem 2.1 we saw that $\#\mathcal{S}_n = n!$ for all $n \in \mathbb{N}$, so that the

exponential generating function for the class of permutations is

$$S(x) := \sum_{n=0}^{\infty} (\#\mathcal{S}_n) \frac{x^n}{n!} = \sum_{n=0}^{\infty} n! \frac{x^n}{n!} = \frac{1}{1-x}.$$

Example 11.4. Second, consider the class $\mathcal{C}$ of *cyclic permutations*. Recall from Exercise 2.2 that $\#\mathcal{C}_0 = 0$ and $\#\mathcal{C}_n = (n-1)!$ for all $n \geq 1$. Condition (i) is easy, and you should think about how to verify condition (ii). Therefore, $\mathcal{C}$ is a class and its exponential generating function is

$$C(x) := \sum_{n=0}^{\infty} (\#\mathcal{C}_n) \frac{x^n}{n!} = \sum_{n=1}^{\infty} (n-1)! \frac{x^n}{n!} = \log\left(\frac{1}{1-x}\right)$$

by Example 7.9(a).

Example 11.5. Third, consider the class $\mathcal{E}$ of *finite sets*: to each finite set X this associates the set $\mathcal{E}_X := \{X\}$ containing only X. This seems like much ado about nothing, but it will be very useful in a little while. That is, $\mathcal{E}$ is the class of finite sets with no additional structure. Conditions (i) and (ii) are clear in this case. The exponential generating function for the class $\mathcal{E}$ is

$$E(x) := \sum_{n=0}^{\infty} (\#\mathcal{E}_n) \frac{x^n}{n!} = \sum_{n=0}^{\infty} \frac{x^n}{n!} = \exp(x)$$

by Example 7.9(b).

Notice that we have the relation

$$\frac{1}{1-x} = \exp\left(\log\left(\frac{1}{1-x}\right)\right)$$

among these power series. Using the names of the exponential generating functions, that is $S(x) = E(C(x))$. This suggests that some combinatorial relation exists among the classes $\mathcal{S}$, $\mathcal{C}$, and $\mathcal{E}$ – a relation which it would be sensible to denote by something like $\mathcal{S} \equiv \mathcal{E}[\mathcal{C}]$. In fact this is the case – a permutation is equivalent to a finite set of pairwise disjoint cyclic permutations. Our first task is to develop enough of the theory to make sense of an expression like $\mathcal{S} \equiv \mathcal{E}[\mathcal{C}]$.

The usefulness of this theory stems from the ability to identify combinatorial relations among classes – as above – and then to translate these into functional equations for the corresponding exponential generating functions. After that, one can apply algebraic techniques such as the Lagrange Implicit Function Theorem to extract the coefficients of these generating functions, thereby solving the relevant enumeration problems. To realize this program we need to discuss several operations on classes of structures. Then we will have developed enough technique to analyze some nontrivial problems and derive some interesting results. (On a first pass through these constructions it might help to read the first several carefully

and then skim quickly through the rest. After seeing how they are applied in a few problems one can return and reread them all carefully.)

Definition 11.6 (Equivalence of Classes)**.** Two classes $\mathcal{A}$ and $\mathcal{B}$ are *(numerically) equivalent* if $\#\mathcal{A}_X = \#\mathcal{B}_X$ for every finite set X. Of course, this is the case if and only if their exponential generating functions are equal: $A(x) = B(x)$. We use the notation $\mathcal{A} \equiv \mathcal{B}$ to denote that $\mathcal{A}$ and $\mathcal{B}$ are equivalent.

(This concept of equivalence will be superseded in Section 12 by the much more interesting concept of "natural equivalence", but this will do for now.)

Definition 11.7 (Local Finiteness and Sums of Classes)**.** Let $(\mathcal{A}^{(1)}, \mathcal{A}^{(2)}, \ldots)$ be a (finite or infinite) sequence of classes. We say that this sequence is *locally finite* provided that for every finite set X, at most finitely many of the sets $(\mathcal{A}^{(i)}_X : \ i \geq 1)$ are not empty. If this is the case then the set

$$\mathcal{B}_X := \bigcup_{i=1}^{\infty} \Big(\{i\} \times \mathcal{A}^{(i)}_X \Big)$$

is a finite union of finite sets, so that $\mathcal{B}_X$ is a finite set. A typical element of $\mathcal{B}_X$ is an ordered pair (i, α) with $i \geq 1$ and $\alpha \in \mathcal{A}^{(i)}_X$. The presence of the first coordinate ensures that the sets $\{i\} \times \mathcal{A}^{(i)}_X$ are pairwise disjoint. Since the union defining $\mathcal{B}_X$ is a disjoint union, we have

$$\#\mathcal{B}_X = \sum_{i=1}^{\infty} \#\mathcal{A}^{(i)}_X.$$

Conditions (i) and (ii) of Definition 12.1 can now be verified easily for $\mathcal{B}$. In summary, for a locally finite sequence of classes $(\mathcal{A}^{(i)} : \ i \geq 1)$ the class $\mathcal{B}$ is well–defined, and is called the *sum* of the sequence. The exponential generating function of $\mathcal{B}$ is

$$B(x) = \sum_{i=1}^{\infty} A(x).$$

The sum of classes $\mathcal{A}^{(i)}$ for $i \geq 1$ is usually denoted by

$$\bigoplus_{i=1}^{\infty} \mathcal{A}^{(i)}.$$

We could use another set of indices for the classes $\mathcal{A}^{(i)}$ as well, rather than $\{1, 2, \ldots\}$. For example, with only two classes we would just write $\mathcal{A}^{(1)} \oplus \mathcal{A}^{(2)}$, and so on.

As an example of this definition, consider the class $\mathcal{A} \oplus \mathcal{A}$ (in which $\mathcal{A}$ is any class). For a finite set X, a typical element of $(\mathcal{A} \oplus \mathcal{A})_X$ is of the form (i, α) with $i \in \{1, 2\}$ and $\alpha \in \mathcal{A}_X$. So $\mathcal{A} \oplus \mathcal{A}$ is the class of $\mathcal{A}$–structures each of which has been given one of two "colours": $i = 1$ means "red" while $i = 2$ means "blue". The exponential generating function of $\mathcal{A} \oplus \mathcal{A}$ is $2A(x)$ as it should be.

Definition 11.8 (Subclasses and Difference of Classes)**.** Two classes $\mathcal{A}$ and $\mathcal{B}$ are such that $\mathcal{A}$ *is a subclass of* $\mathcal{B}$ provided that for every finite set X, $\mathcal{A}_X \subseteq \mathcal{B}_X$. In this case we can define a class $\mathcal{B} \setminus \mathcal{A}$, called $\mathcal{B}$ *minus* $\mathcal{A}$, by saying that

$$(\mathcal{B} \setminus \mathcal{A})_X := \mathcal{B}_X \setminus \mathcal{A}_X$$

for every finite set X. The exponential generating function of $\mathcal{B} \setminus \mathcal{A}$ is $B(x) - A(x)$.

Now, properly speaking, $\mathcal{A}$ is not a subclass of $\mathcal{A} \oplus \mathcal{B}$. However, for any finite set X there is an injective function

$$\begin{array}{rcl} \mathcal{A}_X & \to & (\mathcal{A} \oplus \mathcal{B})_X \\ \alpha & \mapsto & (1, \alpha) \end{array}$$

so $\mathcal{A}$ is equivalent to a subclass of $\mathcal{A} \oplus \mathcal{B}$. From the point of view of exponential generating functions this is good enough. There is some subtlety to this way of comparing classes of structures, to which we return in Section 12.

Definition 11.9 (Superposition of Classes)**.** Let $\mathcal{A}$ and $\mathcal{B}$ be any two classes. The *superposition* of $\mathcal{A}$ and $\mathcal{B}$ is the class $\mathcal{A}\&\mathcal{B}$ defined as follows: for any finite set X,

$$(\mathcal{A}\&\mathcal{B})_X := \mathcal{A}_X \times \mathcal{B}_X.$$

That is, an $(\mathcal{A}\&\mathcal{B})$–structure on X is an ordered pair (α, β) in which α is an $\mathcal{A}$–structure on X and β is a $\mathcal{B}$–structure on X. Certainly $\#(\mathcal{A}\&\mathcal{B})_X = (\#\mathcal{A}_X)(\#\mathcal{B}_X)$, which implies condition (ii), and condition (i) is also easily verified. There is no elementary formula for the exponential generating function of $\mathcal{A}\&\mathcal{B}$ in terms of $A(x)$ and $B(x)$. Nonetheless, superposition is sometimes useful and will be important in Section 12.

Definition 11.10 (Products of Classes)**.** Let $\mathcal{A}$ and $\mathcal{B}$ be any two classes. The *product* of $\mathcal{A}$ and $\mathcal{B}$ is the class $\mathcal{A} * \mathcal{B}$ defined as follows: for any finite set X,

$$(\mathcal{A} * \mathcal{B})_X := \bigcup_{S \subseteq X} \left(\mathcal{A}_S \times \mathcal{B}_{(X \setminus S)} \right).$$

That is, an $(\mathcal{A} * \mathcal{B})$–structure on the set X is an ordered pair (α, β) in which α is an $\mathcal{A}$–structure on some subset $S \subseteq X$, and β is a $\mathcal{B}$–structure on the complementary subset $X \setminus S$ of X. Notice that condition (i) of the definition of classes $\mathcal{A}$ and $\mathcal{B}$ implies that the union defining $(\mathcal{A} * \mathcal{B})_X$ is a disjoint union. Thus, using condition (ii) as well, we calculate that the cardinality of this set is, for an n–element set X,

$$\#(\mathcal{A} * \mathcal{B})_X = \sum_{k=0}^{n} \binom{n}{k} (\#\mathcal{A}_k)(\#\mathcal{B}_{n-k}).$$

Since this depends only on $\#X = n$, $\mathcal{A} * \mathcal{B}$ satisfies condition (ii) of the definition of a class. Verification of condition (i) for $\mathcal{A} * \mathcal{B}$ is left as a good exercise. From multiplying the above equation by $x^n/n!$ and summing over all $n \in \mathbb{N}$, one easily

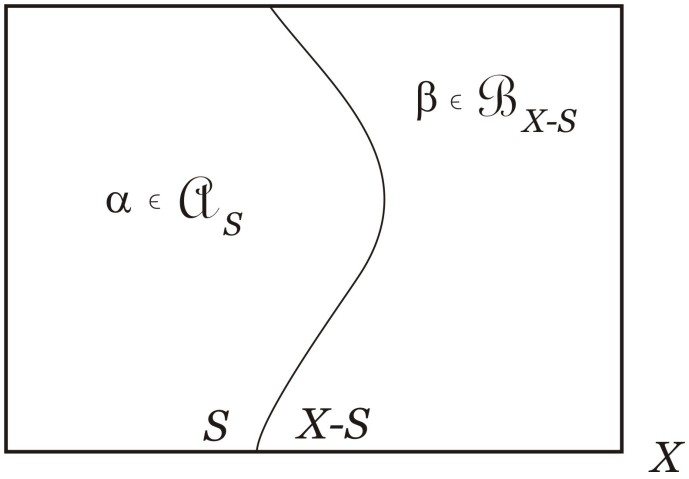


FIGURE 11.1. a structure from the class $\mathcal{A} * \mathcal{B}$.

calculates that the exponential generating function of $\mathcal{A} * \mathcal{B}$ is $A(x)B(x)$. Figure 11.1 illustrates the generic form of a structure from the class $\mathcal{A} * \mathcal{B}$.

Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ be classes of structures. The classes $(\mathcal{A} * \mathcal{B}) * \mathcal{C}$ and $\mathcal{A} * (\mathcal{B} * \mathcal{C})$ are not equal, but they are equivalent. That is, for any finite set X we have $\#((\mathcal{A} * \mathcal{B}) * \mathcal{C})_X = \#(\mathcal{A} * (\mathcal{B} * \mathcal{C}))_X$ because of the following bijection:

$$\begin{aligned} (\mathcal{A} * \mathcal{B}) * \mathcal{C} &\equiv \mathcal{A} * (\mathcal{B} * \mathcal{C}) \\ ((\mathcal{A} * \mathcal{B}) * \mathcal{C})_X &\rightleftharpoons (\mathcal{A} * (\mathcal{B} * \mathcal{C}))_X \\ ((\alpha, \beta), \gamma) &\leftrightarrow (\alpha, (\beta, \gamma)) \end{aligned}$$

This extends similarly to any finite number of factors, so that we can speak about iterated products such as $\mathcal{A} * \mathcal{B} * \mathcal{C} * \cdots * \mathcal{D}$ unambiguously, at least modulo the equivalence relation. (The concept of natural equivalence is used in the next section to strengthen this sense in which we can say that the product $*$ is associative.)

Definition 11.11 (Powers of Classes)**.** Let $\mathcal{A}$ be a class of structures. By iterating the product construction we may define the *powers of* $\mathcal{A}$ to be the products of $\mathcal{A}$ with itself any finite number of times. That is, $\mathcal{A}^1 := \mathcal{A}$ and for all $k \geq 1$, $\mathcal{A}^{k+1} := \mathcal{A}^k * \mathcal{A}$. For all $k \geq 1$, the exponential generating function of $\mathcal{A}^k$ is $A(x)^k$, as can be seen by induction on k. We would like to define $\mathcal{A}^0$ as well – this should have exponential generating function $A(x)^0 = 1$ and be such that $\mathcal{A}^0 * \mathcal{A} \equiv \mathcal{A}$. The class $\mathcal{E}_0$ defined by putting

$$(\mathcal{E}_0)_X := \begin{cases} \{\varnothing\} & \text{if } X = \varnothing, \\ \varnothing & \text{if } X \neq \varnothing. \end{cases}$$

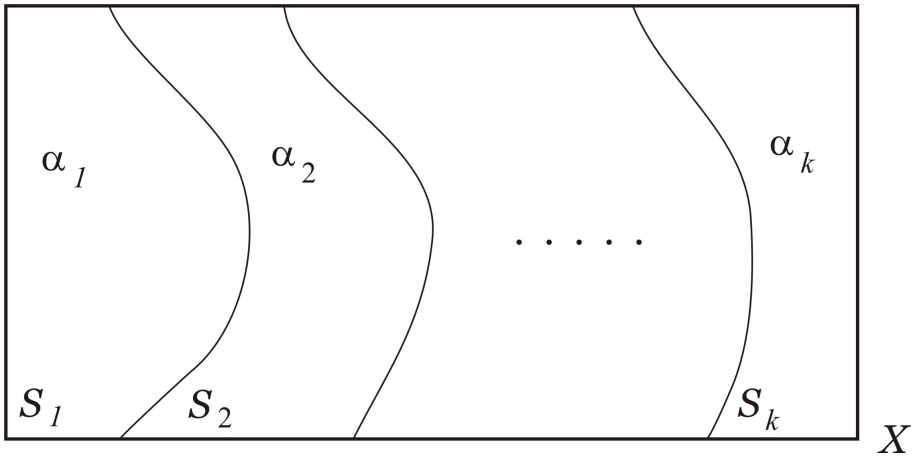


FIGURE 11.2. a structure from the class $\mathcal{A}^k$.

for each finite set X does have exponential generating function $E_0(x) = 1$. This class $\mathcal{E}_0$ is known as the class of *empty*, or *null* structure. By the product formula the classes $\mathcal{E}_0 * \mathcal{A}$ and $\mathcal{A}$ have the same generating function, so they are equivalent. For any natural numbers $j, k \in \mathbb{N}$ there is an equivalence $\mathcal{A}^j * \mathcal{A}^k \equiv \mathcal{A}^{j+k}$. An $\mathcal{A}^k$–structure on a set X is an ordered sequence of k $\mathcal{A}$–structures $(\alpha_1, \ldots, \alpha_k)$ which are pairwise disjoint and cover the set X. Figure 11.2 illustrates the generic form of a structure from the class $\mathcal{A}^k$.

Definition 11.12 (Finite Strings and Connected Classes)**.** Let $\mathcal{A}$ be a class of structures, and consider the sequence $(\mathcal{A}^j : \ j \in \mathbb{N})$ of powers of the class $\mathcal{A}$. We can form the sum of this sequence if (and only if) it is locally finite. This, however, need not be the case. For instance, if $\mathcal{A}_\varnothing \neq \varnothing$ then let $\alpha \in \mathcal{A}_\varnothing$. In this case, for each $k \in \mathbb{N}$, the sequence $(\alpha, \ldots, \alpha)$ of length k is an element of $\mathcal{A}^k_\varnothing$. This shows that if $\mathcal{A}_\varnothing \neq \varnothing$ then the sequence of powers of $\mathcal{A}$ is not locally finite. The converse is also true, and is left as an important exercise: if $\mathcal{A}_\varnothing = \varnothing$ then the sequence of powers of $\mathcal{A}$ is locally finite. We say that the class $\mathcal{A}$ is *connected* when $\mathcal{A}_\varnothing = \varnothing$. (This rather odd–sounding choice of terminology is explained after Example 11.19.) If $\mathcal{A}$ is a connected class then the powers of $\mathcal{A}$ form a locally finite sequence, and we define the *class of finite strings of $\mathcal{A}$–structures* to be the class

$$\mathcal{A}^* := \bigoplus_{k=0}^{\infty} \mathcal{A}^k.$$

By what has gone before, the exponential generating function of $\mathcal{A}^*$ is

$$A^*(x) := \sum_{k=0}^{\infty} A(x)^k = \frac{1}{1 - A(x)}.$$

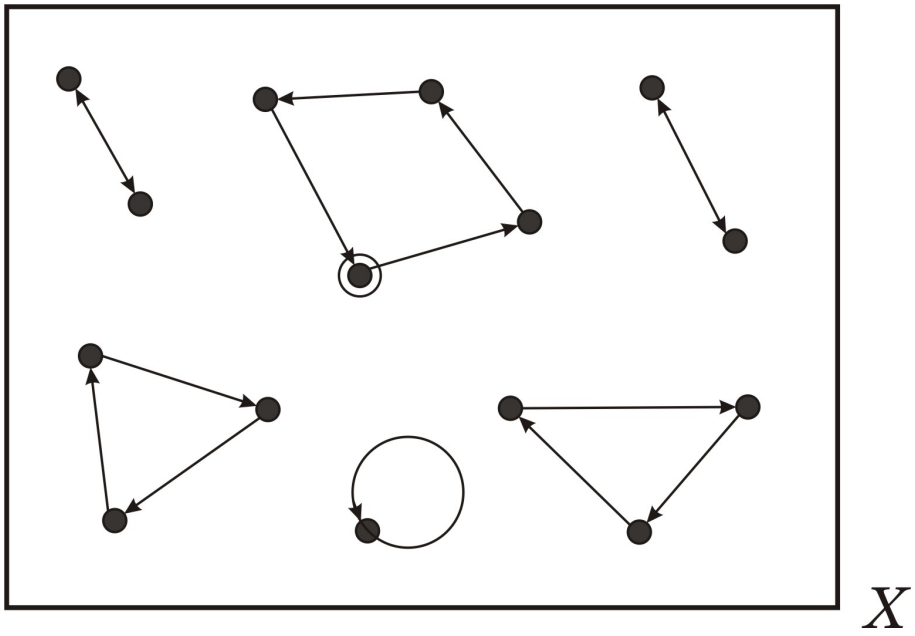


FIGURE 11.3. a rooted permutation.

Notice that the condition that $\mathcal{A}$ is connected corresponds to the condition that $[x^0]A(x) = 0$, which is exactly what is required for the composition of $A(x)$ into $1/(1-x)$ to be well–defined in the ring $\mathbb{Q}[[x]]$ of formal power series.

Definition 11.13 (Rooted Structures)**.** Let $\mathcal{A}$ be a class of structures. The class of *rooted $\mathcal{A}$–structures* is denoted by $\mathcal{A}^\bullet$ and defined as follows: for every finite set X,

$$\mathcal{A}^\bullet_X := \mathcal{A}_X \times X.$$

That is, a rooted $\mathcal{A}$–structure on the set X is an ordered pair (α, v) with $\alpha \in \mathcal{A}_X$ and $v \in X$. Conditions (i) and (ii) are easily verified for $\mathcal{A}^\bullet$. The exponential generating function of $\mathcal{A}^\bullet$ is

$$A^\bullet(x) := \sum_{n=0}^{\infty} (\#\mathcal{A}^\bullet_n)\frac{x^n}{n!} = \sum_{n=0}^{\infty} n(\#\mathcal{A}_n)\frac{x^n}{n!} = x\frac{d}{dx}A(x).$$

Notice that $\mathcal{A}^\bullet$ is always a connected class. We picture a structure in $(\alpha, v) \in \mathcal{A}^\bullet_X$ as an $\mathcal{A}$–structure α on the set X with one "special" or *root* vertex $v \in X$ circled. For example, Figure 11.3 illustrates a structure from the class $\mathcal{S}^\bullet$ – that is, a "rooted permutation".

Example 11.14 (The Classes of k–Sets)**.** For each natural number $k \in \mathbb{N}$, define a class $\mathcal{E}_k$ as follows. For every finite set X,

$$(\mathcal{E}_k)_X := \begin{cases} \{X\} & \text{if } \#X = k, \\ \varnothing & \text{if } \#X \neq k. \end{cases}$$

Notice that in the case $k = 0$ this agrees with the previous definition of $\mathcal{E}_0$. Conditions (i) and (ii) are easily verified, as is the fact that the exponential generating function of $\mathcal{E}_k$ is $E_k(x) = x^k/k!$. This $\mathcal{E}_k$ is called the class of *k–sets*. The intuitive content of this definition is that, given $k \in \mathbb{N}$ and a finite set X, there is exactly one way for X to be a k–element set if $\#X = k$ (it is what it is), and there is no way for X to be a k–element set if $\#X \neq k$. The sequence $(\mathcal{E}_k : \ k \in \mathbb{N})$ is locally finite. Comparing exponential generating functions we see that the sum $\bigoplus_{k=0}^{\infty} \mathcal{E}_k$ is equivalent to the class $\mathcal{E}$ of Example 11.5.

The class $\mathcal{E}_1$ of 1–sets, or *singletons*, is used so frequently that it deserves special attention. Since $\mathcal{E}_1(x) = x$ we also use the notation $\mathcal{X} := \mathcal{E}_1$ for this class.

Definition 11.15 (k–Sets of Structures)**.** Let $\mathcal{A}$ be a class of structures. We define the class $\mathcal{E}_k[\mathcal{A}]$ as follows. For any finite set X, the finite set $(\mathcal{E}_k[\mathcal{A}])_X$ is the image of the set $\mathcal{A}_X^k$ under the following function:

$$\begin{array}{rcl} \mathcal{A}_X^k & \longrightarrow & (\mathcal{E}_k[\mathcal{A}])_X \\ (\alpha_1, \ldots, \alpha_k) & \mapsto & \{\alpha_1, \ldots, \alpha_k\} \end{array}$$

That is, a structure in $(\mathcal{E}_k[\mathcal{A}])_X$ is an unordered k–element set of pairwise disjoint $\mathcal{A}$–structures which cover X. Each element of $(\mathcal{E}_k[\mathcal{A}])_X$ is the image of $k!$ different elements of $\mathcal{A}_X^k$ under this function, from which it follows that $\#(\mathcal{E}_k[\mathcal{A}])_X = (\#\mathcal{A}_X^k)/k!$. Conditions (i) and (ii) are easily verified, as is the fact that the exponential generating function of $\mathcal{E}_k[\mathcal{A}]$ is $A(x)^k/k!$. Figure 11.4 illustrates the generic form of a structure from the class $\mathcal{E}_k[\mathcal{A}]$.

Theorem 11.16 (The Exponential Formula)**.** *Let $\mathcal{A}$ be a class of structures. If $\mathcal{A}$ is connected then $(\mathcal{E}_k[\mathcal{A}] : \ k \in \mathbb{N})$ is a locally finite sequence of classes. The class*

$$\mathcal{E}[\mathcal{A}] := \bigoplus_{k=0}^{\infty} \mathcal{E}_k[\mathcal{A}]$$

has exponential generating function $\exp(A(x))$.

Proof. Since $\mathcal{A}$ is connected, $\mathcal{A}_\varnothing = \varnothing$. That is, every $\mathcal{A}$–structure α uses at least one vertex. Therefore, if X is a finite set and $\{\alpha_1, \ldots, \alpha_k\} \in (\mathcal{E}_k[\mathcal{A}])_X$, then $k \leq \#X$. The reason for this is that the α_i (for $1 \leq i \leq k$) have pairwise disjoint vertex–sets which cover X, and each of these vertex–sets has at least one element. Therefore, if $k > \#X$ then $(\mathcal{E}_k[\mathcal{A}])_X = \varnothing$. This shows that the sequence $(\mathcal{E}_k[\mathcal{A}] : \ k \in \mathbb{N})$ is locally finite. The formula for the exponential generating function of $\mathcal{E}[\mathcal{A}]$ follows from Definitions 11.7 and 11.15. $\square$

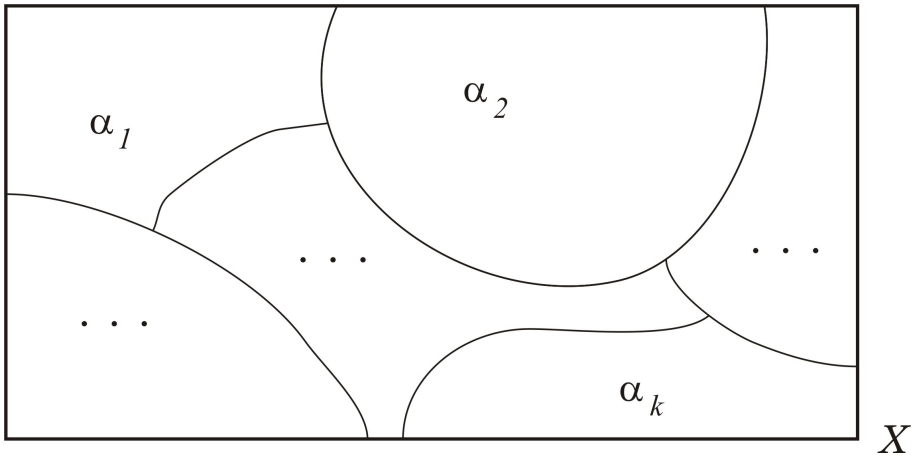


FIGURE 11.4. a structure from the class $\mathcal{E}_k[\mathcal{A}]$.

We have finally developed enough technology to explain the cryptic formula $\mathcal{S} \equiv \mathcal{E}[\mathcal{C}]$ in the paragraph after Example 11.5. We will develop some more theory later in this section and in the next, but we can already do quite a bit with what we have.

Example 11.17. Let $\mathcal{J}$ be the class of (simple, undirected) graphs in which every connected component is a cycle. (These are sometimes called "two–factors".) See Figure 11.5 for an example. Let $\mathcal{H}$ be the class of graphs which are cycles. (Check conditions (i) and (ii) for these classes.) Since each graph in $\mathcal{J}$ can be uniquely decomposed as a disjoint union of an unordered set of cycles, $\mathcal{J} \equiv \mathcal{E}[\mathcal{H}]$. We can obtain the exponential generating function $H(x)$ directly, as follows. Since a (simple) graph cycle must have at least three vertices, we have $\#\mathcal{H}_0 = \#\mathcal{H}_1 = \#\mathcal{H}_2 = 0$. For each $n \geq 3$, a graph cycle may be directed consistently in one of two ways, each of which yields a cyclic permutation. This leads to the equations $\#\mathcal{H}_n = (\#\mathcal{C}_n)/2 = (n-1)!/2$ for all $n \geq 3$. Therefore

$$\begin{aligned} H(x) &= \sum_{n=3}^{\infty} \frac{(n-1)!}{2} \frac{x^n}{n!} = \frac{1}{2} \sum_{n=3}^{\infty} \frac{x^n}{n} \\ &= \frac{1}{2} \log\left(\frac{1}{1-x}\right) - \frac{x}{2} - \frac{x^2}{4}. \end{aligned}$$

By the Exponential Formula we have $J(x) = \exp(H(x))$, and we conclude that

$$J(x) = \frac{\exp(-x/2 - x^2/4)}{\sqrt{1-x}}.$$

Cool! Getting an answer to the enumeration problem $\#\mathcal{J}_n = n![x^n]J(x)$ remains a challenge, however. (It is not really difficult, but the answer is slightly unpleasant.)

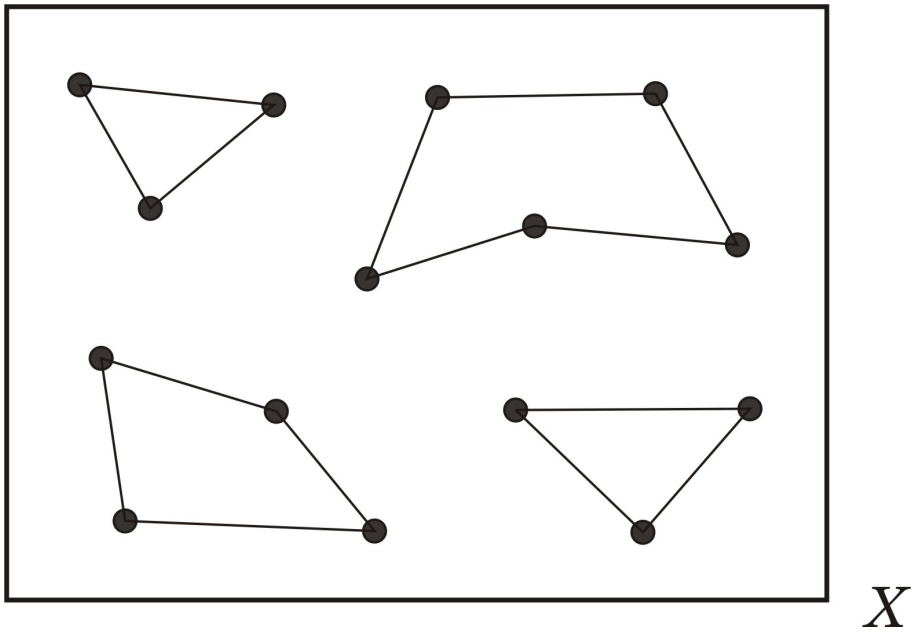


FIGURE 11.5. illustration of Example 11.17.

Example 11.18 (Labelled Trees)**.** Let $\mathcal{T}$ be the class of graphs which are trees. (Check conditions (i) and (ii). Verifying (ii) is not trivial at this point – how would you do it? The ideas of Chapter 12 provide the method.) Using Definition 11.13, $\mathcal{T}^\bullet$ is the class of rooted trees. We may delete the root vertex from a rooted tree and, for each connected component of the remaining graph, root that component at its unique vertex that was adjacent to the deleted vertex. We obtain an unordered set of pairwise disjoint rooted trees, none of which uses the deleted vertex. See Figure 11.6 for an example. Conversely, from a designated vertex v and a set of rooted trees which are pairwise disjoint and do not use v, we may join v to the root of each tree by an edge and root the resulting tree at v. That is, we have bijections for each finite set X:

$$\begin{array}{rcl} \mathcal{T}^\bullet_X & \rightleftharpoons & (\mathcal{X} * \mathcal{E}[\mathcal{T}^\bullet])_X \\ (T, v) & \leftrightarrow & (v, \{(S_1, w_1), \ldots, (S_k, w_k)\}) \end{array}$$

To be more precise, in passing from the LHS to the RHS we let $\{S_1, \ldots, S_k\}$ be the connected components of $T \setminus \{v\}$, and for each $1 \leq i \leq k$ we let w_i be the unique vertex of S_i which is adjacent to v in T. Conversely, in passing from the RHS to the LHS we already have v, and we let

$$V(T) := \{v\} \cup V(S_1) \cup \cdots \cup V(S_k)$$

and

$$E(T) := E(S_1) \cup \cdots \cup E(S_k) \cup \{\{v, w_i\} : \ 1 \leq i \leq k\}.$$

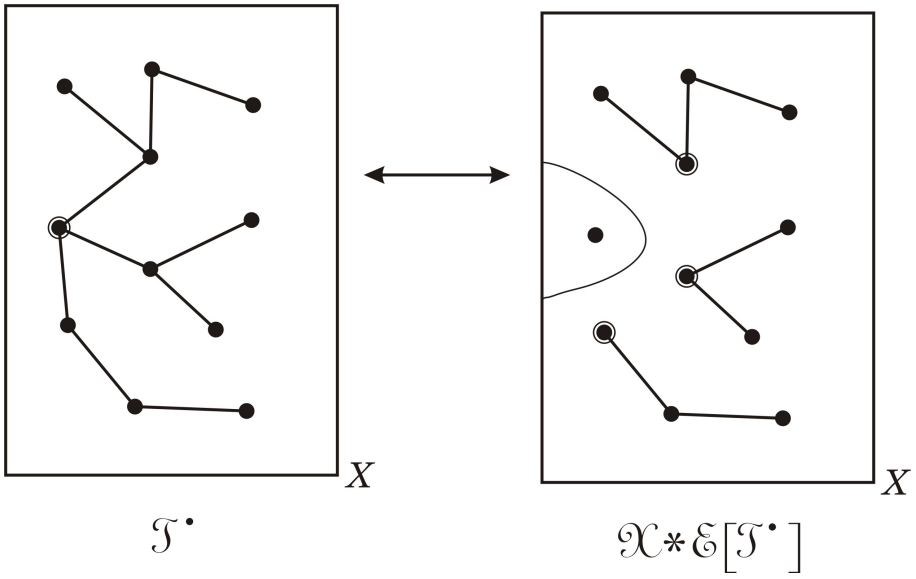


FIGURE 11.6. illustration of Example 11.18.

These bijections establish the following equivalence of classes:

$$\mathfrak{T}^\bullet \equiv \mathfrak{X} * \mathcal{E}[\mathfrak{T}^\bullet].$$

By applying the forgoing theory, we obtain a functional equation for the exponential generating function:

$$T^\bullet(x) = x \exp(T^\bullet(x)).$$

The number of rooted trees on any n–element set is $\#\mathfrak{T}_n^\bullet = n![x^n]T^\bullet(x)$. We can obtain this coefficient by applying the Lagrange Implicit Function Theorem, in this case with $R(x) = T^\bullet(x)$, $F(u) = u$, and $G(u) = \exp(u)$. Thus, we calculate that

$$\begin{aligned} \#\mathfrak{T}_n^\bullet &= n![x^n]T^\bullet(x) = (n-1)![u^{n-1}]\exp(u)^n \\ &= (n-1)![u^{n-1}]\exp(nu) = n^{n-1}. \end{aligned}$$

Since each tree with n vertices can be rooted at any of its n vertices, we conclude that the number of (unrooted) trees on a set of n vertices is $\#\mathfrak{T}_n = n^{n-2}$ for each $n \geq 1$.

Example 11.19 (Stirling numbers of the second kind)**.** Recall from Example 3.5 that for $n \in \mathbb{N}$ and $0 \leq k \leq n$, $S(n,k)$ denotes the number of set partitions of N_n with exactly k parts. We define the class $\mathcal{P}tn$ of set partitions by saying that for every finite set X, $\mathcal{P}tn_X$ is the set of all set partitions of X. (Check that this satisfies conditions (i) and (ii) of Definition 11.1.) To keep track of the number of

parts in a set partition, we will use a **bivariate** generating function

$$Ptn(x,y) := \sum_{n=0}^{\infty} \left(\sum_{\pi \in \mathcal{P}tn_n} y^{\#\pi} \right) \frac{x^n}{n!}.$$

(Notice that this is "exponential in x" and "ordinary in y".) A set partition of X is a finite set of pairwise disjoint nonempty finite sets which partition X. That is, letting

$$\mathcal{E}_{\geq 1} := \bigoplus_{k=1}^{\infty} \mathcal{E}_k$$

denote the class of nonempty sets, we have an equivalence

$$\mathcal{P}tn \equiv \mathcal{E}[\mathcal{E}_{\geq 1}]$$

of classes. Since the exponential generating function of $\mathcal{E}$ is $\exp(x)$, by considering how the indeterminate y enters the formula we find that

$$Ptn(x,y) = \exp(y \exp(x) - y).$$

For any $n, k \in \mathbb{N}$, the fact that

$$S(n,k) = n![x^n y^k] \exp(y \exp(x) - y)$$

can be used to give another proof of Excercise 3.11.

Examples 11.17 and 11.19 illustrate the reason why a class $\mathcal{A}$ for which $\mathcal{A}_\varnothing = \varnothing$ is said to be "connected". In Example 11.17, the connected components of the graphs from the class $\mathcal{J}$ were graphs from the connected class $\mathcal{H}$. In Example 11.19, the parts ("connected components") of the set partitions from the class $\mathcal{P}tn$ were from the connected class $\mathcal{E}_{\geq 1}$. The classes $\mathcal{J}$ and $\mathcal{P}tn$ are not connected. The structures from these classes are built up as disjoint unions of pieces, each of which is from a connected class. In short, structures from a connected class often serve as the connected components out of which structures from other classes are constructed.

Definition 11.20 (Composition of Classes)**.** The exponential formula is the prototypical special case of composition of classes. Let $\mathcal{A}$ and $\mathcal{B}$ be classes of structures, with $\mathcal{A}$ connected (*i.e.* $\mathcal{A}_\varnothing = \varnothing$). We define the *composition of* $\mathcal{A}$ *into* $\mathcal{B}$ to be the class $\mathcal{B}[\mathcal{A}]$ defined as follows. Fix a finite set X. A $\mathcal{B}[\mathcal{A}]$–structure on X consists of a pair (ξ, β) such that
• ξ is an $\mathcal{E}[\mathcal{A}]$–structure on X, and
• β is a $\mathcal{B}$–structure on ξ.
Remember, ξ is a finite set (of $\mathcal{A}$–structures), so this makes sense. In set–theoretic notation, the definition is

$$\mathcal{B}[\mathcal{A}]_X := \bigcup_{\xi \in \mathcal{E}[\mathcal{A}]_X} \left(\{\xi\} \times \mathcal{B}_\xi \right).$$

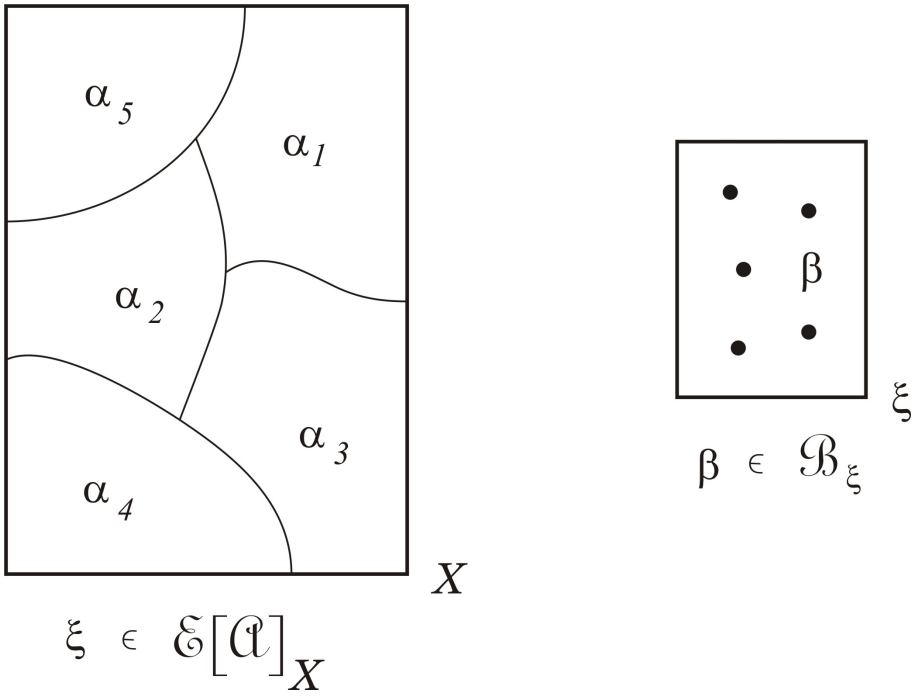


FIGURE 11.7. a structure from the class $\mathcal{B}[\mathcal{A}]$.

Verification of condition (i) for $\mathcal{B}[\mathcal{A}]$ is left as an exercise. If ξ is an element of $(\mathcal{E}_k[\mathcal{A}])_X$ then ξ is a k–element set, so there are $\#\mathcal{B}_k$ different $\mathcal{B}$–structures on ξ. Therefore

$$\#\mathcal{B}[\mathcal{A}]_X = \sum_{k=0}^{\infty} (\#\mathcal{E}_k[\mathcal{A}]_X)(\#\mathcal{B}_k).$$

This verifies condition (ii) for $\mathcal{B}[\mathcal{A}]$.

(The notations $\mathcal{E}_k[\mathcal{A}]$ and $\mathcal{B}[\mathcal{A}]$ are slightly inconsistent when $\mathcal{B} = \mathcal{E}_k$, but the two constructions are "naturally equivalent" in the sense of Section 12.) Figure 11.7 attempts to illustrate the generic form of a structure from the class $\mathcal{B}[\mathcal{A}]$.

Theorem 11.21 (The Compositional Formula)**.** *Let $\mathcal{A}$ and $\mathcal{B}$ be classes with $\mathcal{A}$ connected. Then the exponential generating function of $\mathcal{B}[\mathcal{A}]$ is $B(A(x))$.*

Proof. We calculate that

$$\begin{aligned}\sum_{n=0}^{\infty}(\#\mathcal{B}[\mathcal{A}]_n)\frac{x^n}{n!} &= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}(\#\mathcal{E}_k[\mathcal{A}]_n)(\#\mathcal{B}_k)\frac{x^n}{n!}\\ &= \sum_{k=0}^{\infty}(\#\mathcal{B}_k)\sum_{n=0}^{\infty}(\#\mathcal{E}_k[\mathcal{A}]_n)\frac{x^n}{n!}\\ &= \sum_{k=0}^{\infty}(\#\mathcal{B}_k)\frac{A(x)^k}{k!} = B(A(x)).\end{aligned}$$

□

Example 11.22. Let $\mathcal{Q}$ be the class of graphs which are connected, have exactly one cycle, have maximum degree at most three, and are such that each vertex of degree three is on the unique cycle. We will determine $\#\mathcal{Q}_n$ for all $n \in \mathbb{N}$. Fix a finite set X and consider a graph $\gamma \in \mathcal{Q}_X$. Directing each cut–edge of γ towards the unique cycle C of γ shows how we may regard γ as a collection of nonempty directed paths which have been fit together "inside" the graph cycle C. See Figure 11.8 for an example. Conversely, given an unordered set of pairwise disjoint nonempty directed paths partitioning X and a graph cycle on this set of paths, a graph in $\mathcal{Q}_X$ is constructed by connecting the terminal vertices of these paths according to the graph cycle. A nonempty directed path is equivalent to a structure from the class $\mathcal{X} * \mathcal{X}^*$, so that this analysis gives an equivalence of classes

$$\mathcal{Q} \equiv \mathcal{H}[\mathcal{X} * \mathcal{X}^*]$$

with the class $\mathcal{H}$ as in Example 11.17. Therefore

$$\begin{aligned}Q(x) &= H\left(\frac{x}{1-x}\right)\\ &= \frac{1}{2}\left[\log\left(\frac{1}{1-x/(1-x)}\right) - \frac{x}{1-x} - \frac{x^2}{2(1-x)^2}\right]\\ &= \frac{1}{2}\left[\log\left(\frac{1}{1-2x}\right) - \log\left(\frac{1}{1-x}\right) - \frac{x}{1-x} - \frac{x^2}{2(1-x)^2}\right]\\ &= \frac{1}{2}\left[\sum_{n=1}^{\infty}\frac{2^n x^n}{n} - \sum_{n=1}^{\infty}\frac{x^n}{n} - \sum_{j=0}^{\infty}x^{j+1} - \frac{1}{2}\sum_{j=0}^{\infty}\binom{j+1}{j}x^{j+2}\right]\\ &= \frac{1}{2}\left[(2-1-1)x + \sum_{n=2}^{\infty}\left(\frac{2^n-1}{n} - 1 - \frac{n-1}{2}\right)x^n\right]\\ &= \sum_{n=2}^{\infty}\left(\frac{(2^{n+1}-n^2-n-2)(n-1)!}{2}\right)\frac{x^n}{n!}.\end{aligned}$$

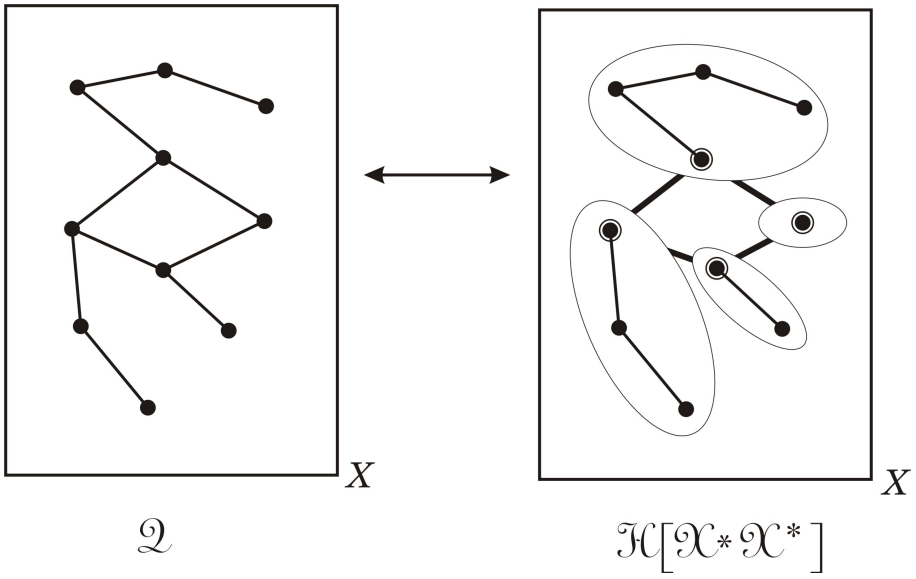


FIGURE 11.8. illustration for Example 11.22.

Hence, the number of graphs in $\mathcal{Q}$ with vertex set $\{1, 2, \ldots, n\}$ is $\#\mathcal{Q}_0 = \#\mathcal{Q}_1 = 0$, and

$$\#\mathcal{Q}_n = (n-1)! \left(2^n - 1 - \binom{n+1}{2} \right),$$

for all $n \geq 2$. In particular, $\#\mathcal{Q}_2 = 0$ as well, as it should be.

Example 11.23 (Endofunctions and Doubly–Rooted Trees)**.** In this example we obtain the formula $\#\mathcal{T}_n = n^{n-2}$ (for all $n \geq 1$) of Example 11.18 by a more direct combinatorial argument.

Let $\mathfrak{F}$ denote the class of endofunctions. It is clear that $\#\mathfrak{F}_n = n^n$ for all $n \in \mathbb{N}$. Let $\mathcal{R} = \mathcal{T}^\bullet$ be the class of rooted trees, and fix a finite set X. Directing each edge of a rooted tree $\gamma \in \mathcal{R}_X$ towards the root of γ, and putting a directed loop at the root of γ, determines the functional directed graph of an endofunction on X. This determines an injective function $\mathcal{R}_X \to \mathfrak{F}_X$ for every finite set X. Thus, $\mathcal{R}$ may be regarded as a subclass of $\mathfrak{F}$.

The functional directed graph of an arbitrary endofunction is the disjoint union of an unordered set of (weakly) connected components. Letting $\mathfrak{C}$ denote the class of endofunctions for which the functional directed graph is connected, we have $\mathfrak{F} \equiv \mathcal{E}[\mathfrak{C}]$.

The functional directed graph of an endofunction in $\mathfrak{C}$ may be uniquely decomposed as the disjoint union of an unordered set of endofunctions in $\mathcal{R}$, with the loops at the roots replaced by a cyclic permuation of the roots. This gives an equivalence

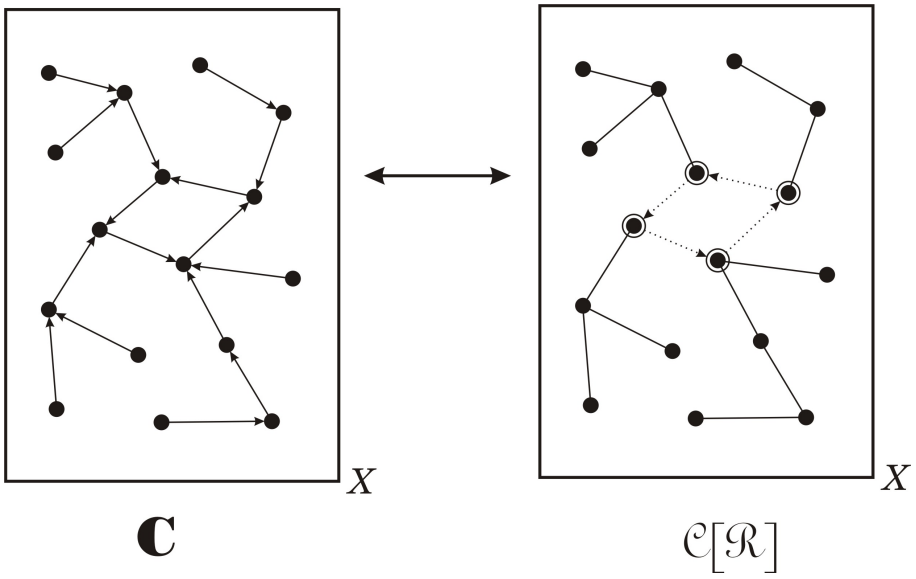


FIGURE 11.9. the equivalence $\mathfrak{C} \equiv \mathcal{C}[\mathcal{R}]$.

$\mathfrak{C} \equiv \mathcal{C}[\mathcal{R}]$, in which $\mathcal{C}$ is the class of cyclic permutations. See Figure 11.9 for an example.

Thus we have $\mathfrak{F} \equiv \mathcal{E}[\mathcal{C}[\mathcal{R}]]$, and so

$$F(x) = \exp\left(\log\left(\frac{1}{1-R(x)}\right)\right) = \frac{1}{1-R(x)}.$$

Since $\#\mathfrak{F}_n = n^n$ for all $n \in \mathbb{N}$, we conclude that

$$\frac{1}{1-R(x)} = F(x) = \sum_{n=0}^{\infty} n^n \frac{x^n}{n!}.$$

Now consider the class $\mathcal{R}^\bullet = (\mathcal{T}^\bullet)^\bullet$ of *doubly–rooted* trees. The rootings are done sequentially, so the root vertices are ordered: one first, the other second. We denote the first root with a circle and the second root with a triangle. Also, the two root vertices may coincide. If γ is a doubly–rooted tree on the set X then γ contains a unique directed path ℓ from the second root to the first root. Every edge of γ not on ℓ may be directed towards ℓ; also put a directed loop at each vertex on ℓ. This decomposes γ uniquely as the disjoint union of a nonempty unordered set of rooted trees in $\mathcal{R}$, and a total order on the set of their roots. See Figure 11.10 for an example.

This gives a natural equivalence

$$\mathcal{R}^\bullet \equiv (\mathcal{X} * \mathcal{X}^*)[\mathcal{R}] \equiv \bigoplus_{k=1}^{\infty} \mathcal{R}^k,$$

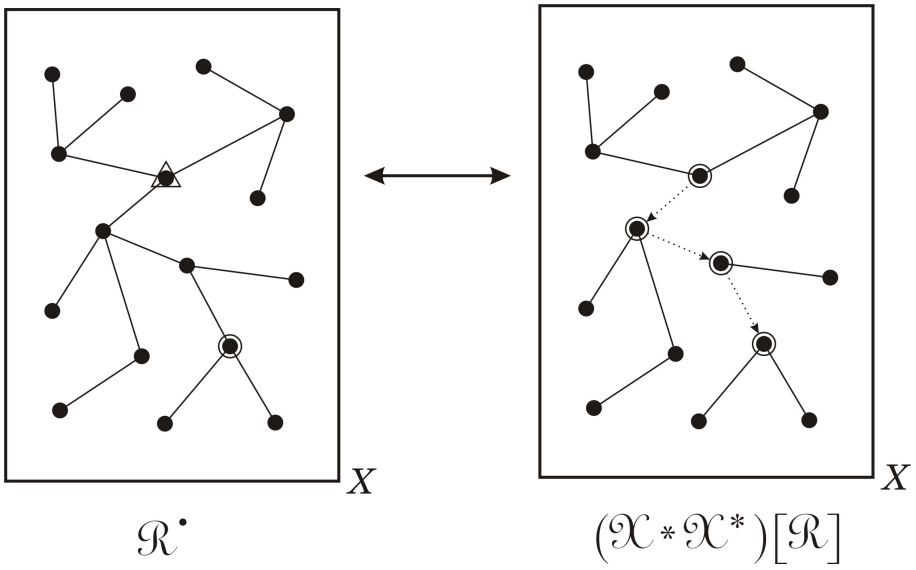


FIGURE 11.10. the equivalence $\mathcal{R}^\bullet \equiv (\mathcal{X} * \mathcal{X}^*)[\mathcal{R}]$.

and hence

$$T^{\bullet\bullet}(x) = R^\bullet(x) = \frac{R(x)}{1 - R(x)} = F(x) - 1.$$

Comparing this with the above formulas, we get

$$\left(x\frac{d}{dx}\right)^2 T(x) = \sum_{n=1}^{\infty} n^2 (\#\mathcal{T}_n)\frac{x^n}{n!} = \sum_{n=1}^{\infty} n^n \frac{x^n}{n!}.$$

Therefore, for all $n \geq 1$ we have $\#\mathcal{T}_n = n^{n-2}$.

Example 11.24 (Expected Number of Leaves in a Tree)**.** Let $\mathcal{T}$ be the class of trees, let X be a finite set, and let $T \in \mathcal{T}_X$ be a tree with vertex–set X. Denote by $\ell(T)$ the number of leaves (vertices of degree at most one) in T. Assume that $\#X = n$. Among all the n^{n-2} trees in $\mathcal{T}_X$, what is the average value of $\ell(T)$? That is, how many leaves should we expect to see on a tree with n vertices? First of all, since the designation of a root vertex $v \in X$ does not change the number of leaves of T, we may just as well compute the average of $\ell(T)$ as (T, v) varies over all n^{n-1} rooted trees in $\mathcal{T}_n^\bullet$. This allows us to use the recursive structure $\mathcal{R} \equiv \mathcal{X} * \mathcal{E}[\mathcal{R}]$ of the class $\mathcal{R} = \mathcal{T}^\bullet$ of rooted trees.

Consider the bivariate generating function

$$R(x, y) := \sum_{n=0}^{\infty} \left(\sum_{(T,v)\in\mathcal{R}_n} y^{\ell(T)} \right) \frac{x^n}{n!}.$$

For any $n \geq 1$,

$$L_n := n![x^n]\ \frac{\partial}{\partial y} R(x,y)\bigg|_{y=1} = \sum_{(T,v)\in\mathcal{R}_n} \ell(T)$$

is the total number of leaves among all the rooted trees in $\mathcal{R}_n$. Thus, the answer to our question is L_n/n^{n-1}.

Notice that for $(T,v) \in \mathcal{R}_n$, the leaves of T are:
• those vertices which have no children in (T,v), that is, the terminal vertices of (T,v), and
• if the root vertex has at most one child, then the root is also a leaf.
This special case for the root vertex is kind of annoying, so let's ignore it for the moment. For a rooted tree $(T,v) \in \mathcal{R}_n$, let $\tau(T,v)$ be the number of vertices which have no children in T, and let

$$B(x,y) := \sum_{n=0}^{\infty} \left(\sum_{(T,v)\in\mathcal{R}_n} y^{\tau(T,v)} \right) \frac{x^n}{n!}.$$

To obtain an expression for $B(x,y)$, we use the recursive structure $\mathcal{R} \equiv \mathcal{X} * \mathcal{E}[\mathcal{R}]$. In this equivalence, if (T,v) corresponds to $(v, \{(S_1,w_1),\ldots,(S_k,w_k)\})$, then

$$\tau(T,v) = \begin{cases} 1 & \text{if } k=0, \\ \tau(S_1,w_1)+\cdots+\tau(S_k,w_k) & \text{if } k \geq 1. \end{cases}$$

Keeping track of $\tau(T,v)$ through the equivalence $\mathcal{R} \equiv \mathcal{X} * \mathcal{E}[\mathcal{R}]$ yields the functional equation

$$\begin{aligned} B(x,y) &= x\left[y + \sum_{k=1}^{\infty} \frac{B(x,y)^k}{k!} \right] \\ &= x(y + \exp(B(x,y)) - 1) \end{aligned}$$

for the generating function $B(x,y)$.

Now we can handle the special case of the root vertex, by noticing that if (T,v) corresponds to $(v, \{(S_1,w_1),\ldots,(S_k,w_k)\})$ as above, then

$$\ell(T) = \begin{cases} 1 & \text{if } k=0, \\ 1+\tau(S_1,v_1) & \text{if } k=1, \\ \tau(S_1,v_1)+\cdots+\tau(S_k,v_k) & \text{if } k \geq 2. \end{cases}$$

From this, we see that

$$\begin{aligned} R(x,y) &= x\left[y + yB(x,y) + \sum_{k=2}^{\infty} \frac{B(x,y)}{k!} \right] \\ &= x(y + yB(x,y) + \exp(B(x,y)) - 1 - B(x,y)). \end{aligned}$$

The functional equation for $B(x,y)$ and the equation for $R(x,y)$ in terms of $B(x,y)$ are of the form to which the Lagrange Implicit Function Theorem applies, in this case with $\mathbb{K} = \mathbb{Q}(y)$, $G(u) = \exp(u) + y - 1$, and $F(u) = \exp(u) + (y-1)(1+u)$. Notice that $B(x,y) = xG(B(x,y))$ and $R(x,y) = xF(B(x,y))$, and that $F'(u) = (d/du)F(u) = \exp(u) + y - 1 = G(u)$. Thus we calculate that

$$\begin{aligned}
L_n &= n![x^n]\,\frac{\partial}{\partial y}R(x,y)\bigg|_{y=1} = n!\,\frac{\partial}{\partial y}[x^n]xF(B(x,y))\bigg|_{y=1} \\
&= n!\,\frac{\partial}{\partial y}[x^{n-1}]F(B(x,y))\bigg|_{y=1} = n(n-2)!\,\frac{\partial}{\partial y}[u^{n-2}]F'(u)G(u)^{n-1}\bigg|_{y=1} \\
&= n(n-2)![u^{n-2}]\,\frac{\partial}{\partial y}(\exp(u)+y-1)^n\bigg|_{y=1} \\
&= n(n-2)![u^{n-2}]n\exp((n-1)u) \\
&= \frac{n^2(n-2)!(n-1)^{n-2}}{(n-2)!} = n^2(n-1)^{n-2}.
\end{aligned}$$

Hence, finally, we see that the average number of leaves among all trees on the set $\{1,2,\ldots,n\}$ is

$$\frac{n^2(n-1)^{n-2}}{n^{n-1}} = (n-1)\left(1-\frac{1}{n}\right)^{n-3} \sim \frac{n}{e},$$

asymptotically as $n \to \infty$. Informally speaking, in a large random tree one expects that something close to the fraction $1/e \approx 0.36787944\ldots$ of the vertices are leaves.

---

11. Exercises.

---

1. Let $(\mathcal{A}^{(1)}, \mathcal{A}^{(2)}, \ldots)$ be a locally finite sequence of classes. Show that $\mathcal{B} := \bigoplus_{i=1}^{\infty} \mathcal{A}^{(i)}$ satisfies condition (i) of Definition 11.1.

---

2. Let $\mathcal{A}$ and $\mathcal{B}$ be classes of structures. Show that $\mathcal{A} * \mathcal{B}$ satisfies condition (i) of Definition 11.1.

---

3. Show that if $\mathcal{A}$ is a connected class (*i.e.* $\mathcal{A}_\varnothing = \varnothing$) then the powers of $\mathcal{A}$ form a locally finite sequence.

---

4. Let $\mathcal{A}$ and $\mathcal{B}$ be classes, with $\mathcal{A}$ connected. Show that $\mathcal{B}[\mathcal{A}]$ satisfies condition (i) of Definition 11.1.

---

5. Recall that a derangement is a permutation with no fixed points. Let $\mathcal{D}$ be the class of derangments.

(a) Derive the exponential generating function

$$D(x) = \frac{\exp(-x)}{1-x}.$$

(b) Use part (a) to give another solution for Example 2.6.

---

6. For a permutation $\sigma \in \mathcal{S}_n$, let $c(\sigma)$ be the number of cycles of σ. What is the average value of $c(\sigma)$ among all $n!$ permutations in $\mathcal{S}_n$?

---

7. Use the formula of Example 11.19 to give another solution for Exercise 3.11.

---

8. Fix a positive integer k. For a permutation σ, let $c(\sigma, k)$ be the number of cycles in σ of length exactly k.

(a) Obtain an algebraic formula for the bivariate exponential generating function

$$S(x,y) = \sum_{n=0}^{\infty} \left(\sum_{\sigma \in \mathcal{S}_n} y^{c(\sigma,k)} \right) \frac{x^n}{n!}.$$

(b) Show that the average number of cycles of length k among all $n!$ permutations in $\mathcal{S}_n$ is

$$\begin{cases} 1/k & \text{if } k \leq n, \\ 0 & \text{if} k > n. \end{cases}$$

---

9. Let $\mathcal{Y}$ be the class of (nonrooted) labelled trees in which each vertex has degree either 1 or 3. Show that for all $k \geq 0$, $\#\mathcal{Y}_{2k+1} = 0$ and

$$\#\mathcal{Y}_{2k+2} = \frac{(2k)!}{2^k} \binom{2k+2}{k}.$$

10(a) Let $\mathcal{A}$ be the class of rooted labelled trees such that each vertex has at most two children. Show that the exponential generating function for $\mathcal{A}$ is

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

(b) Let $\mathcal{B}$ be the class of rooted labelled trees which are in $\mathcal{A}$ and are such that the root vertex has at most one child. Show that the exponential generating function for $\mathcal{B}$ is

$$B(x) = 1 - \sqrt{1 - 2x - x^2}.$$

(c) Let $\mathcal{U}$ be the class of endofunctions $f : X \to X$ such that for every $v \in X$, $\#f^{-1}(v) \leq 2$. Show that the exponential generating function for $\mathcal{U}$ is

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

(d) Use part (c) to obtain a formula for the number of endofunctions $f : N_n \to N_n$ in the class $\mathcal{U}$, for every natural number $n \in \mathbb{N}$.

11(a) Derive the following formula

$$\sum_{k=0}^{\infty} \frac{x^{2k+1}}{2k+1} = \frac{1}{2} \log\left(\frac{1+x}{1-x}\right).$$

(b) Let $\mathcal{Q}$ be the class of endofunctions in which each cycle has odd length. Explain the following formulas which implicitly determine the exponential generating function $Q(x)$ of $\mathcal{Q}$:

$$\begin{cases} Q(x) &= \sqrt{\frac{1+R(x)}{1-R(x)}}, \\ R(x) &= x \, \exp(R(x)). \end{cases}$$

---

12. A *triangle–tree* is a connected graph in which every edge is in exactly one cycle, and this cycle has length three (see Figure 11.11). Show that the number of triangle–trees with vertex–set N_n is 0 when n is even, and is

$$\frac{(2k)!(2k+1)^{k-1}}{k!2^k}$$

when $n = 2k + 1$ is odd.
(Hint: Describe the recursive structure of the class of rooted triangle–trees.)

---

13. Let $\mathfrak{F}$ be the class of endofunctions, and for $\phi \in \mathfrak{F}_X$, let $p(\phi)$ denote the number of fixed points of ϕ: that is, the number of $v \in X$ such that $\phi(v) = v$.

(a) Obtain functional equations which determine the exponential generating function

$$F(x, y) = \sum_{n=0}^{\infty} \left(\sum_{\phi \in \mathfrak{F}_n} y^{p(\phi)} \right) \frac{x^n}{n!}.$$

(b) Among all endofunctions $\phi \in \mathfrak{F}_n$, what is the average value of $p(\phi)$?

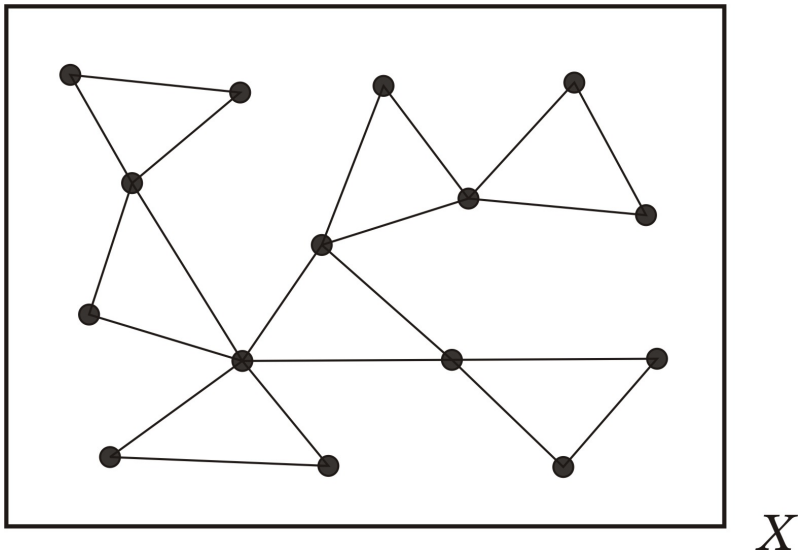


FIGURE 11.11. a triangle–tree.

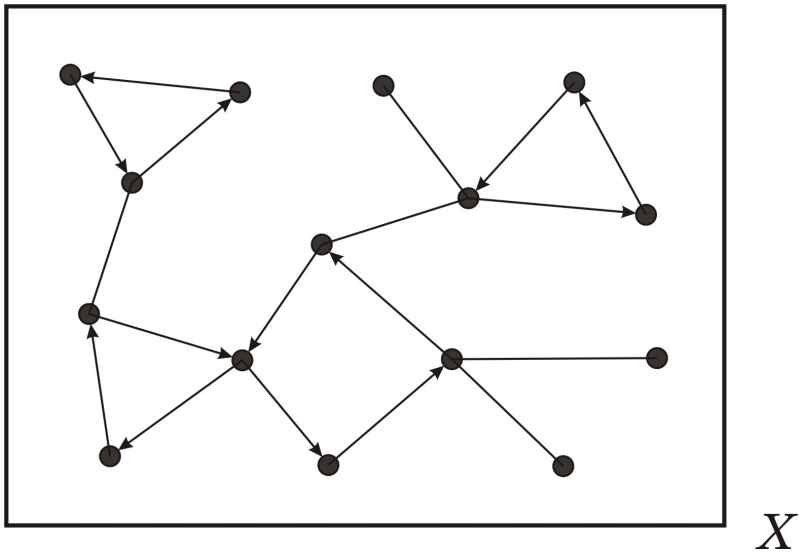


FIGURE 11.12. an oriented cactus.

---

14. A *cactus* is a connected graph such that each edge is in at most one cycle. Equivalently, it is a connected graph in which each block (2–connected component) is either an edge or a cycle. An *oriented cactus* is a cactus in which each cycle has been directed in one of its two strongly connected orientations. (See Figure 11.12.)

(a) Show that for all $n \geq 1$, the number of oriented cacti on the set $\{1, \ldots, n\}$ is

$$(n-1)! \sum_{k=0}^{n-1} \frac{n^{k-1}}{k!} \binom{n-2}{n-1-k}.$$

(b) Derive a functional equation that implicitly determines the exponential generating function for the class of rooted non–oriented cacti.

(c) For each $n \in \mathbb{N}$, what is the number of (non–rooted, non–oriented) cacti on the set N_n?

15(a) Show that the number of rooted trees on the set N_n which have exactly k terminal vertices is

$$(n-k)! \binom{n}{k} S(n-1, n-k).$$

(b)* Find a combinatorial (bijective) proof of this result.

16.* Revisit Exercise 8 of Chapter 8. Find a combinatorial proof of Exercise 8.8(b).

17.* For $n, k \in \mathbb{N}$, let $q(n,k)$ be the number of connected graphs with k edges and vertex–set $\{1, 2, \ldots, n\}$; also let $Q_n(t) := \sum_{k=0}^{n(n-1)/2} q(n,k) t^k$.

(a) Explain an efficient algorithm for computing $Q_n(t)$.
(Hint: the generating function $\sum_{n=0}^{\infty} Q_n(t) x^n/n!$ is related to an easily determined series.)

(b) If you know MAPLE or another computer algebra application, write some code and crank out $Q_8(t)$. (Or do it by pencil and paper! ;-)

---

11. Endnotes.

Here are three books that treat exponential generating functions in detail:

• I.P. Goulden and D.M. Jackson, "Combinatorial Enumeration," John Wiley & Sons, New York, 1983.

• R.P. Stanley, "Enumerative Combinatorics, vol. II," Cambridge U.P., Cambridge, 1999.

- H.S. Wilf, "Generatingfunctionology," Academic Press, New York, 1994.

That the number of labelled trees with n vertices is n^{n-2} is attributed to Cayley, in 1889. His solution is a bit sketchy, however. A bijective proof was given in 1918 by Prüfer. See page 51 of

- N.L. Biggs, E.K. Lloyd, and R.J. Wilson, "Graph theory: 1736–1936," Clarendon Press, Oxford, 1976.

There are other kinds of generating functions besides the ordinary generating functions and exponential generating functions we have discussed. For example, in number theory it is frequently useful to encode a sequence $a_1, a_2, \ldots$ of integers by means of its *Lambert series*:

$$\sum_{k=1}^{\infty} a_k \frac{x^k}{1-x^k}.$$

It is not difficult to verify that for every $n \geq 1$, the coefficient of x^n in this series is

$$\sum_{d|n} a_d,$$

the sum being over all positive divisors of n. For an introduction to several of the various forms of generating functions, see

- P. Doubilet, G.-C. Rota, and R.P. Stanley, *On the foundations of combinatorial theory. VI. The idea of generating function*, in "Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability Vol. II: Probability theory," Univ. California Press, Berkeley, 1972.

- R.P. Stanley, *Generating functions*, in "Studies in Combinatorics" (G.-C. Rota, ed.), Math. Assoc. America, Washington, 1978.

12. Foundations of Exponential Generating Functions.

In Section 11 we introduced the theory of exponential generating functions and applied it to solve several enumeration problems. There were some shortcomings of that discussion, though, which we address in this section. The main point of dissatisfaction is that the concept of equivalence used in Section 11, while adequate for numerical purposes, is much too coarse to discern the more interesting properties of classes of structures. The underlying problem is that Definition 11.1 does not really provide an adequate foundation for the theory even though, as we saw, much of it can be developed satisfactorily at that level of detail. Thus, we begin with a more sophisticated definition of a "natural" class of structures.

Definition 12.1 (Natural Classes of Structures)**.** A *natural class of structures* $\mathcal{A}$ associates to each finite set X another finite set $\mathcal{A}_X$, so that the following conditions hold.
[***V**] There is an algorithm $\mathsf{V}_{\mathcal{A}}$ which takes as input an $\mathcal{A}$–structure $\alpha \in \mathcal{A}_X$ and returns as output the set X on which α is defined.
[***C1**] For finite sets X and Y and a bijection $f : X \to Y$, there is an *induced bijection* $f_* : \mathcal{A}_X \to \mathcal{A}_Y$.
[***C2**] For the identity bijection $\mathrm{id}_X : X \to X$, the induced bijection is the identity bijection

$$(\mathrm{id}_X)_* = \mathrm{id}_{\mathcal{A}_X} : \mathcal{A}_X \to \mathcal{A}_X.$$

[***C3**] For bijections $f : X \to Y$ and $g : Y \to Z$, the induced bijections are compatible with composition of functions:

$$(g \circ f)_* = g_* \circ f_* : \mathcal{A}_X \to \mathcal{A}_Z.$$

If we want to emphasize the class $\mathcal{A}$ which is being used to produce the induced bijection f_* then we will write $f_{\mathcal{A}}$ instead.

This looks like a lot to require, but we will see that **everything** in Section 11 satisfies these much more demanding conditions.

First of all, I want to explain the intuitive content of the axioms. The axiom [*V] says that if one is handed an arbitrary structure α from the class $\mathcal{A}$ then there is a computation $\mathsf{V}_{\mathcal{A}}$ which can be performed with this input in order to determine the set (of vertices) on which α is defined. The axiom [*C1] says that given a bijection $f : X \to Y$, if we take the set of $\mathcal{A}$–structures defined on X and "relabel the vertices according to f" then we obtain a bijection f_* from $\mathcal{A}_X$ to $\mathcal{A}_Y$. The key idea is that f_* is the result of changing the names of the vertices only, leaving the additional

structure (from the class $\mathcal{A}$) unchanged. Axiom [*C2] is then an obvious requirement – if we do not change the names of the vertices at all then each $\mathcal{A}$–structure on X must be mapped to itself. Axiom [*C3] is likewise a necessary requirement for this interpretation of the induced bijections, which I leave to you to ponder.

Example 12.2 (Graphs)**.** As an example, consider the class $\mathcal{G}$ of graphs. Let's verify the axioms to show that $\mathcal{G}$ is a natural class. Given a graph $\gamma = (V, E)$, we can determine its vertex–set $\mathsf{V}_{\mathcal{G}}(\gamma) := V$, so that axiom [*V] holds. For the remaining axioms, let $f : X \to Y$ be a bijection, and consider $\gamma = (X, E) \in \mathcal{G}_X$. We define

$$f_*(\gamma) := (Y, \{\{f(v), f(w)\} : \{v, w\} \in E\}).$$

This defines a function $f_* : \mathcal{G}_X \to \mathcal{G}_Y$. It is now a routine if somewhat tedious matter to check that this definition satisfies axioms [*C1], [*C2], and [*C3].

Example 12.3 (Endofunctions)**.** As another example, consider the class $\mathfrak{F}$ of endofunctions. Let's verify the axioms to show that $\mathfrak{F}$ is a natural class. Given an endofunction ϕ, we can determine its vertex–set $\mathsf{V}_{\mathfrak{F}}(\phi) := \mathrm{dom}(\phi)$, since the domain of a function is implicit in its definition. Thus, axiom [*V] holds. For the remaining axioms, let $f : X \to Y$ be a bijection, and consider $\phi \in \mathfrak{F}_X$. We define

$$f_*(\phi) := f \circ \phi \circ f^{-1} : Y \to Y$$

This defines a function $f_* : \mathfrak{F}_X \to \mathfrak{F}_Y$. It is now a routine if somewhat tedious matter to check that this definition satisfies axioms [*C1], [*C2], and [*C3].

The next issue is to show that these axioms imply the conditions of Definition 11.1.

Proposition 12.4. *Let $\mathcal{A}$ be a natural class of structures. Then $\mathcal{A}$ satisfies the conditions of Definition 11.1.*

Proof. Let X and Y be finite sets, and assume that $\alpha \in \mathcal{A}_X \cap \mathcal{A}_Y$. From axiom [*V] we have $\mathsf{V}_{\mathcal{A}}(\alpha) = X$ and $\mathsf{V}_{\mathcal{A}}(\alpha) = Y$, so that $X = Y$. This implies condition (i). For the second condition, let X and Y be finite sets such that $\#X = \#Y$. By Proposition 1.1 there is a bijection $f : X \to Y$. By axiom [*C1] there is thus a bijection $f_* : \mathcal{A}_X \to \mathcal{A}_Y$. By Proposition 1.1 again, it follows that $\#\mathcal{A}_X = \#\mathcal{A}_Y$, verifying condition (ii). □

The main problem with Section 11 is that the concept of equivalence of classes used there was much too weak to be really interesting. The correct idea of equivalence takes some getting used to but enables us to see some beautiful subtleties which were invisible before. In order to define it we first need to properly describe a few things that we've already done.

Definition 12.5 (Natural Transformations)**.** Let $\mathcal{A}$ and $\mathcal{B}$ be natural classes. A *natural transformation τ from $\mathcal{A}$ to $\mathcal{B}$* is denoted by $\tau : \mathcal{A} \Rightarrow \mathcal{B}$ and is defined as follows. For each finite set X there is a function $\tau_X : \mathcal{A}_X \to \mathcal{B}_X$, and these satisfy

the following axiom:
[***T**] Let $f : X \to Y$ be a bijection between finite sets. Then the diagram

$$\begin{array}{ccc} \mathcal{A}_X & \xrightarrow{f_{\mathcal{A}}} & \mathcal{A}_Y \\ \tau_X \downarrow & & \downarrow \tau_Y \\ \mathcal{B}_X & \xrightarrow{f_{\mathcal{B}}} & \mathcal{B}_Y \end{array}$$

commutes. In less visual terms, this means that $\tau_Y \circ f_{\mathcal{A}} = f_{\mathcal{B}} \circ \tau_X$ as functions from $\mathcal{A}_X$ to $\mathcal{B}_Y$.

The intuitive content of this definition is as follows. Each τ_X is a "procedure for changing an $\mathcal{A}$–structure on X into a $\mathcal{B}$–structure on X". Commutativity of the diagram says that τ_X and τ_Y really are the same procedure: only the names of the elements of the underlying vertex–set have been changed according to the bijection $f : X \to Y$. This does not change the effect of the transformation τ. That is, the transformation τ does not depend on the names of the elements of the set underlying the structures on which it acts.

Example 12.6. Let $\mathcal{R} := \mathcal{T}^\bullet$ be the class of rooted trees, and let $\mathfrak{F}$ be the class of endofunctions. In Example 11.22 we considered $\mathcal{R}$ to be a subclass of $\mathfrak{F}$. More precisely, we were considering a natural transformation $\tau : \mathcal{R} \Rightarrow \mathfrak{F}$ defined as follows. For a finite set X, the function $\tau_X : \mathcal{R}_X \to \mathfrak{F}_X$ takes as input a rooted tree $(T, v) \in \mathcal{R}_X$ and returns as output the following endofunction $\tau_X(T, v) := \phi \in \mathfrak{F}_X$. We let $\phi(v) := v$, and for every other $w \in X$ we let $\phi(w)$ be the unique parent of w in the rooted tree (T, v). Care must be taken to verify the axiom [*T], but again it follows directly from the definitions. All the functions τ_X are injective in this example, so we can consider $\mathcal{R}$ as a subsclass of $\mathfrak{F}$ *via* the natural transformation τ.

Example 12.7. We define a natural transformation η from the class $\mathfrak{F}$ of endofunctions to the class $\mathcal{G}$ of graphs, as follows. For any finite set X, the function $\eta_X : \mathfrak{F}_X \to \mathcal{G}_X$ is defined as follows. For any endofunction $\phi \in \mathfrak{F}_X$, the graph $\eta_X(\phi)$ is defined as follows: $\eta_X(\phi) := (X, E)$ in which

$$E := \{\{v, w\} \subseteq X : \ v \neq w \text{ and either } w = \phi(v) \text{ or } v = \phi(w)\}.$$

It is a good exercise to check that axiom [*T] holds in this example. Notice that in general the functions η_X are neither surjective nor injective.

Definition 12.8 (Natural Equivalence)**.** Let $\mathcal{A}$ and $\mathcal{B}$ be natural classes. A *natural equivalence* between $\mathcal{A}$ and $\mathcal{B}$ is a pair of natural transformations $\tau : \mathcal{A} \Rightarrow \mathcal{B}$ and $\rho : \mathcal{B} \Rightarrow \mathcal{A}$ such that for every finite set X, the functions $\tau_X : \mathcal{A}_X \to \mathcal{B}_X$ and $\rho_X : \mathcal{B}_X \to \mathcal{A}_X$ are mutually inverse bijections. If there exists a natural equivalence between the classes $\mathcal{A}$ and $\mathcal{B}$ then we say that these classes are *naturally equivalent*, and denote this relation by $\mathcal{A} \equiv \mathcal{B}$.

This supersedes the notation used in Section 11 but does not contradict it – in every case in which the symbol $\equiv$ was used in Section 11 the classes are in fact **naturally** equivalent.

Next, we revisit the constructions of Section 11 for natural classes.

Definition 12.9 (Sums of Classes)**.** Let $(\mathcal{A}^{(j)} : j \geq 1)$ be a locally finite sequence of natural classes. Then the sum $\mathcal{B} := \bigoplus_{j=1}^{\infty} \mathcal{A}^{(j)}$ is a natural class. We check the axioms of Definition 12.1 for $\mathcal{B}$. For an arbitrary $\mathcal{B}$–structure $\beta = (i, \alpha)$ we let $\mathsf{V}_{\mathcal{B}}(\beta) := \mathsf{V}_{\mathcal{A}^{(i)}}(\alpha)$. For a finite set X, if $\beta \in \mathcal{B}_X$ then $\alpha \in \mathcal{A}^{(i)}_X$, so that $\mathsf{V}_{\mathcal{B}}(\beta) = X$ as desired, by axiom [*V] for $\mathcal{A}^{(i)}$. This verifies [*V] for $\mathcal{B}$. Consider a bijection $f : X \to Y$ between finite sets. We define the induced bijection $f_{\mathcal{B}} : \mathcal{B}_X \to \mathcal{B}_Y$ as follows: for each $\beta = (i, \alpha) \in \mathcal{B}_X$, let $f_{\mathcal{B}}(\beta) := (i, f_{\mathcal{A}^{(i)}}(\alpha))$. Since each $f_{\mathcal{A}^{(i)}}$ is a bijection from $\mathcal{A}^{(i)}_X$ to $\mathcal{A}^{(i)}_Y$, it follows that $f_{\mathcal{B}}$ is a bijection from $\mathcal{B}_X$ to $\mathcal{B}_Y$. This establishes [*C1] for $\mathcal{B}$. Axioms [*C2] and [*C3] follow from the construction of the induced bijections $f_{\mathcal{B}}$, as can be checked.

Definition 12.10 (Subclasses and Difference of Classes)**.** For natural classes $\mathcal{A}$ and $\mathcal{B}$, in order to say that $\mathcal{A}$ *is a subclass of* $\mathcal{B}$ we not only need $\mathcal{A}_X \subseteq \mathcal{B}_X$ for every finite set X, but also we require that $\mathsf{V}_{\mathcal{A}}(\alpha) = \mathsf{V}_{\mathcal{B}}(\alpha) = X$ for all $\alpha \in \mathcal{A}_X$, and that $f_{\mathcal{A}}(\alpha) = f_{\mathcal{B}}(\alpha) \in \mathcal{A}_Y$ for all $\alpha \in \mathcal{A}_X$ and $f : X \to Y$. In this case, the difference $\mathcal{B} \setminus \mathcal{A}$ is again a natural class, with operations $\mathsf{V}_{\mathcal{B}\setminus\mathcal{A}}$ and $f_{\mathcal{B}\setminus\mathcal{A}}$ induced from the corresponding operations on $\mathcal{B}$. Thus, $\mathcal{B} \setminus \mathcal{A}$ is also a subclass of $\mathcal{B}$.

Definition 12.11 (Superposition of Classes)**.** For natural classes $\mathcal{A}$ and $\mathcal{B}$, the superposition $\mathcal{A}\&\mathcal{B}$ is also a natural class. For an $(\mathcal{A}\&\mathcal{B})$–structure (α, β) we may define $\mathsf{V}_{\mathcal{A}\&\mathcal{B}}(\alpha, \beta) := \mathsf{V}_{\mathcal{A}}(\alpha)$ to satisfy axiom [*V]. If $(\alpha, \beta) \in (\mathcal{A}\&\mathcal{B})_X$ and $f : X \to Y$ is a bijection, then

$$f_{\mathcal{A}\&\mathcal{B}}(\alpha, \beta) := (f_{\mathcal{A}}(\alpha), f_{\mathcal{B}}(\beta))$$

satisfies axioms [*C1], [*C2], and [*C3].

Definition 12.12 (Products and Powers of Classes)**.** For natural classes $\mathcal{A}$ and $\mathcal{B}$, the product $\mathcal{A} * \mathcal{B}$ is also a natural class. For an $(\mathcal{A} * \mathcal{B})$–structure (α, β) we may define $\mathsf{V}_{\mathcal{A}*\mathcal{B}}(\alpha, \beta) := \mathsf{V}_{\mathcal{A}}(\alpha) \cup \mathsf{V}_{\mathcal{B}}(\beta)$ to satisfy axiom [*V]. If $(\alpha, \beta) \in (\mathcal{A} * \mathcal{B})_X$ and $f : X \to Y$ is a bijection, then

$$f_{\mathcal{A}*\mathcal{B}}(\alpha, \beta) := (f_{\mathcal{A}}(\alpha), f_{\mathcal{B}}(\beta))$$

satisfies axioms [*C1], [*C2], and [*C3]. Iterating this construction shows that for any natural class $\mathcal{A}$, each of the powers $\mathcal{A}^k$ is also a natural class.

Definition 12.13 (Finite Strings and Connected Classes)**.** If $\mathcal{A}$ is a connected natural class then the sequence $(\mathcal{A}^k : k \in \mathbb{N})$ of powers of $\mathcal{A}$ is locally finite, and it follows from Definitions 12.9 and 12.12 that $\mathcal{A}^*$ is a natural class.

Definition 12.14 (Rooted Structures)**.** For a natural class $\mathcal{A}$, the class $\mathcal{A}^\bullet$ is also natural. For an $\mathcal{A}^\bullet$–structure (α, v) we may define $\mathsf{V}_{\mathcal{A}^\bullet}(\alpha, v) := \mathsf{V}_{\mathcal{A}}(\alpha)$ to satisfy axiom [*V]. If $(\alpha, v) \in \mathcal{A}^\bullet_X$ and $f : X \to Y$ is a bijection, then

$$f_{\mathcal{A}^\bullet}(\alpha, v) := (f_{\mathcal{A}}(\alpha), f(v))$$

satisfies axioms [*C1], [*C2], and [*C3].

Definition 12.15 (Composition of Classes)**.** If $\mathcal{A}$ and $\mathcal{B}$ are natural classes, with $\mathcal{A}$ connected, then $\mathcal{B}[\mathcal{A}]$ is a natural class. First consider the special case of the class $\mathcal{E}[\mathcal{A}]$. For an $\mathcal{E}[\mathcal{A}]$–structure ξ we may define

$$\mathsf{V}_{\mathcal{E}[\mathcal{A}]}(\xi) := \bigcup_{\alpha \in \xi} \mathsf{V}_{\mathcal{A}}(\alpha)$$

to satisfy axiom [*V]. If $f : X \to Y$ is a bijection, then the induced bijection $f_{\mathcal{E}[\mathcal{A}]} : \mathcal{E}[\mathcal{A}]_X \to \mathcal{E}[\mathcal{A}]_Y$ may be defined as follows. For each $\xi \in \mathcal{E}[\mathcal{A}]_X$ let

$$f_{\mathcal{E}[\mathcal{A}]}(\xi) := \{(f|_{\mathsf{V}_{\mathcal{A}}(\alpha)})_{\mathcal{A}}(\alpha) : \ \alpha \in \xi\}.$$

That is, for each $\mathcal{A}$–structure α in ξ, let $f|_{\mathsf{V}_{\mathcal{A}}(\alpha)}$ be the restriction of f to the subset $\mathsf{V}_{\mathcal{A}}(\alpha) \subseteq X$. This gives a bijection from $\mathsf{V}_{\mathcal{A}}(\alpha)$ to some subset of Y, and we take the image of α under the corresponding induced bijection $(f|_{\mathsf{V}_{\mathcal{A}}(\alpha)})_{\mathcal{A}}$ of $\mathcal{A}$–structures. One can check the axioms [*C1], [*C2], and [*C3] for this construction. Moreover, there is another bijection to be considered, since both $\xi \in \mathcal{E}[\mathcal{A}]_X$ and $f_{\mathcal{E}[\mathcal{A}]}(\xi)$ are finite sets. Namely, there is a unique bijection $\widehat{f} : \xi \to f_{\mathcal{E}[\mathcal{A}]}(\xi)$ with the property that for all $\alpha \in \xi$,

$$\mathsf{V}_{\mathcal{A}}(\widehat{f}(\alpha)) = \{f(v) : \ v \in \mathsf{V}_{\mathcal{A}}(\alpha)\}.$$

This merely says that $\widehat{f}(\alpha)$ is the part of $f_{\mathcal{E}[\mathcal{A}]}(\xi)$ which corresponds to α under the renaming of vertices $f : X \to Y$. Now we can check the axioms in the general case $\mathcal{B}[\mathcal{A}]$ of composition of classes. Given a $\mathcal{B}[\mathcal{A}]$–structure (ξ, β) we may define $\mathsf{V}_{\mathcal{B}[\mathcal{A}]}(\xi, \beta) := \mathsf{V}_{\mathcal{E}[\mathcal{A}]}(\xi)$ to satisfy axiom [*V]. If $(\xi, \beta) \in \mathcal{B}[\mathcal{A}]_X$ and $f : X \to Y$ is a bijection, then we let

$$f_{\mathcal{B}[\mathcal{A}]}(\xi, \beta) := (f_{\mathcal{E}[\mathcal{A}]}(\xi), \widehat{f}_{\mathcal{B}}(\beta)).$$

We have used the bijection $\widehat{f} : \xi \to f_{\mathcal{E}[\mathcal{A}]}(\xi)$ to induce the bijection $\widehat{f}_{\mathcal{B}} : \mathcal{B}_\xi \to \mathcal{B}_{f_{\mathcal{E}[\mathcal{A}]}(\xi)}$ which is then applied to the $\mathcal{B}$–structure $\beta \in \mathcal{B}_\xi$. One can check the axioms [*C1], [*C2], and [*C3] for this construction, but it is rather involved.

In summary, we have seen that all the constructions of Section 11 can be carried through more precisely for natural classes, and yield natural classes in return. This level of detail is not always appropriate when solving particular problems, but it is important to establish the foundations of the theory. It is a worthwhile exercise to review Section 11 and understand why each of the equivalences discussed there is in fact a natural equivalence.

There are classes which are equivalent but are not naturally equivalent – so from now on we speak of *numerical equivalence* when referring to the weaker relation. Notice that the class $\mathcal{S}$ of permutations and the class $\mathcal{L} := \mathcal{X}^*$ of total orders are numerically equivalent, since they both have exponential generating function $S(x) = (1-x)^{-1} = L(x)$. The obvious question arises as to whether or not these classes are naturally equivalent. In fact they are not naturally equivalent, and much more is true: **there are no natural transformations from $\mathcal{S}$ to $\mathcal{L}$**. This is a very strong statement! There seem to be many possibilities for attempting to define a natural transformation from $\mathcal{S}$ to $\mathcal{L}$. How can we be sure that none of them will work? The key idea is that with a natural transformation $\tau : \mathcal{A} \Rightarrow \mathcal{B}$ information can only be lost (or at best preserved – never added) when passing from $\alpha \in \mathcal{A}_X$ to $\tau_X(\alpha) \in \mathcal{B}_X$. As a consequence of this, $\tau_X(\alpha)$ has at least as many "symmetries" as α has. We'll next make these cryptic comments precise, and then apply this strategy to the classes $\mathcal{S}$ and $\mathcal{L}$.

Definition 12.16 (Automorphisms)**.** Let $\mathcal{A}$ be a natural class, X a finite set, and $\alpha \in \mathcal{A}_X$. An *automorphism* of α is a permutation $\sigma \in \mathcal{S}_X$ such that $\sigma_*(\alpha) = \alpha$. Here we are regarding $\sigma : X \to X$ as a bijection, and considering the induced bijection $\sigma_* : \mathcal{A}_X \to \mathcal{A}_X$ guaranteed by axiom [*C1]. Let $\mathrm{aut}(\alpha)$ denote the set of all automorphisms of α.

Proposition 12.17. *Let $\mathcal{A}$ be a natural class, X a finite set, and $\alpha \in \mathcal{A}_X$. Then* $\mathrm{aut}(\alpha)$ *is a subgroup of* $\mathcal{S}_X$*.*

Proof. Certainly $\mathrm{aut}(\alpha)$ is a subset of $\mathcal{S}_X$, and hence is finite. By axiom [*C2] we have $(\mathrm{id}_X)_* = \mathrm{id}_{\mathcal{A}_X} : \mathcal{A}_X \to \mathcal{A}_X$, so that $(\mathrm{id}_X)_*(\alpha) = \alpha$ and therefore $\mathrm{id}_X \in \mathrm{aut}(\alpha)$. If $\pi, \sigma \in \mathrm{aut}(\alpha)$ then – using axiom [*C3] – we have

$$(\pi \circ \sigma)_*(\alpha) = \pi_*(\sigma_*(\alpha)) = \pi_*(\alpha) = \alpha,$$

so that $\pi \circ \sigma \in \mathrm{aut}(\alpha)$ as well. This shows that $\mathrm{aut}(\alpha)$ is a finite set of permutations which contains the identity permutation and is closed under functional composition. From this it follows that $\mathrm{aut}(\alpha)$ is a group, necessarily a subgroup of $\mathcal{S}_X$. □

Definition 12.18. Let $\mathfrak{P}$ denote the class of *permutation groups*: for any finite set X, $\mathfrak{P}_X$ is the set of all subgroups Γ of $\mathcal{S}_X$. Given a permutation group Γ, we can determine its identity element ι and define $\mathsf{V}_{\mathfrak{P}}(\Gamma) := \mathrm{dom}(\iota)$, the domain of the bijection ι. This verifies axiom [*V]. We leave as an exercise the verification of the axioms [*C1] through [*C3], noting only that for a bijection $f : X \to Y$, the induced bijection is defined by

$$f_*(\Gamma) := \{f \circ \pi \circ f^{-1} : \ \pi \in \Gamma\}.$$

Proposition 12.19. *For any natural class $\mathcal{A}$, the construction* $\mathrm{aut} : \mathcal{A} \Rightarrow \mathfrak{P}$ *is a natural transformation.*

Proof. We check the axiom [*T] for aut. Let $f : X \to Y$ be a bijection between finite sets, and let $\alpha \in \mathcal{A}_X$. We must chack that $f_{\mathfrak{P}}(\mathrm{aut}(\alpha)) = \mathrm{aut}(f_{\mathcal{A}}(\alpha))$. Consider an arbitrary permutation $\sigma \in \mathcal{S}_Y$. Then $\sigma \in \mathrm{aut}(f_{\mathcal{A}}(\alpha))$ if and only if $\sigma_{\mathcal{A}}(f_{\mathcal{A}}(\alpha)) = f_{\mathcal{A}}(\alpha)$. On the other hand, $\sigma \in f_{\mathfrak{P}}(\mathrm{aut}(\alpha))$ if and only if $\sigma = f \circ \pi \circ f^{-1}$ for some $\pi \in \mathrm{aut}(\alpha)$. This condition implies that

$$\sigma_{\mathcal{A}}(f_{\mathcal{A}}(\alpha)) = (f_{\mathcal{A}} \circ \pi_{\mathcal{A}} \circ f_{\mathcal{A}}^{-1} \circ f_{\mathcal{A}})(\alpha) = f_{\mathcal{A}}(\alpha),$$

using axioms [*C2] and [*C3]. This shows that $f_{\mathfrak{P}}(\mathrm{aut}(\alpha))$ is a subset of $\mathrm{aut}(f_{\mathcal{A}}(\alpha))$. Conversely, for any $\sigma \in \mathrm{aut}(f_{\mathcal{A}}(\alpha))$, let $\pi := f^{-1} \circ \sigma \circ f$. Now

$$\pi(\alpha) = f_{\mathcal{A}}^{-1}(\sigma_{\mathcal{A}}(f_{\mathcal{A}}(\alpha))) = f_{\mathcal{A}}^{-1}(f_{\mathcal{A}}(\alpha)) = \alpha,$$

so that $\pi \in \mathrm{aut}(\alpha)$. Also, since

$$f \circ \pi \circ f^{-1} = f \circ f^{-1} \circ \sigma \circ f \circ f^{-1} = \sigma,$$

this shows that $\mathrm{aut}(f_{\mathcal{A}}(\alpha))$ is a subset of $f_{\mathfrak{P}}(\mathrm{aut}(\alpha))$ as well. This completes the proof. □

Proposition 12.20. *Let $\mathcal{A}$ and $\mathcal{B}$ be natural classes, and let $\tau : \mathcal{A} \Rightarrow \mathcal{B}$ be a natural transformation. For any finite set X and $\alpha \in \mathcal{A}_X$,* $\mathrm{aut}(\alpha)$ *is a subgroup of* $\mathrm{aut}(\tau(\alpha))$.

Proof. Consider any $\sigma \in \mathrm{aut}(\alpha)$, so that $\sigma_{\mathcal{A}}(\alpha) = \alpha$. By axiom [*T], we have $\sigma_{\mathcal{B}}(\tau(\alpha)) = \tau(\sigma_{\mathcal{A}}(\alpha)) = \tau(\alpha)$, so that $\sigma \in \mathrm{aut}(\tau(\alpha))$ as well. □

Example 12.21. Now we can finally show that there is no natural transformation from $\mathcal{S}$ to $\mathcal{L}$. Suppose that there were such a transformation $\tau : \mathcal{S} \Rightarrow \mathcal{L}$. Fix $n \in \mathbb{N}$ and let $\iota \in \mathcal{S}_n$ be the identity permutation on N_n. Then $\ell := \tau(\iota)$ is a total order on N_n and, by Proposition 12.20, $\mathrm{aut}(\iota)$ is a subgroup of $\mathrm{aut}(\ell)$. However, as is easily seen, $\mathrm{aut}(\iota) = \mathcal{S}_n$ and $\mathrm{aut}(\ell) = \{\iota\}$. For all $n \geq 2$ it is impossible for $\mathcal{S}_n$ to be a subgroup of $\{\iota\}$, and therefore the hypothetical τ does not exist.

We close our discussion of exponential generating functions with an example which could have been included at the end of Section 11. However, the combinatorics is a little complicated and the algebra is **very** complicated, so we have postponed it until now. We limit ourselves to merely sketching the main steps, leaving the verification of details and the substantial algebraic manipulations as good exercises.

Definition 12.22 (Nested Set Systems)**.** A *nested set system on the set X* is a pair (X, Δ) in which X is a finite set and Δ is a set of subsets of X such that
• If $A, B \in \Delta$ then either $A \subseteq B$ or $B \subseteq A$ or $A \cap B = \varnothing$.
(An example is illustrated in Figure 12.1.) Let $\mathcal{N}_X$ denote the set of all nested set systems on the set X. Axioms [*V], [*C1], [*C2], [*C3] are easily verified, so that this defines the natural class $\mathcal{N}$ of nested set systems.

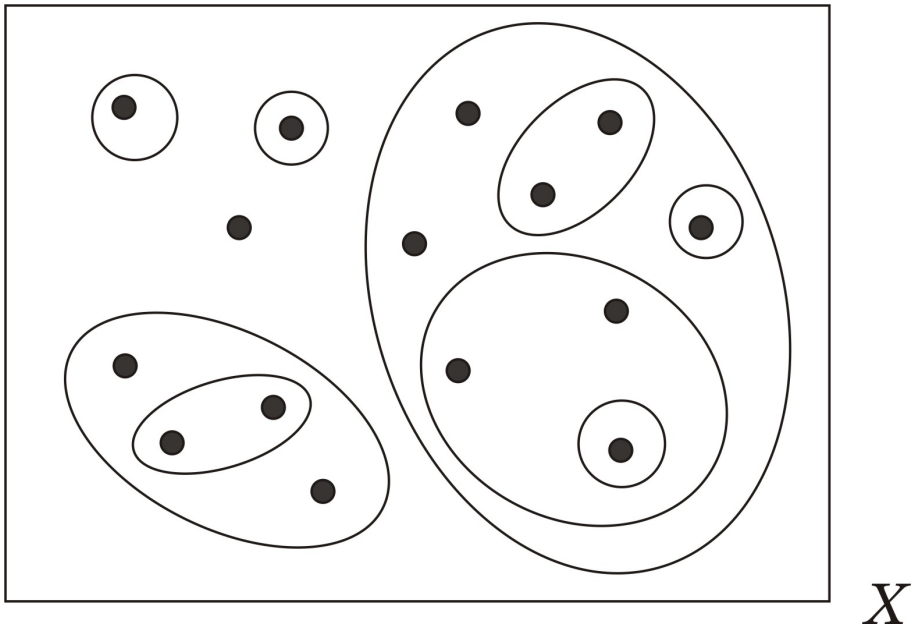


FIGURE 12.1. a nested set system.

To analyze the class $\mathcal{N}$, first notice that $(\varnothing, \varnothing)$ and $(\varnothing, \{\varnothing\})$ are two different $\mathcal{N}$–structures on the empty set. Also, if (X, Δ) is a nested set sytem and $v \in X$ is such that $\{v\} \notin \Delta$, then $(X, \Delta \cup \{\{v\}\})$ is also a nested set system. Let's say that a nested set sytem (X, Δ) is *proper* if $A \in \Delta$ implies $\#A \geq 2$, and denote by $\mathcal{M}$ the subclass of $\mathcal{N}$ consisting of the proper set systems. The *proper part* of $(X, \Delta) \in \mathcal{N}_X$ is (X, Δ°) in which $\Delta^\circ := \{A \in \Delta : \ \#A \geq 2\}$. Let $\mathcal{P}$ be the natural class such that for any finite set X,

$$\mathcal{P}_X := \{(X, A) : \ A \subseteq X\}.$$

For each finite set X, define a function

$$\begin{array}{rcl} \tau_X : \mathcal{N}_X & \to & \mathcal{N}_\varnothing \times \mathcal{P}_X \times \mathcal{M}_X \\ (X, \Delta) & \mapsto & ((\varnothing, \{A \in \Delta : \ A = \varnothing\}), (X, \{v \in X : \ \{v\} \in \Delta\}), (X, \Delta^\circ)). \end{array}$$

We leave it as an exercise to show that this construction defines a natural transformation $\tau : \mathcal{N} \Rightarrow \mathcal{N}_\varnothing * (\mathcal{P}\&\mathcal{M})$, and that moreover this transformation is one part of a natural equivalence $\mathcal{N} \equiv \mathcal{N}_\varnothing * (\mathcal{P}\&\mathcal{M})$.

To further analyze the class $\mathcal{M}$, let's say that a nested set system (X, Δ) is a *cell* if it is proper and $X \in \Delta$. Let $\mathcal{Q}$ be the class of cells. If (X, Δ) is a cell then $(X, \Delta \smallsetminus \{X\})$ is a proper nested set system such that:
• $\#X \geq 2$ (since if $\#X \leq 1$ then (X, Δ) would not have been proper), and
• X is not in $\Delta \smallsetminus \{X\}$.

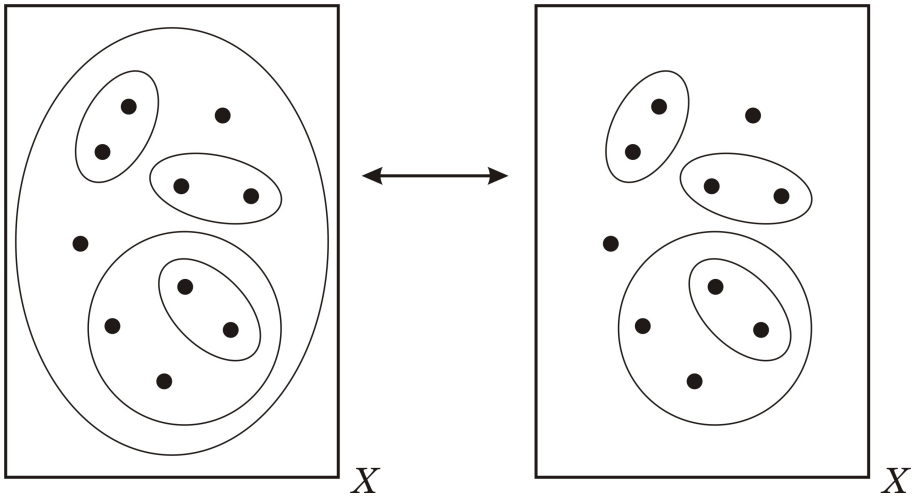


FIGURE 12.2. the equivalence $\mathcal{Q} \equiv \mathcal{M} \setminus (\mathcal{E}_0 \oplus \mathcal{E}_1 \oplus \mathcal{Q})$.

A little thought shows that this gives rise to a natural equivalence

$$\mathcal{Q} \equiv \mathcal{M} \setminus (\mathcal{E}_0 \oplus \mathcal{E}_1 \oplus \mathcal{Q}).$$

See Figure 12.2 for an example of this equivalence.

Finally, an arbitrary proper nested set system can be expressed uniquely as the disjoint union of a collection of cells and singleton vertices, so that

$$\mathcal{M} \equiv \mathcal{E} * \mathcal{E}[\mathcal{Q}].$$

These recursive relations among the classes $\mathcal{N}$, $\mathcal{M}$, and $\mathcal{Q}$ lead to functional equations relating the exponential generating function

$$N(x,y) := \sum_{n=0}^{\infty} \left(\sum_{(N_n,\Delta)\in\mathcal{N}_n} y^{\#\Delta} \right) \frac{x^n}{n!}$$

to the analogous exponential generating functions $M(x,y)$ and $Q(x,y)$ for the subclasses $\mathcal{M}$ and $\mathcal{Q}$, respectively. The remainder of the solution to this enumeration problem (determining $\#\mathcal{N}_n$ for all $n \in \mathbb{N}$) is relegated to a series of exercises.

12. Exercises.

1. In the following, $\mathcal{A}$ denotes a connected class, $\mathcal{C}$ is the class of cycles, $\mathfrak{F}$ is the class of endofunctions, $\mathcal{L} := \mathcal{X}^*$ is the class of total orders, $\mathcal{R} := \mathcal{T}^\bullet$ is the class of rooted trees, and $\mathcal{S}$ is the class of permutations. Prove the following natural equivalences.

(a) $\mathcal{C}^\bullet \equiv \mathcal{X} * \mathcal{L}$.

(b) $\mathcal{L}^\bullet \equiv \mathcal{L} * \mathcal{X} * \mathcal{L}$.

(c) $\mathcal{E}[\mathcal{A}]^\bullet \equiv \mathcal{E}[\mathcal{A}] * \mathcal{A}^\bullet$.

(d) $\mathcal{C}[\mathcal{R}]^\bullet \equiv \mathcal{L}^\bullet[\mathcal{R}]$.

(e) What expression for $\mathfrak{F}^\bullet$ follows from (c) and (d)?

(f) $\mathcal{S}\&\mathcal{L} \equiv \mathcal{L}\&\mathcal{L}$.

---

2. The classes $\mathcal{T}^{\bullet\bullet\bullet}$ of triply–rooted trees and $\mathfrak{F}^\bullet$ of rooted endofunctions are numerically equivalent, since $\#\mathcal{T}^{\bullet\bullet\bullet} = n^{n+1} = \#\mathcal{F}_n^\bullet$ for all $n \in \mathbb{N}$. Are these classes naturally equivalent?

3.* Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{D}$ be natural classes such that $\mathcal{A} \oplus \mathcal{D} \equiv \mathcal{B} \oplus \mathcal{D}$. Prove that $\mathcal{A} \equiv \mathcal{B}$. (This is trivial for numerical equivalence! For natural equivalence it is quite subtle.)

Exercises 4, 5, and 6 could have been put in Chapter 11, but as they are rather more difficult than the eaarlier ones I've chosen to postpone them until now.

---

4. A *labelled plane tree* (LPT) is a tree with vertex–set $\{1, 2, \ldots, n\}$ (for some $n \in \mathbb{N}$) which is embedded in the plane as a planar graph. Two embeddings are considered the same if and only if they are "ambient isotopic"; this means that the whole plane may be stretched and squished like a rubber sheet to bring one embedding onto the other. Folding or tearing is not allowed. For example, of the LPTs pictured in Figure 12.3, the center one is equivalent to the one on the left, but not to the one on the right. Let $h(n)$ be the number of (equivalence classes of) LPTs with vertex–set $\{1, 2, .., n\}$, for each $n \in \mathbb{N}$. The first few values are $h(0) = 0$, $h(1) = 1$, $h(2) = 1$, $h(3) = 9$, and $h(4) = 20$. Determine $h(n)$ for all $n \in \mathbb{N}$.

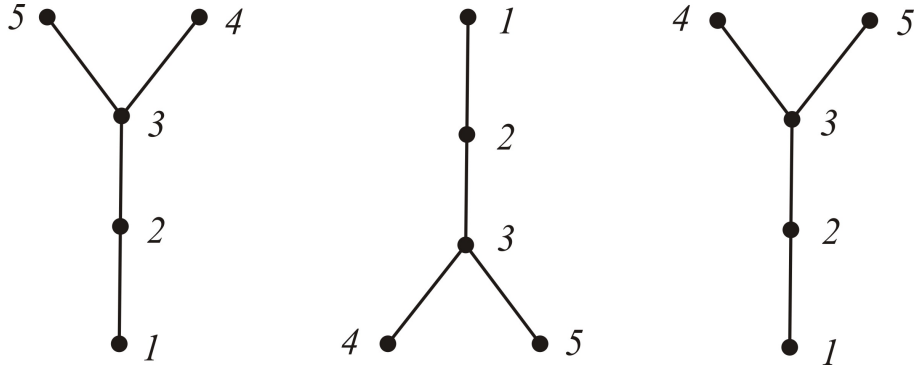


FIGURE 12.3. labelled plane trees.

5. For a finite graph $G = (V, E)$, the *Wiener index* of G is defined to be

$$W(G) := \sum_{v \neq w} \mathrm{dist}_G(v, w),$$

in which the sum is over all unordered pairs of distinct vertices of G, and $\mathrm{dist}_G(v, w)$ denotes the distance between v and w in G. Give a formula for the average value of $W(T)$ as T ranges over the set of all trees with vertex-set $\{1, 2, \ldots, n\}$, for each $n \in \mathbb{N}$.

---

6. For a finite graph $G = (V, E)$, the *productivity* of G is defined to be

$$\pi(G) := \prod_{v \in V} \deg_G(v),$$

in which the product is over all vertices of G, and $\deg_G(v)$ denotes the number of edges incident with v in G. (This is the number of endofunctions $\phi : V \to V$ such that $\{v, \phi(v)\} \in E$ for all $v \in V$.) Give a formula for the average value of $\pi(T)$ as T ranges over the set of all trees with vertex–set $\{1, 2, \ldots, n\}$, for each $n \in \mathbb{N}$.

The remaining exercises outline a solution to the enumeration of nested set–systems.

---

7. Show that the class $\mathcal{N}$ of nested set systems is a natural class.

---

8. Show that the functions $\tau_X : \mathcal{N}_X \to \mathcal{N}_\varnothing \times \mathcal{P}_X \times \mathcal{M}_X$ constructed above define a natural transformation which is one part of a natural equivalence $\mathcal{N} \equiv \mathcal{N}_\varnothing * (\mathcal{P}\&\mathcal{M})$.

---

9. Give detailed justifications of the natural equivalences

$$\mathcal{Q} \equiv \mathcal{M} \setminus (\mathcal{E}_0 \oplus \mathcal{E}_1 \oplus \mathcal{Q})$$

and

$$\mathcal{M} \equiv \mathcal{E} * \mathcal{E}[\mathcal{Q}]$$

discussed above.

---

10. Derive functional equations relating $N(x,y)$, $M(x,y)$ and $Q(x,y)$ from the natural equivalences among $\mathcal{N}$, $\mathcal{M}$, and $\mathcal{Q}$.

---

11. Use the functional equations of Exercise 10 to show that

$$N(x,y) = \frac{(1+y)^2}{y} R\left(\frac{y}{1+y}\exp\left(\frac{x-y+xy}{1+y}\right)\right),$$

in which $R(t) = t\exp(R(t))$. (Hint: the change of variables $z := y/(1+y)$ is very useful.)

---

12. A *normed ring* is a commutative ring R with a *norm function* $|\cdot| : R \to [0,\infty)$ that satisfies the following axioms:
• $|0| = 0$ and $|1| = 1$,
• for all $a, b \in R$, $|ab| = |a||b|$, and
• for all $a, b \in R$, $|a+b| \leq |a| + |b|$.
For example, $\mathbb{C}$ with the usual modulus function is a normed ring.

When R is a normed ring, we may relax the definition of convergence of a sequence of formal power series $f_k(x)$ as $k \to \infty$, as follows. We require that for all $n \in \mathbb{N}$ there exists a constant $A_n \in R$ such that for every real $\varepsilon > 0$ there exists a $K = K(n,\varepsilon)$ such that for all $k \geq K$,

$$|A_n - [x^n]f_k(x)| < \varepsilon.$$

If this holds then the formal power series $F(x) = \sum_{n=0}^{\infty} A_n x^n$ is the limit of the sequence $(f_k(x) : \ k \geq 1)$.

(a) Show that if $(f_k(x) : \ k \geq 1)$ converges in the sense of Section 7, then it converges in the above sense.

(b) Give an example of a sequence of formal power series in $\mathbb{R}[[x]]$ which converges in the above sense but not in the sense of Section 7.

(c) Extend this definition of convergence to sequences of formal Laurent series over a normed ring.

13(a) By setting $y = 1$ in Exercise 11, obtain an expression for $\sum_{n=0}^{\infty}(\#\mathcal{N}_n)x^n/n!$ as the limit of a sequence of formal power series in $\mathbb{R}[[x]]$ which converges in the sense of Exercise 12.

13(b) Deduce that for all $n \in \mathbb{N}$:

$$\#\mathcal{N}_n = 4 \sum_{k=1}^{\infty} \frac{k^{n+k-1}}{k! 2^k e^{k/2}}.$$

---

12. Endnotes.

Our approach to the foundations of exponential generating functions follows the ground–breaking paper of Joyal:

• A. Joyal, *Une théorie combinatoire des séries formelles,* Adv. in Math. **42** (1981), 1–82.

Further developments of this theory are explained in

• F. Bergeron, G. Labelle, and P. Leroux, "Combinatorial Species and Tree–Like Structures," Cambridge U.P., Cambridge, 1998.

Nested set–systems are naturally equivalent to certain other combinatorial structures. Their enumeration was first given in

• J. P. Hayes, *Enumeration of fanout–free Boolean functions*, J. ACM, **23** (1976), 700-709.

See also

• K. L. Kodandapani and S. C. Seth, *On combinational networks with restricted fan–out*, IEEE Trans. Computers, **27** (1978), 309-318.

• L. R. Foulds and R. W. Robinson, *Determining the asymptotic number of phylogenetic trees*, pp. 110-126 of "Combinatorial Mathematics VII (Newcastle, August 1979)", ed. R. W. Robinson, G. W. Southern and W. D. Wallis. Lect. Notes Math., **829**. Springer, 1980.

These last three references were found with the help of Neil Sloane's *On–Line Encyclopedia of Integer Sequences*:

`http://www.research.att.com/~njas/sequences/index.html`

13. A Combinatorial Proof of the Lagrange Implicit Function Theorem.

In this section we give a proof of LIFT which uses slightly weaker hypotheses than that in Section 8. It also gives some combinatorial insight into the "meaning" of the formula which is not apparent from the previous algebraic proof.

Theorem 13.1 (LIFT)**.** *Let $\mathbb{K}$ be a commutative ring which contains the rational numbers $\mathbb{Q}$. Let $F(u)$ and $G(u)$ be formal power series in $\mathbb{K}[[u]]$ with $[u^0]G(u) \neq 0$.*
(a) *There is a unique (nonzero) formal power series $R(x)$ in $\mathbb{K}[[x]]$ such that*

$$R(x) = x\,G(R(x)).$$

(b) *The constant term of $R(x)$ is 0,and for all $n \geq 1$,*

$$[x^n]F(R(x)) = \frac{1}{n}[u^{n-1}]F'(u)G^n(u).$$

(Notice that we do not require $[u^0]G(u)$ to be invertible in $\mathbb{K}$.)

Proof. We prove this combinatorially by interpreting these formal power series as exponential generating functions for "generic" classes of structures. More precisely, we will interpret both sides combinatorially and define bijections which imply that

$$n![x^n]F(R(x)) = (n-1)![u^{n-1}]F'(u)G^n(u)$$

for all $n \geq 1$. To do this, let $f_0, f_1, f_2, \ldots$ and $g_0, g_1, g_2, \ldots$ be infinitely many indeterminates which commute pairwise and are algebraically independent over $\mathbb{K}$. Form the power series

$$F(u) = \sum_{n=0}^{\infty} f_n \frac{u^n}{n!} \quad \text{and} \quad G(u) = \sum_{n=0}^{\infty} g_n \frac{u^n}{n!}.$$

Thinking of these as exponential generating functions for classes $\mathcal{F}$ and $\mathcal{G}$, we see that f_n represents all $\mathcal{F}$–type structures on an n–element set; analogously for g_n as well. The fact that the f_i–s and g_j–s are indeterminates means that these series do not satisfy any special identities – any algebraic formula which can be proved valid for them must also be true if the indeterminates are specialized to have particular values; for example, if $f_n = \#\mathcal{A}_n$ for a particular class $\mathcal{A}$ of structures. It is in this sense that $F(u)$ and $G(u)$ are *generic exponential generating functions*.

In what sense, though, can we talk about a *generic class of structures*, which is what we want $\mathcal{F}$ (and $\mathcal{G}$) to be? In fact, a generic class $\mathcal{F}$ is just the class $\mathcal{E}$ of finite sets – the only difference is in what goes into the exponential generating function. We just agree to mark an n–element set with the indeterminate f_n, for each $n \in \mathbb{N}$.

If we specialize $f_n = \#\mathcal{A}_n$ for some class $\mathcal{A}$, then indeed there are f_n choices for putting an $\mathcal{A}$–structure on an n–element set, and the generic $F(u)$ specializes to the particular $A(u)$ in this case.

Now consider the natural equivalence $\mathcal{R} \equiv \mathcal{X} * \mathcal{G}[\mathcal{R}]$; this (implicitly) defines a class $\mathcal{R}$ for which the exponential generating function satisfies $R(x) = xG(R(x))$. An $\mathcal{R}$–structure on the set X consists of a rooted tree T with vertex–set X and, for each vertex $w \in X$, a $\mathcal{G}$–structure on the set of children of w in T. But the generic class $\mathcal{G}$ is just $\mathcal{E}$, so $\mathcal{R}$ is the class of all rooted labelled trees (RLTs). The only novelty is in how the indeterminates g_j enter the formula for the exponential generating function of $\mathcal{R}$.

Let $c(T, v, w)$ denote the number of children of the vertex w in the RLT (T, v). Recalling that a $\mathcal{G}$–structure on a j–element set is marked by g_j, we see that the equation $R(x) = xG(R(x))$ has the unique solution

$$R(x) = \sum_{n=0}^{\infty} \left(\sum_{(T,v)\in\mathcal{R}_n} \mathbf{g}^{(T,v)} \right) \frac{x^n}{n!},$$

in which

$$\mathbf{g}^{(T,v)} := \prod_{w\in \mathsf{V}_{\mathcal{R}}(T,v)} g_{c(T,v,w)}.$$

Notice that $[x^1]R(x) = g_0 \neq 0$ so that $R(x)$ is nonzero, and from $R(x) = xG(R(x))$ it follows that $[x^0]R(x) = 0$. This proves statement (a) in LIFT.

To prove statement (b), we analyze that formula combinatorially. The power series $F(R(x))$ may be interpreted using composition of classes. This must be the exponential generating function for $\mathcal{F}[\mathcal{R}]$, the class of forests of rooted labelled trees. In this generating function, f_r indicates a forest with exactly r connected components. Thus we have arrived at a combinatorial interpretation of the LHS of LIFT: $n![x^n]F(R(x))$ is the sum over all forests of RLTs with vertex–set $\{1, 2, .., n\}$, in which each forest φ contributes the monomial

$$M(\varphi) := f_{\#\varphi} \prod_{(T,v)\in\varphi} \mathbf{g}^{(T,v)}.$$

(Here we think of φ as a set of RLTs.) That is,

$$F(R(x)) = \sum_{n=0}^{\infty} \left(\sum_{\varphi\in\mathcal{F}[\mathcal{R}]_n} M(\varphi) \right) \frac{x^n}{n!}.$$

The next step is to find a similar interpretation for the RHS of LIFT. But this is easy! $F'(u)G^n(u)$ is u^{-1} times the exponential generating function of the class $\mathcal{F}^\bullet * \mathcal{G}^n$. A structure from this class on a finite set X is (naturally equivalent to) an ordered $(n+2)$–tuple $\sigma = (A, v, B_1, \ldots, B_n)$ in which $A, B_1, \ldots, B_n$ are pairwise

disjoint subsets of X which have X as their union, and v is a designated element of A. The contribution of σ to the exponential generating function is the monomial

$$m(\sigma) := f_{\#A} \prod_{i=1}^{n} g_{\#B_i}.$$

Therefore,

$$\begin{aligned}(n-1)![u^{n-1}]F'(u)G^n(u) &= n^{-1}n![u^n]uF'(u)G^n(u) \\ &= \frac{1}{n} \sum_{\sigma \in (\mathcal{F}^\bullet * \mathcal{G}^n)_n} m(\sigma)\end{aligned}$$

is $1/n$ times the sum of $m(\sigma)$ over all σ from the class $\mathcal{F}^\bullet * \mathcal{G}^n$ on the set $\{1, 2, \ldots, n\}$. That factor of $1/n$ is kind of annoying, but we can move it to the LHS by considering the class $\mathcal{F}[\mathcal{R}]^\bullet$ instead.

In summary, our strategy for proving LIFT is to show that for $n \geq 1$,

$$n!n[x^n]F(R(x)) = n![u^n]uF'(u)G^n(u)$$

by constructing a bijection between two sets. On the LHS is the set $\mathcal{F}[\mathcal{R}]_n^\bullet$ of pairs (φ, w) in which φ is a forest of RLTs with vertex–set $\{1, 2, .., n\}$ and w is a designated vertex of φ. On the RHS is the set $(\mathcal{F}^\bullet * \mathcal{G}^n)_n$ with elements σ as described above in the case $X = \{1, 2, \ldots, n\}$. Moreover, in this bijection, if (φ, w) corresponds to σ, then we require that $M(\varphi) = m(\sigma)$. In tabular form:

$$\begin{aligned}\mathcal{F}[\mathcal{R}]_n^\bullet &\rightleftharpoons (\mathcal{F}^\bullet * \mathcal{G}^n)_n \\ (\varphi, w) &\leftrightarrow \sigma = (A, v, B_1, \ldots, B_n) \\ M(\varphi) &= m(\sigma)\end{aligned}$$

Such a bijection will suffice to prove LIFT.

Before defining the bijection we're looking for, I want to remark that its construction is **not natural** in the sense of Section 12 – it will use the numerical order of the labels $\{1, 2, \ldots, n\}$ of the vertices in the underlying set. But this doesn't bother me too much. Notice that we are not trying to show that the classes $\mathcal{F}[\mathcal{R}]^\bullet$ and $\mathcal{F}^\bullet * \mathcal{G}^n$ are naturally equivalent – they are not even numerically equivalent! We're just relating one coefficient of $F(R(x))$ to one coefficient of $F'(u)G^n(u)$.

First think about defining a function from $\mathcal{F}[\mathcal{R}]_n^\bullet$ to $(\mathcal{F}^\bullet * \mathcal{G}^n)_n$. Let φ be a forest of RLTs with vertex–set $\{1, 2, \ldots, n\}$, and let w be a vertex of φ. The corresponding $\sigma = (A, v, B_1, \ldots, B_n)$ will be defined below (eventually). Notice that since we require that $m(\sigma) = M(\varphi)$, this gives us a big hint as to how to construct σ from (φ, w). Since we must have $\#A = \#\varphi$, let A be the set of root vertices of the components of φ, and let v be the root vertex of the component of φ which contains w. To define the sets B_i is a little bit tricky, and this is where the clever combinatorics in this proof comes into play.

We'll start by defining a function BFS from $\mathcal{F}[\mathcal{R}]_n$ to $(\mathcal{F} * \mathcal{G}^n)_n$ using the following algorithm, which makes use of a list L and a (first–in first–out) queue Q.

```
FUNCTION:  BFS from F[R]_n to (F * G^n)_n;
INPUT:  φ;
initially L and Q are empty, and i := 1;
let A be the set of root vertices of the components of φ;
put the vertices in A on the list L in ascending numerical order;
repeat while L is not empty:
  copy the first vertex of L onto Q;
  delete the first vertex from L;
  repeat while Q is not empty:
    let C_i be the set of children of the first vertex of Q;
    put the vertices of C_i onto Q in ascending numerical order;
    increment i ← i + 1;
    delete the first vertex from Q;
  end repeat;
end repeat;
OUTPUT:  (A, C_1, ..., C_n).
```

What this is doing is breadth–first search on each component of φ, taking the components in ascending order of root labels, and recording the set of children of each vertex. (I suggest that you run the algorithm by hand on an arbitrary example with 15 vertices and three components.) Observe that since each vertex in $\{1, 2, \ldots, n\}$ is deleted from the queue exactly once, a sequence of n sets $(C_1, \ldots, C_n)$ is produced. Also notice that if $(B_1, \ldots, B_n)$ is a listing of $(C_1, \ldots, C_n)$ in any order and $\sigma = (A, v, B_1, \ldots, B_n)$, then $m(\sigma) = M(\varphi)$ as required.

From the output $(A, C_1, \ldots, C_n)$ we can recover the forest φ by the following inverse algorithm, again using a list L and a queue Q.

```
FUNCTION:  FOREST from a subset of (F * G^n)_n to F[R]_n;
INPUT:  (A, C_1, ..., C_n);
initially φ has vertices N_n and no edges;
initially L and Q are empty, and i := 1;
put the vertices in A on the list L in ascending numerical order;
repeat while L is not empty:
  copy the first vertex of L onto Q;
  delete the first vertex from L;
  mark the first vertex of Q as a root vertex of φ;
  repeat while Q is not empty:
    join the first vertex of Q to each vertex of C_i by an edge of φ;
```

```
    put the vertices of C_i onto Q in ascending numerical order;
    increment i ← i + 1;
    delete the first vertex from Q;
  end repeat;
end repeat;
OUTPUT:  φ.
```

The algorithm $FOREST$ is not well–defined on all of $(\mathcal{F} * \mathcal{G}^n)_n$. It is amusing and instructive to find an example input with $n = 10$, say, which causes the algorithm to malfunction. Nonetheless, it is not difficult to verify that for any $\varphi \in \mathcal{F}[\mathcal{R}]_n$, we have

$$FOREST(BFS(\varphi)) = \varphi.$$

In order to construct the bijection for proving LIFT, it will help to understand exactly which $(n+1)$–tuples in $(\mathcal{F} * \mathcal{G}^n)_n$ are in the image of the function BFS.

For this, let's start with the construction of the sets $(C_1, \ldots, C_k)$ for just the first tree T_1 of φ (the one with the smallest root label). Here, k is the number of vertices of T_1. This corresponds to the first pass through the outer `repeat` loop in the algorithm defining BFS. For each $1 \leq i \leq k$, let $c_i := \#C_i - 1$. Notice that after the i–th vertex has been deleted from the queue Q, the number of vertices remaining on Q is $1 + c_1 + \cdots + c_i$; this is true initially (the case $i = 0$) and the elements of C_i are appended to Q just before the i–th vertex is deleted from Q. So the sequence $(c_1, \ldots, c_k)$ satisfies the following conditions:

- each entry is an integer $c_i \geq -1$;
- if $1 \leq i < k$ then the partial sum $c_1 + \cdots + c_i$ is nonnegative; and
- $c_1 + \cdots + c_k = -1$.

Such a sequence is called a *simple Raney sequence*. We shall also say that a sequence $(C_1, \ldots, C_k)$ of sets is a *simple Raney sequence* when $(\#C_1 - 1, \ldots, \#C_k - 1)$ is. As a result, we see that if $(A, C_1, \ldots, C_n) = BFS(\varphi)$ for some $\varphi \in \mathcal{F}[\mathcal{R}]_n$, then $(C_1, \ldots, C_n)$ is the concatenation of $\#A$ simple Raney sequences.

Lemma 13.2. *Let $(b_1, \ldots, b_k)$ be a sequence of integers such that each $b_i \geq -1$ and $b_1 + \cdots + b_k = -1$. Then there is exactly one cyclic shift of the b_i which is a simple Raney sequence.*

Proof. For $0 \leq i \leq k$, let $s_i := b_1 + \cdots + b_i$, so that $s_0 := 0$ and $s_k := -1$. Let s^* denote the minimum of $\{s_0, \ldots, s_k\}$, and let j be the first index at which $s_j = s^*$. The cyclic shift $c_i := b_{i+j}$ (subscripts modulo k for $1 \leq i \leq k$) of the b_i is easily seen to be a simple Raney sequence. (See Figure 13.1.) Any other cyclic shift of the b_i is seen to have some partial sum which is at most -2, and hence fails to be a simple Raney sequence. □

Definition 13.3. A finite sequence $(c_1, \ldots, c_n)$ of integers is an *r–fold Raney sequence* provided that the following conditions hold:

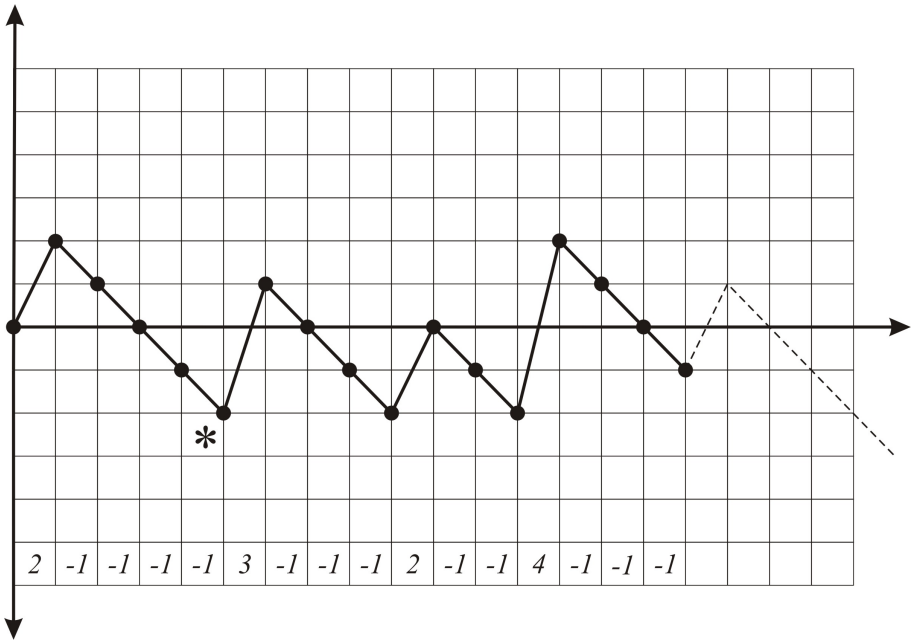


FIGURE 13.1. the unique cyclic shift of Lemma 13.2.

• each entry is an integer $c_i \geq -1$;
• if $1 \leq i < n$ then $c_1 + \cdots + c_i > -r$; and
• $c_1 + \cdots + c_k = -r$.
We shall also say that a sequence $(C_1, \ldots, C_n)$ of sets is an *r–fold Raney sequence* when $(\#C_1 - 1, \ldots, \#C_k - 1)$ is.

It is easy to see that the concatenation of r simple Raney sequences is an r–fold Raney sequence. In part, the following lemma asserts that the converse is also true. (The proof is left as an exercise.)

Lemma 13.4. *Let $\theta = (c_1, \ldots, c_n)$ be an r–fold Raney sequence, for some $r \geq 1$.*
(a) *Then θ has a unique expression as the concatenation $\theta = \rho_1 \cdots \rho_r$ of r simple Raney sequences, called the* blocks *of θ.*
(b) *A cyclic shift of θ is an r–fold Raney sequence if and only if it is obtained by a cyclic shift of the blocks of θ; that is, $\rho_{j+1} \cdots \rho_r \rho_1 \cdots \rho_j$ for some $0 \leq j \leq r-1$.*
(c) *Let $\beta = (b_1, \ldots, b_n)$ be a sequence of integers $b_i \geq -1$ such that $b_1 + \cdots + b_n = -r$. Then there are exactly r cyclic shifts of β which are r–fold Raney sequences.*

Lemma 13.4 gives us enough leverage to finish the proof of LIFT. The idea is to define $(B_1, \ldots, B_n)$ to be a cyclic shift of $(C_1, \ldots, C_n)$ which, along with v, encodes the choice of the root vertex w of (φ, w).

To see this, consider any rooted forest $(\varphi, w) \in \mathcal{F}[\mathcal{R}]^\bullet_n$. Let v be the root of the tree containing w, and construct $(A, C_1, \ldots, C_n)$ by applying the algorithm BFS to

φ. We have seen that $(C_1, \ldots, C_n)$ is an r–fold Raney sequence. Let $\rho_1\rho_2\cdots\rho_r$ be the factorization of this r–fold Raney sequence into its blocks, corresponding to the components of φ. It remains to decide which of the sets C_i is to become the first set B_1 – then the cyclic shift taking the C_i–s to the B_i–s is determined, and we have constructed $\sigma = (A, v, B_1, \ldots, B_n)$.

Now, if v is the s–th vertex of A in ascending numerical order ($1 \leq s \leq r$) then consider the block ρ_s. If w is the p–th vertex (in ascending numerical order) of the s–th component T_s (in ascending order of root labels) of φ, then B_1 is chosen to be the p–th set in the s–th block of $(C_1, \ldots, C_n)$.

Conversely, given $\sigma = (A, v, B_1, \ldots, B_n)$ we must decide which of the sets B_i to choose for the first set C_1. Given $(B_1, \ldots, B_n)$, thought of as a cyclic list of sets, let $\rho_1\rho_2\cdots\rho_r$ be the cyclic list of simple Raney sequences forming its block decomposition. If v is the s–th vertex of A then choose the indexing of these blocks so that B_1 is in the block ρ_s. Now let C_1 be the first set in ρ_1. This determines $(C_1, \ldots, C_n)$ and by applying the algorithm $FOREST$ we construct φ. Now if B_1 is the p–th set in the block ρ_s, let w be the p–th vertex of the s–th component of φ.

The constructions given above provide a pair of mutually inverse bijections between the sets $\mathcal{F}[\mathcal{R}]_n^\bullet$ and $(\mathcal{F}^\bullet * \mathcal{G}^n)_n$, such that if (φ, w) corresponds to σ then $M(\varphi) = m(\sigma)$. There are a number of details to check – that the hypotheses of the lemmas are satisfied when required, that the algorithms always terminate properly on the given input, that they produce the expected output, etc. – but these are left to the diligent reader. This completes the combinatorial proof of LIFT. □

13. Exercises.

1. Prove Lemma 13.4.

13. Endnotes.

This proof of LIFT is a variation of that given by Raney:

• G.N. Raney, *Functional composition patterns and power series reversion*, Trans. Amer. Math. Soc. **94** (1960), 441–451.

There are multivariate versions of LIFT as well. For a state–of–the–art version and some pointers to prior literature, see

• I.P. Goulden and D.M. Kulkarni, *Multivariable Lagrange inversion, Gessel–Viennot cancellation, and the matrix tree theorem*, J. Combin. Theory Ser. A **80** (1997), 295–308.

14. Enumeration and Symmetry.

Another type of enumeration problem – to which neither ordinary nor exponential generating functions apply directly – involves enumerating "symmetry classes" of combinatorial objects. (In fact, some of the theory of exponential generating functions can be adapted to these problems, but we will not develop the subject that far.) Before giving the general theory, let's look at a simple example of the kind of problem involved. (It is rather arbitrary but does illustrate the issues which arise in more interesting examples.)

Imagine a 3–by–3 square array of dots:

$$\begin{array}{ccc} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{array}$$

Such an array can be rotated by 90, 180, or 270 degrees to yield the same array. Also, there are four possible reflections (in the horizontal, vertical, and two diagonal axes) which leave the shape of the array unchanged. Together with the identity function this gives a set of eight symmetries of the diagram. This set D_4 of symmetries is in fact a group, called the *dihedral group of degree 4*.

Now we allow ourselves to colour each of the dots in the array either *filled* $\bullet$ or *open* $\circ$. For each $0 \leq k \leq 9$ we may ask for the number of ways of colouring these dots filled or open so that exactly k of the dots are filled. The answer is easy – the number of choices is $\binom{9}{k}$ since we are merely designating a k–element subset of the 9–element set of dots. The more subtle question we address in this section is the following: for each $0 \leq k \leq 9$, how many **inequivalent** ways are there to colour exactly k of these dots filled? We consider two colourings to be equivalent if one can be changed into the other by applying some symmetry in the dihedral group D_4. For example, when $k = 0$ there is clearly only one such colouring, since all the dots must be open. When $k = 1$ there are three inequivalent colourings:

$$\begin{array}{ccc} \bullet & \circ & \circ \\ \circ & \circ & \circ \\ \circ & \circ & \circ \end{array} \qquad \begin{array}{ccc} \circ & \bullet & \circ \\ \circ & \circ & \circ \\ \circ & \circ & \circ \end{array} \qquad \begin{array}{ccc} \circ & \circ & \circ \\ \circ & \bullet & \circ \\ \circ & \circ & \circ \end{array}$$

as is easily seen. The case $k = 2$ can also be handled readily by a case analysis, with the result that there are eight inequivalent colourings. For the cases $k = 3$ and $k = 4$ the case analysis becomes rather tedious, and we are motivated to find a more systematic procedure. (The cases with $5 \leq k \leq 9$ can be reduced to the cases with $0 \leq k \leq 4$ by the additional symmetry which exchanges the colouring $\bullet \leftrightarrow \circ$ of all the dots.)

We will return to this example in a little while, but now we introduce the general hypotheses and solution to such problems. Given is a finite set Ω of "combinatorial objects" and a subgroup G of the group $\mathcal{S}_\Omega$ of all permutations of Ω. This group G is regarded as the group of "symmetries" underlying the set Ω. (In the example above, Ω is the set of ways to colour the diagram with k filled dots, and G is the dihedral group D_4 acting on this set of coloured diagrams.) The following terminology is essential. For each $g \in G$ we let

$$\mathrm{fix}(g) := \{v \in \Omega : \ g(v) = v\}$$

be the *fixed–point set of* g. For each $v \in \Omega$ we let

$$\mathrm{stab}(v) := \{g \in G : \ g(v) = v\}$$

be the *stablilizer group of* v. (It is not difficult to verify that $\mathrm{stab}(v)$ really is a subgroup of G.) For each $v \in \Omega$ we let

$$\mathrm{orb}(v) := \{g(v) : \ g \in G\}$$

be the *orbit of* v *under* G. Two objects $v, v' \in \Omega$ are related by some symmetry in G – that is $g(v) = v'$ for some $g \in G$ – if and only if $\mathrm{orb}(v) = \mathrm{orb}(v')$. The orbits of Ω under G are thus precisely the equivalence classes of objects in Ω under the group of symmetries G. Thus, the general enumeration problem we wish to solve in this section is to determine the number of orbits of Ω under the group G.

The theorem which solves this problem is known as Burnside's Lemma, although it was first proved by Frobenius. First we need two easy lemmas.

Lemma 14.1. *Let Ω be a finite set and let G be a subgroup of $\mathcal{S}_\Omega$. For any $v \in \Omega$ and $v' \in \mathrm{orb}(v)$ there are exactly $\#\mathrm{stab}(v)$ elements $g \in G$ such that $g(v) = v'$.*

Proof. Let $G(v, v')$ be the set of elements $g \in G$ such that $g(v) = v'$. Since $v' \in \mathrm{orb}(v)$ we know that $G(v, v')$ is not empty. Let $h \in G(v, v')$. Consider the following function:

$$\begin{aligned} \mathrm{stab}(v) &\to G(v, v') \\ g &\mapsto h \circ g \end{aligned}$$

This is a bijection and the inverse function is given by

$$\begin{aligned} G(v, v') &\to \mathrm{stab}(v) \\ g &\mapsto h^{-1} \circ g \end{aligned}$$

Therefore $\#G(v, v') = \#\mathrm{stab}(v)$, as was to be shown. □

Lemma 14.2. *Let Ω be a finite set and let G be a subgroup of $\mathcal{S}_\Omega$. For any $v \in \Omega$ we have*

$$(\#\mathrm{stab}(v))(\#\mathrm{orb}(v)) = \#G.$$

Proof. Consider any $v \in \Omega$, and define a function $F : G \to \mathrm{orb}(v)$ by letting $F(g) := g(v)$ for all $g \in G$. By Lemma 14.1 we have $\#F^{-1}(v') = \#\mathrm{stab}(v)$ for all $v' \in \mathrm{orb}(v)$. Proposition 1.3 now implies the result. □

Theorem 14.3 (Burnside's Lemma)**.** *Let Ω be a finite set and let G be a subgroup of $\mathcal{S}_\Omega$. The number of orbits of Ω under G is*

$$\frac{1}{\#G}\sum_{g\in G}\#\mathrm{fix}(g),$$

the average number of fixed–points among all elements of G.

Proof. We calculate that

$$\begin{aligned}\frac{1}{\#G}\sum_{g\in G}\#\mathrm{fix}(g) &= \frac{1}{\#G}\sum_{g\in G}\sum_{v\in\mathrm{fix}(g)}1 = \frac{1}{\#G}\sum_{v\in\Omega}\sum_{g\in\mathrm{stab}(v)}1\\ &= \sum_{v\in\Omega}\frac{\#\mathrm{stab}(v)}{\#G} = \sum_{v\in\Omega}\frac{1}{\#\mathrm{orb}(v)},\end{aligned}$$

using Lemma 14.2. Now each orbit U of Ω under G has $\#U$ elements, and each element of U contributes $1/\#U$ to this summation. Therefore, the total contribution of each orbit of Ω under G to this sum is 1. That is, this sum gives the total number of orbits of Ω under G, as desired. □

Example 14.4. Let's return to the case $k = 3$ of the example with which we began this section. Thus, Ω is the set of patterns with three filled dots, and $G = D_4$ is the dihedral group of degree 4. We may now do a case analysis on the symmetries $g \in D_4$ to determine their numbers of fixed points. For $g = \iota$ the identity permutation, all $\binom{9}{3} = 84$ patterns are fixed. For the rotations by 90 or 270 degrees, there are no patterns in Ω left fixed by these rotations. For the rotation by 180 degrees there are 4 patterns left fixed. For the reflections in the horizontal or vertical axes there are 10 patterns left fixed. For the reflections in the two diagonal axes there are also 10 patterns left fixed. The average number of fixed patterns is therefore

$$\frac{1}{8}\left(84+0+0+4+10+10+10+10\right) = \frac{128}{8} = 16.$$

By Burnside's Lemma, this is the number of orbits of Ω under G – that is, there are 16 inequivalent ways to fill in three of the dots in that 3–by–3 array. (The case $k = 4$ is left as an exercise.)

Example 14.5. Consider an arrangement of 12 dots in a regular dodecagon. We fix a positive integer k and allow ourselves to colour each of the dots with one of k colours. Taking rotational symmetries of the diagram into account, in how many inequivalent ways can this be done? To solve this we can apply Burnside's Lemma

with Ω being the set of functions $f : \mathbb{Z}_{12} \to N_k$ and $G = \mathbb{Z}_{12}$. A group element $[b] \in G$ acts on an object $f \in \Omega$ by the rule

$$([b]f)([x]) := f([x + b])$$

for all $[x] \in \mathbb{Z}_{12}$. (After a little thought you will see that this corresponds to the intuitive notion of "rotating" a colouring of the dots by b steps.) We must calculate the average number of fixed objects among all elements of G. For example, consider the group element $[9] \in \mathbb{Z}_{12}$. The set $\mathbb{Z}_{12}$ splits up into three cycles of length four under the action of this group element: $\{[0], [9], [6], [3]\}$, $\{[1], [10], [7], [4]\}$, and $\{[2], [11], [8], [5]\}$. If a function $f : \mathbb{Z}_{12} \to N_k$ is to be left fixed by the action of $[9]$, that is, if $[9]f = f$, then we must have $([9]f)([x]) = f([x + 9]) = f([x])$ for all $[x] \in \mathbb{Z}_{12}$. That is, f must be constant on each cycle of $[9]$ acting on $\mathbb{Z}_{12}$. Since there are three cycles and k choices of colour for the dots in each cycle, the group element $[9]$ has k^3 fixed objects in Ω. Now we repeat this argument for each group element $[b] \in \mathbb{Z}_{12}$. The general pattern is given below, but this little table can be worked out by case analysis.

g	$[0]$	$[1]$	$[2]$	$[3]$	$[4]$	$[5]$	$[6]$	$[7]$	$[8]$	$[9]$	$[10]$	$[11]$
#(cycles)	12	1	2	3	4	1	6	1	4	3	2	1

Here, #(cycles) denotes the number of cycles of the corresponding group element acting on $\mathbb{Z}_{12}$. In conclusion, we find that the number of rotationally inequivalent ways of colouring the circular arrangement of 12 dots with k colours is

$$\frac{1}{12}(k^{12} + k^6 + 2k^4 + 2k^3 + 2k^2 + 4k).$$

Generalizing the previous example, let's consider the case of a finite cyclic group acting on itself.

Proposition 14.6. *Fix a positive integer n and an element $[b]$ of $\mathbb{Z}_n$. The permutation $[x] \mapsto [x + b]$ acting on $\mathbb{Z}_n$ has exactly $\gcd(b, n)$ cycles, each of length $n/\gcd(b, n)$.*

Proof. Let $d := \gcd(b, n)$. From MATH 135, recall that the linear Diophantine equation $bs + nt = d$ has a solution $s, t \in \mathbb{Z}$. Therefore $d \equiv bs \pmod{n}$. That is, d and 0 are in the same cycle of $\mathbb{Z}_n$ under the element $[b]$. This cycle consists of the elements $[d], [2d], \ldots, [md] = [n] = [0]$ in some order, in which $m = n/d$. All the other cycles of $[b]$ acting on $\mathbb{Z}_n$ are of the form $\{[j], [d+j], [2d+j], \ldots, [(m-1)d+j]\}$ for some $1 \leq j \leq d - 1$, as is easily checked. This proves the claim. □

Proposition 14.6 provides the key to the following theorem, the proof of which is left as an exercise.

Theorem 14.7. *Let n and k be positive integers, and let Ω be the set of functions $f : \mathbb{Z}_n \to N_k$. Under the action of $\mathbb{Z}_n$ on Ω, the number of orbits is*

$$\frac{1}{n} \sum_{d|n} \varphi\left(\frac{n}{d}\right) k^d,$$

in which $d|n$ denotes that d is a positive divisor of n, and $\varphi(\cdot)$ is the Euler totient function.

---

14. Exercises.

---

1. Solve the case $k = 4$ of Example 14.4.

---

2. Consider an m-by-m square array of dots, each of which can be coloured with one of k colours. Relative to the dihedral group D_4 of symmetries, count the inequivalent colourings of this figure.

---

3. Prove Theorem 14.7.

4(a) For a positive integer n, the *dihedral group* D_n can be thought of as acting on $\mathbb{Z}_n$ with the following symmetries: for each $[b] \in \mathbb{Z}_n$ there is a "rotation" $[x] \mapsto [x+b]$ and a "reflection" $[x] \mapsto [b - x]$. Show that this set D_n is a group (*i.e*, that it is closed under functional composition).

(b) Fix positive integers n and k, and let Ω be the set of functions $f : \mathbb{Z}_n \to N_k$. Derive a formula for the number of orbits of the dihedral group D_n acting on Ω.

---

5. Imagine a regular cube Q in three–dimensional Euclidean space.

(a) Show that Q has exactly 24 rotational symmetries (including the identity). (Hint: first show that for a nontrivial rotation, the axis of rotation must be a line through the center of the cube and either a vertex, the center of an edge, or the center of a face.)

(b) For each positive integer k, derive a formula for the number of inequivalent ways of colouring the faces of Q with k colours, up to rotational symmetries of Q.

(c) For each positive integer k, derive a formula for the number of inequivalent ways of colouring the vertices of Q with k colours, up to rotational symmetries of Q.

(d) For each positive integer k, derive a formula for the number of inequivalent ways of colouring the edges of Q with k colours, up to rotational symmetries of Q.

---

6. Let $\mathcal{A}$ be a natural class of structures (as in Section 12). For each finite set X define

$$\widetilde{\mathcal{A}}_X := \{(\alpha, \sigma) : \ \alpha \in \mathcal{A}_X \text{ and } \sigma \in \mathrm{aut}(\alpha)\}.$$

(a) Show that this defines a natural subclass of $\mathcal{A}\&\mathcal{S}$.

(b) For each $n \in \mathbb{N}$, let a_n be the number of orbits of $\mathcal{S}_n$ acting on $\mathcal{A}_n$. Show that the exponential generating function of $\widetilde{\mathcal{A}}$ is

$$\widetilde{A}(x) = \sum_{n=0}^{\infty} a_n x^n,$$

the ordinary generating function for the numbers a_n.

---

14. Endnotes.

For further development of the theory of enumeration and symmetry, see

• G. Pólya and R.C. Read, "Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds," Springer–Verlag, New York, 1987.

15. Conway's Checker–Jumping Game.

The topic of this section is an amusing but not terribly important factoid in the realm of Recreational Mathematics. Nonetheless, it is instructive in that it illustrates the usefulness of thinking "outside the box". Plus, what the heck, I think it's a lot of fun.

John Horton Conway, world–famous group theorist and generally brilliant guy, is also well–known for making up mathematical games and puzzles. This is the kind of thing that sometimes occurs to him....

Imagine the integer plane $\mathbb{Z} \times \mathbb{Z}$ as representing an infinite checker board by visualizing a unit square centered at each point $(a, b) \in \mathbb{Z} \times \mathbb{Z}$. We allow ourselves to start the game by putting at most one checker on each of the squares with $b \leq 0$, that is, in the lower half–plane

$$H := \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : \ b \leq 0\}.$$

Thereafter, the checkers may be jumped over one another in the following manner. Find a strip of three (horizontally or vertically) consecutive squares, in which the center square and one of the end squares are occupied by checkers while the other end square is vacant. Then move the end checker to the other (vacant) end, and remove the center checker. See Figure 15.1 for an illustration of this.

Now we set ourselves the following hierarchy of goals. For each $g \in \mathbb{N}$, the *g–th level game* is to begin with enough checkers in the lower half–plane ($b \leq 0$) so that, by some sequence of checker jumps, we may finish with a checker on the *g–th goal square* $(0, g)$. If we can do this then we win, but we will also want to minimize the number $N(g)$ of checkers required at the beginning. If we cannot find a starting position which enables us to reach the goal square then we lose and $N(g)$ is not defined. (I leave it up to you to decide on the stakes for which we are playing.)

Before reading on, please pause for a minute and think about the function $N(g)$. What kind of rate of growth do you expect it to have?

The 0–th level game is trivial, since we may begin with a checker on the 0–th goal square $(0, 0)$. That is, we win and $N(0) = 1$.

The 1–st level game is very easy, since we may begin with checkers on the squares $(0, 0)$ and $(0, -1)$. One jump then puts a checker on the 1–st goal square $(0, 1)$, so that we win and $N(1) = 2$.

The 2–nd level game is easy, since we may begin with checkers on the squares $(0, 0)$, $(0, -1)$, $(1, 0)$, and $(2, 0)$. A sequence of three jumps then puts a checker on the 2–nd goal square $(0, 2)$, so that we win and $N(2) = 4$. (It is easy to see that $N(2) > 3$.)

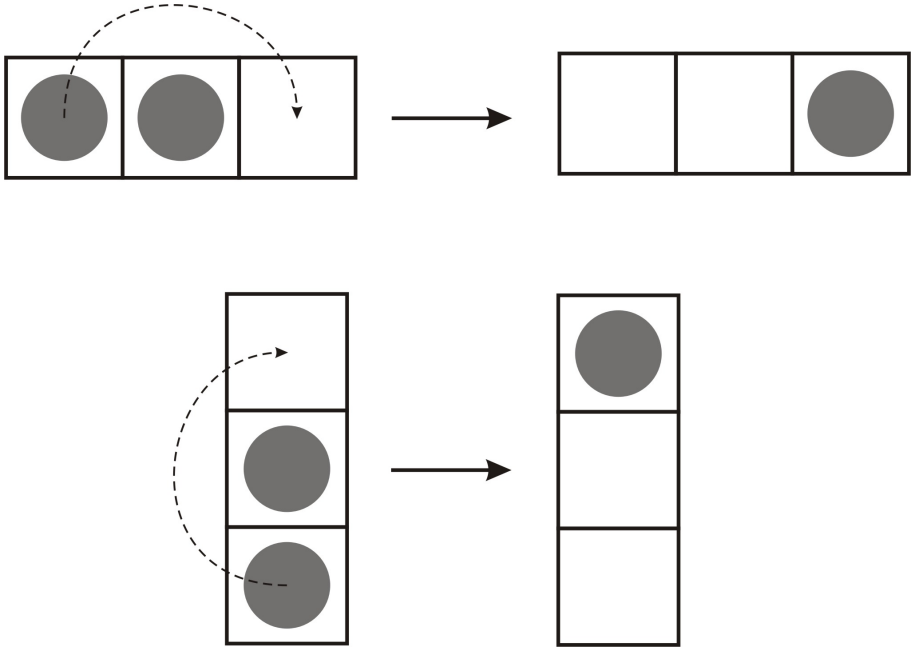


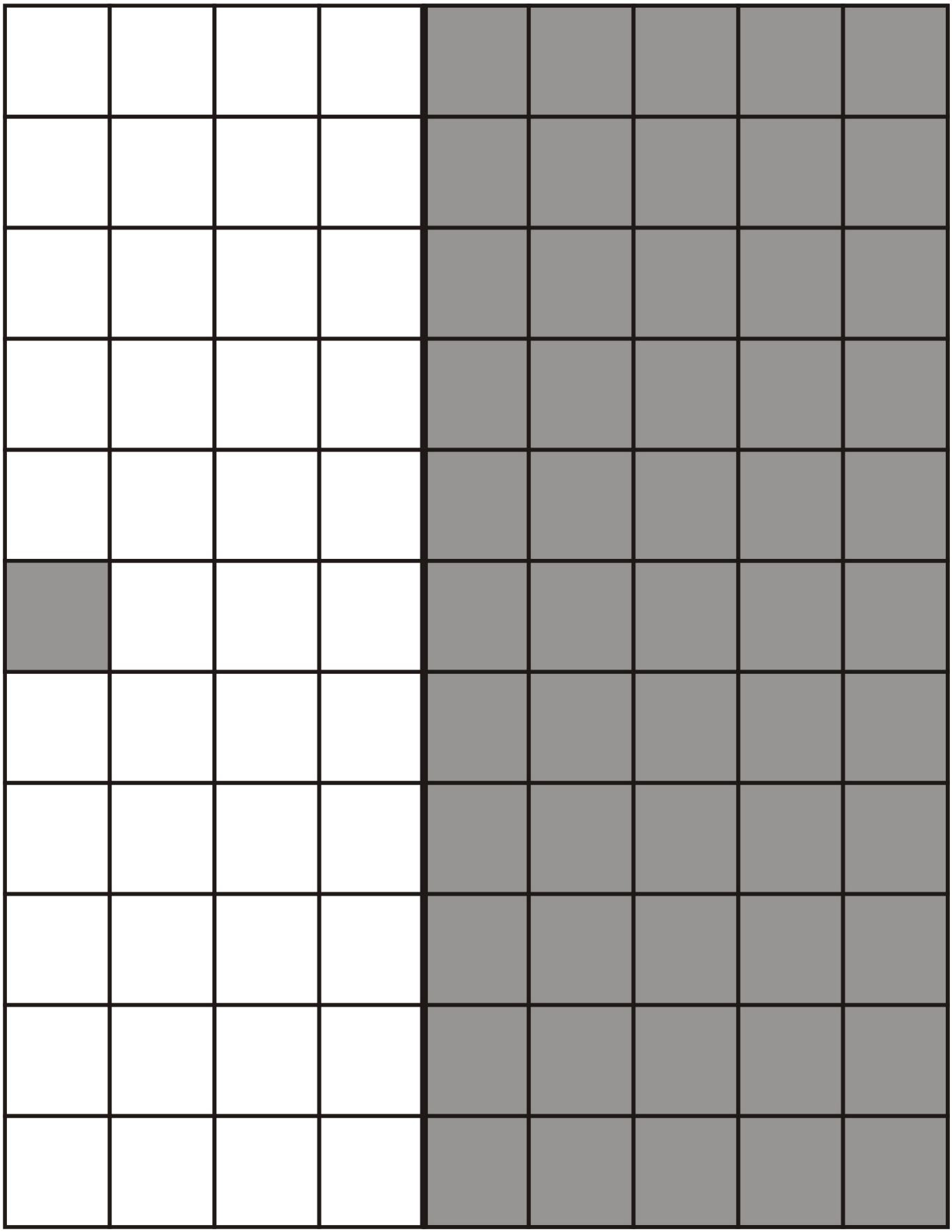
FIGURE 15.1. the checker–jumping rules.

The 3–rd level game is starting to get interesting, but I'll still tell you the answer. We may start with checkers on the squares $(-2,0)$, $(-1,0)$, $(0,0)$, $(1,0)$, $(2,0)$, $(-2,-1)$, $(-1,-1)$, and $(0,-1)$. A sequence of seven jumps then puts a checker on the 3–rd goal square $(0,3)$, so that we win and $N(3) \leq 8$. In this case, it is not so obvious that $N(3) = 8$, but this is indeed true. I leave it up to you to convince yourself that no position starting with at most seven checkers can reach the third level.

So far, the pattern $N(g) = 2^g$ holds true for $0 \leq g \leq 3$. Are you confident enough to conjecture that this pattern holds for all $g \in \mathbb{N}$?

The 4–th level game is interesting. I won't give it all away, but we win and $N(4) \leq 20$. Please pause for a while and search for a solution – one with 21 checkers is relatively easy to find. On the facing page the lower half–plane and the 4–th goal square are shaded. Get some pennies and play with it for a little while – perhaps you can find a solution using fewer than 20 checkers. If you used a simple case analysis proof above to show that $N(3) > 7$, then you will quite likely feel that a similar analysis to derive a lower bound for $N(4)$ would be formiddably unpleasant. Fortunately, after examining the 5–th level game we will return to this question with some new ideas.

The 5–th level game is extremely interesting! I encourage you to stop reading now and experiment with it before turning the page....



If you tried playing the 5–th level game before reading this, then no doubt you are feeling frustrated. Or perhaps you have deluded yourself into thinking you found a solution. I am confident in saying this, because **we lose the 5–th level game**! Of course, this means that we lose the g–th level game for all $g \geq 5$. This is somewhat disappointing, but also rather interesting. Having searched for a solution, and having failed to find one, you may be inclined to accept this assertion. Such bald assertions hold little credence in mathematics (except, perhaps, in the *Comptes Rendus*) so we must supply a proof. But how?

Here is the nice idea of Conway's which makes the proof possible. We are going to define a real–valued "value function" V on the finite subsets of $\mathbb{Z} \times \mathbb{Z}$ with the following properties:
(i) The value of the 5–th goal square is one: $V(\{(0,5)\}) = 1$.
(ii) If S is a finite subset of the lower half–plane ($b \leq 0$) then $V(S) < 1$.
(iii) If S' is obtained from S by one checker jump, then $V(S') \leq V(S)$.
Certainly, if such a value function exists, then we lose the 5–th level game. The reason is simple – we begin in a position with value strictly less than one and our position can never increase in value. Therefore, we can never reach the goal position, which has value one.

To define this value function, we begin by defining the "weight" of a square $(a,b) \in \mathbb{Z} \times \mathbb{Z}$ to be its (rectangular) distance to the 5–th goal square $(0,5)$; that is,

$$\omega(a,b) := |a| + |b-5|.$$

Now, for any subset $S \subseteq \mathbb{Z} \times \mathbb{Z}$, we may consider the ordinary generating function for S with respect to ω:

$$\Phi_S(x) := \sum_{(a,b)\in S} x^{\omega(a,b)}.$$

See Figure 15.2 for an illustration – the 5–th goal square and the lower half–plane H are shaded in the figure. Here is the point at which we must think outside the box, and do something which I have repeatedly told you not to do. We are going to substitute a particular real value ξ for the indeterminate x. (Heresy!) This will be done so that the function $V(S) := \Phi_S(\xi)$ has the three properties required above.

Condition (i) will be satisfied for any value of ξ, since $\Phi_{\{(0,5)\}}(x) = 1$ identically.

To ensure condition (ii), let us first compute the generating function of the lower half–plane $H = \{(a,b) \in \mathbb{Z} \times \mathbb{Z} : \ b \leq 0\}$. We find that

$$\begin{aligned} \Phi_H(x) &= \sum_{a=-\infty}^{\infty} \sum_{b=-\infty}^{0} x^{|a|+|b-5|} = \left(\sum_{a=-\infty}^{\infty} x^{|a|}\right)\left(\sum_{c=5}^{\infty} x^c\right) \\ &= \left(\frac{1+x}{1-x}\right)\left(\frac{x^5}{1-x}\right) = \frac{x^5+x^6}{(1-x)^2}. \end{aligned}$$

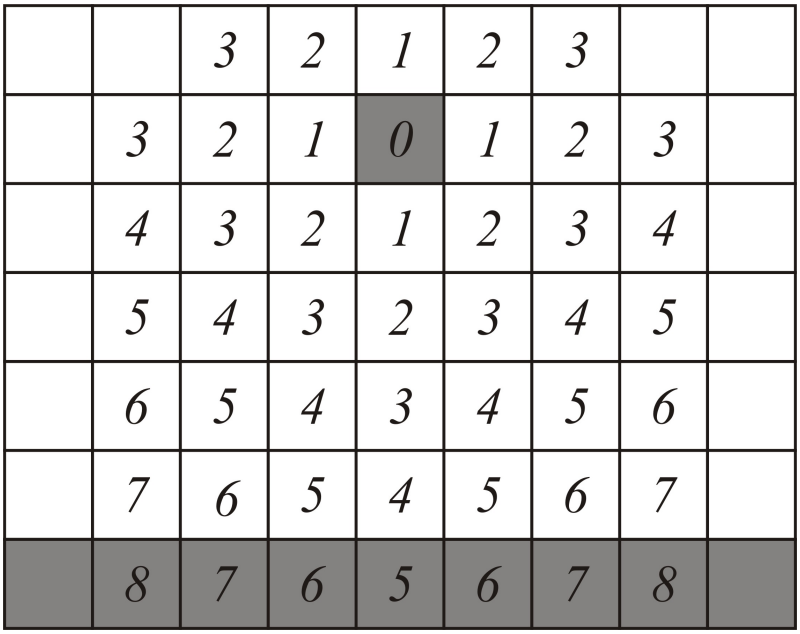


FIGURE 15.2. the weights of the squares of $\mathbb{Z} \times \mathbb{Z}$.

As a power series involving a real variable x, this is convergent as long as $|x| < 1$. Therefore, to ensure condition (ii) we need a real number ξ such that $0 < \xi < 1$ and

$$V(H) = \frac{\xi^5 + \xi^6}{(1-\xi)^2} \leq 1.$$

This suffices for condition (ii) because, for any finite subset $S \subset H$, the fact that $0 < \xi$ implies that $V(S) < V(H) \leq 1$.

Condition (iii) is the heart of the matter. Assume that the finite subsets S and S' of $\mathbb{Z} \times \mathbb{Z}$ are such that S' is obtained from S by one checker jump. Consider the difference $\Delta(x) := \Phi_S(x) - \Phi_{S'}(x)$. Since S and S' agree everywhere except on one (horizontal or vertical) strip of three consecutive squares, we may focus our attention on this strip where they differ. There are three cases, depending upon the location of this strip on the checkerboard $\mathbb{Z} \times \mathbb{Z}$: for some $k \in \mathbb{N}$ we must have one of

(a) $\Delta(x) = x^{k+2} + x^{k+1} - x^k$,
(b) $\Delta(x) = x^{k+1} + x^k - x^{k+1}$, or
(c) $\Delta(x) = x^k + x^{k+1} - x^{k+2}$.

We need to find a real value $0 < \xi < 1$ such that $\Delta(\xi) \geq 0$ in each of these cases.

Since $0 < \xi$, we also have $0 < \xi^k$, so that case (b) is trivial. Since $0 < \xi < 1$, we also have $\xi^{k+1} > \xi^{k+2}$, so that case (c) is satisfied as well. To ensure condition (iii) for the value function, all that remains is to require that $0 < \xi < 1$ satisfies the

inequality in case (a); that is (after factoring out $\xi^k > 0$),

$$\xi^2 + \xi - 1 \geq 0.$$

The values of ξ which attain this with equality are found by the Quadratic Formula:

$$\xi_{\pm} = \frac{-1 \pm \sqrt{5}}{2}.$$

Since the leading coefficient of the quadratic function is positive, the inequality is satisfied if and only if either $\xi \leq (-1 - \sqrt{5})/2$ or $\xi \geq (-1 + \sqrt{5})/2$. The first choice contradicts $0 < \xi$, but the second choice is possible: we are now restricted to considering real values such that $(-1 + \sqrt{5})/2 \leq \xi < 1$. Any such value will ensure condition (iii).

We still need to verify the inequality

$$V(H) = \frac{\xi^5 + \xi^6}{(1-\xi)^2} \leq 1$$

in order to ensure that condition (ii) holds. At this point a minor miracle occurs! The particular (extremal) value $\xi := (-1+\sqrt{5})/2$ has met all the requirements so far, and it also satisfies the equation $\xi^2 + \xi - 1 = 0$. Therefore, $\xi^5 + \xi^6 = \xi^4(\xi + \xi^2) = \xi^4$, and also $(1-\xi)^2 = (\xi^2)^2 = \xi^4$. That is, for this value of ξ we have

$$V(H) = \frac{\xi^5 + \xi^6}{(1-\xi)^2} = \frac{\xi^4}{\xi^4} = 1,$$

satisfying the last of the requirements needed to prove that the value function $V(S) := \Phi_S(\xi)$ meets all of the conditions (i), (ii), and (iii). The existence of this value function suffices to prove that we lose the 5–th level game, as described above.

At this point we have a satisfactory understanding of Conway's Checker Jumping Game, except for the 4–th level game. We know that $N(4) \leq 20$ (if you have also found a solution!) but so far we do not have a lower bound for $N(4)$. The ideas in the above impossibility proof for the 5–th level game can be applied here as well, though.

Let us redefine the weight of the square $(a, b) \in \mathbb{Z} \times \mathbb{Z}$ to be

$$\omega(a, b) := |a| + |b - 4|,$$

the (rectangular) distance to the 4–th goal square. Then the generating function for the lower half–plane H is

$$\Phi_H(x) = \frac{x^4 + x^5}{(1-x)^2},$$

as is easily checked. If we substitute the particular value $x = \xi = (-1 + \sqrt{5})/2$ then conditions (i) and (iii) in the impossibility proof for the 5–th level game are still satisfied for the value function $V(S) := \Phi^{\omega}_S(\xi)$. Condition (ii) no longer holds.

Instead, we obtain a condition on a finite subset $S \subset H$ which is necessary if we are to be able to reach the 4–th goal square starting from S – it is necessary that $V(S) \geq 1$.

Expanding the generating function for H in powers of x we have

$$\Phi_H(x) = x^4 + 3x^5 + 5x^6 + 7x^7 + 9x^8 + \cdots .$$

Since $0 < \xi < 1$, if $S \subset H$ and $\#S \leq 19$, then

$$V(S) \leq \xi^4 + 3\xi^5 + 5\xi^6 + 7\xi^7 + 3\xi^8.$$

Using the fact that $\xi^2 + \xi = 1$, we can simplify this as follows:

$$\begin{aligned} V(S) &\leq \xi^4 + 3\xi^5 + 5\xi^6 + 7\xi^7 + 3\xi^8 \\ &= \xi^4 + 3\xi^5 + 8\xi^6 + 4\xi^7 \\ &= \xi^4 + 7\xi^5 + 4\xi^6 \\ &= 5\xi^4 + 3\xi^5 = 3\xi^3 + 2\xi^4 \\ &= 2\xi^2 + \xi^3 = \xi + \xi^2 = 1. \end{aligned}$$

From this we conclude that $N(4) > 18$, and that the positions starting with 19 checkers which could possibly win the 4–th level game are few in number. We are left with a small ambiguity: that is, $N(4) \in \{19, 20\}$. The case analysis required to resolve this issue is, however, too lengthy to pursue here.

---

15. Exercises.

1(a) Show that $N(4) \leq 21$.
(b)* Show that $N(4) \leq 20$.

---

2. Let's play the same game in any number of dimensions. Show that for any $d \in \mathbb{N}$ there is a $g(d) \in \mathbb{N}$ such that we lose the $g(d)$–th level game on $\mathbb{Z}^d$.

---

3. Let's change the rules of the game slightly. We're back in dimension 2, but we also allow checker jumps along the diagonals. In this case, $N(0) = 1$, $N(1) = 2$, $N(2) = 3$, $N(3) = 5$, $N(4) = 8$, and so on.

(a) Find a value of $g \in \mathbb{N}$ such that we lose the g–th level of this version of the game.

(b) Give a lower bound for $N(g-1)$.

---

4. Let's change the rules of the game slightly in a different way. We're in dimension 2 and only allow horizontal and vertical jumps (as in the original game), but we allow ourselves to stack up to k checkers on any square in $\mathbb{Z} \times \mathbb{Z}$ which is at distance k from the goal square, for each $k \in \mathbb{N}$. (We may stack these checkers initially if the square is in H, and also at any intermediate stage of the sequence of checker jumps.) In this case, $N(0) = 1$, $N(1) = 2$, $N(2) = 3$, $N(3) = 5$, $N(4) \leq 10$, and so on.

(a) Find a value of $g \in \mathbb{N}$ such that we lose the g–th level of this version of the game.

(b) Give a lower bound for $N(g-1)$.

---

5. Let's relax the rules of the game in Question 3 even further. Fix a "capacity function" $C : \mathbb{N} \to \mathbb{N}$, and allow a stack of at most $C(k)$ checkers on any square at distance k from the goal square (at any stage of the game). Let's say that a capacity function C is *victorious* if it is possible to win the g–th level game for every $g \in \mathbb{N}$. Show that there is a constant $\theta > 1$ such that if C is a victorious function then $C(k) > \theta^k$ for infinitely many $k \in \mathbb{N}$. What is the largest value of θ which you can find with this property?

---

6. With the terminology of Question 5, either find a victorious capacity function $C : \mathbb{N} \to \mathbb{N}$ or prove that no such function exists.

---

15. Endnotes.

I first learned of Conway's Checker–Jumping Game while in highschool, by reading one of Martin Gardner's "Mathematical Games" columns which used to appear in *Scientific American*. Many of these columns have been anthologized in his books, all of which are highly entertaining. He did not present the impossibility proof for the 5–th level game, which I found many years later in

• R. Honsberger, "Mathematical Gems II," Math. Assoc. of America, Washington, DC, 1976.

Honsberger has written several books in this vein – rather like Gardner's columns but with more complete mathematical derivations. Again, they are all very entertaining and I recommend them highly.

While we're on the subject of general mathematical reading, I urge you to peruse the classic four–volume tome

- J.R. Newman, "The World of Mathematics," Simon & Schuster, New York, 1956.

This is a wonderful must–have book for anyone who likes mathematics!

I am straying off topic now, but I also have to recommend these two books. The first is a more formal mathematical presentation – the second is a whole bunch of fun.

- J.H. Conway, "On numbers and games" London Mathematical Society Monographs, **6**, Academic Press, London–New York, 1976.

- E.R. Berlekamp, J.H. Conway, and R.K. Guy, "Winning ways for your mathematical plays," Academic Press, London–New York, 1982.

16. The Matrix–Tree Theorem.

In Section 11 we saw two proofs of the fact that there are n^{n-2} trees with vertex–set $N_n = \{1, 2, \ldots, n\}$. This can be interpreted as the number of spanning trees of the complete graph K_n with vertex–set N_n. It is natural to ask for a formula for the number of spanning trees of other graphs G. After a little thought you will realize that the theory of exponential generating functions is ill–equipped to address this problem. However, there is a beautiful solution by other means.

Let $G = (V, E)$ be a finite graph which may contain loops or multiple edges. We denote by $\varkappa(G)$ the number of spanning trees of G, sometimes called the *complexity* of G. If G is not connected then $\varkappa(G) = 0$, so we will assume that G is connected from now on. If G' is obtained from G by removing all the loops of G, then $\varkappa(G') = \varkappa(G)$, since a loop can never occur in a spanning tree. Thus, we may assume that G contains no loops as well. Multiple edges, however, do remain a possibility.

The quantity $\varkappa(G)$ can be computed recursively, using the easily proven formula

$$\varkappa(G) = \varkappa(G \smallsetminus e) + \varkappa(G/e),$$

in which $G \smallsetminus e$ is the graph obtained from G by deleting the edge e and G/e is the graph obtained from G by contracting the edge e (and removing any loops produced). Furthermore, if G and H are connected graphs which intersect in exactly one vertex then

$$\varkappa(G \cup H) = \varkappa(G) \cdot \varkappa(H),$$

as is also easily seen. See Figure 16.1 for an example computation using this method. (For each graph in the figure, an edge which is deleted/contracted is marked with an asterisk.) This leads to an algorithm which in general takes an exponential amount of time to complete, and so it is not suitable for large computations. For some simple kinds of graphs this *deletion–contraction reduction* does provide a useful recurrence – Exercise 16.2 is one example.

Fortunately, there is an easily–computable and general formula for $\varkappa(G)$, which we now derive. To state the formula we need to define some matrices. The *adjacency matrix* of G is the square matrix $A = A(G)$ indexed by $V \times V$, which has as its entries: $A_{vv} := 0$ for all $v \in V$, and if $v \neq w$ in V then A_{vw} is the number of edges of G which have vertices v and w at their ends. The degree $\deg_G(v)$ of a vertex $v \in V$ of G is the number of edges of G which are incident with v. The *degree matrix* of G is the diagonal V–by–V matrix $\Delta = \Delta(G)$ such that $\Delta_{vv} := \deg_G(v)$ for all $v \in V$, and $\Delta_{vw} := 0$ if $v \neq w$. Finally, the *Laplacian matrix* of G is defined to be $Q(G) := \Delta(G) - A(G)$.

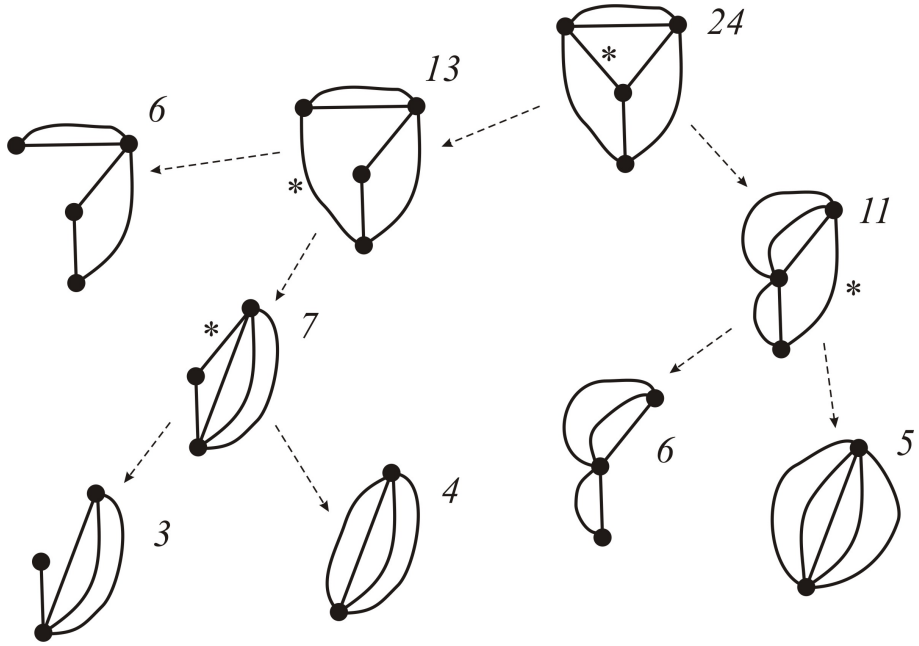


FIGURE 16.1. computing $\varkappa(G)$ by deletion/contraction.

One more piece of notation is required. If M is a matrix and i is a row–index for M and j is a column–index for M, let $M(i|j)$ denote the submatrix of M obtained by deleting row i and column j from M.

Theorem 16.1 (The Matrix–Tree Theorem)**.** *Let $G = (V, E)$ be a connected graph with no loops. Then for any $v \in V$,*

$$\varkappa(G) = \det Q(v|v).$$

That is a fantastic formula! Now... how to prove it?

One could proceed by induction on the number of edges, by showing that the RHS of the formula satisfies the same deletion–contraction reduction as does $\varkappa(G)$. (The initial conditions forming the base case of the induction are easily checked.) However, there is a more informative proof. The essential piece of technology for this proof is a famous determinantal identity – the Binet–Cauchy Formula. Part of the point I want to make is that Linear Algebra does not stop after MATH 235 – there are many interesting and useful facts to pick up one by one as they are required in various subjects. This is one of those occasions.

Let M be an n–by–m matrix, and let P be an m–by–n matrix. The product MP is then a square n–by–n matrix, so its determinant is defined. The Binet–Cauchy Formula expresses this determinant in terms of the factors M and P. Since these are not necessarily square we cannot take their determinants *per se*, so something a bit more complicated is going on. A little more notation is needed to state

the formula. Let M be a matrix, let I be a set of row–indices of M, and let J be a set of column–indices of M. Generalizing the above convention, we let $M(I|J)$ denote the submatrix of M obtained by deleting the rows in I and the columns in J from M. Also, we let $M[I|J]$ denote the submatrix of M obtained by deleting the rows **not** in I and the columns **not** in J from M. The other two possibilities, $M(I|J]$ and $M[I|J)$, are interpreted accordingly. In particular, $M(|J]$ indicates that we use all rows of M but only the columns of M in the set J.

Theorem 16.2 (The Binet–Cauchy Formula)**.** *Let M be an n–by–m matrix, and let P be an m–by–n matrix. Then*

$$\det(MP) = \sum_S \det(M(|S])\det(P[S|))$$

in which the sum is over all n–element subsets of the column indices of M (which are the same as the row indices of P).

Proof. We proceed by induction on n. For the basis of induction, $n = 1$, M is a row vector and P is a column vector of length m. The result follows immediately in this case from the definition of matrix product. For the induction step, assume that the result is true with $n-1$ in place of n.

Consider the Laplace expansion of $\det(MP)$ along row i, for some $1 \le i \le n$. That is,

$$\det(MP) = \sum_{j=1}^{n} (-1)^{i+j} (MP)_{ij} \det((MP)(i|j))$$

We average these among all $1 \le i \le n$ to obtain

$$\begin{aligned}
\det(MP) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(-1)^{i+j}(MP)_{ij}\det((MP)(i|j)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(-1)^{i+j}\left(\sum_{k=1}^{m} M_{ik}P_{kj}\right)\det(M(i|)P(|j)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(-1)^{i+j}\left(\sum_{k=1}^{m} M_{ik}P_{kj}\right)\sum_{U}\det(M(i|U])\det(P[U|j)).
\end{aligned}$$

In the second of these equalities we use the fact that $(MP)(i|j) = M(i|)P(|j)$, and in the third we use the induction hypothesis. The final summation is over all $(n-1)$–element subsets U of the column indices of M.

Continuing with the calculation, we have

$$\begin{aligned} & \det(MP) \\ = \; & \frac{1}{n}\sum_{U}\sum_{k=1}^{m}\left(\sum_{i=1}^{n}(-1)^{i+k}M_{ik}\det(M(i|U])\right)\left(\sum_{j=1}^{n}(-1)^{k+j}P_{kj}\det(P[U|j))\right) \\ = \; & \frac{1}{n}\sum_{U}\sum_{k=1}^{m}(-1)^{n-k}\det(M(|U,k])(-1)^{n-k}\det(P[U,k|)) \end{aligned}$$

In the last of these lines, $M(|U,k]$ is the matrix obtained from $M(|U]$ by adjoining the column $M(|k]$ on the right side. Laplace expansion along the last column shows that

$$(-1)^{n-k}\det(M(|U,k]) = \sum_{i=1}^{n}(-1)^{i+k}M_{ik}\det(M(i|U]).$$

Similarly, $P[U,k|)$ is the matrix obtained from $P[U|)$ by adjoining the row $P[k|)$ on the bottom. Laplace expansion along the last row shows that

$$(-1)^{n-k}\det(P[U,k|)) = \sum_{j=1}^{n}(-1)^{k+j}P_{kj}\det(P[U|j)).$$

Of course, if $k \in U$ then the matrix $M(|U,k]$ has two equal columns, so that $\det(M(|U,k)] = 0$. Thus, in the two outer summations we may restrict attention to pairs (U,k) such that $k \notin U$. In effect, this sums over all n–element subsets $S = U \cup \{k\}$ of the column indices of M and counts each one n times. Since the number of column–exchanges needed to obtain $M(|S]$ from $M(|U,k]$ is equal to the number of row–exchanges needed to obtain $P[S|)$ from $P[U,k|)$, we see that

$$\det(M(|U,k])\det(P[U,k|)) = \det(M(|S])\det(P[S|)).$$

Continuing with the calculation, we conclude that

$$\det(MP) = \sum_{S}\det(M(|S])\det(P[S|))$$

which completes the induction step, and the proof. □

Now consider a connected graph $G = (V,E)$ with no loops, and orient each edge of G arbitrarily by putting an arrow on it pointing towards one of its two ends. The *signed incidence matrix* of G (with respect to this orientation) is the V–by–E indexed matrix D with entries

$$D_{ve} := \begin{cases} 1 & \text{if } v \text{ is at the head of } e, \\ -1 & \text{if } v \text{ is at the tail of } e, \\ 0 & \text{otherwise.} \end{cases}$$

The proof of the Matrix–Tree Theorem is completed by the following two lemmas, the proofs of which are left as exercises.

Lemma 16.3. *Let $G = (V, E)$ be a connected graph with no loops, orient G arbitrarily, and let D be the corresponding signed incidence matrix. Then $DD^\dagger = Q(G)$, in which $D^\dagger$ denotes the transpose of D.*

Lemma 16.4. *Let $G = (V, E)$ be a connected graph with no loops, orient G arbitrarily, and let D be the corresponding signed incidence matrix. Let $v \in V$ and let $S \subseteq E$ be such that $\#S = \#V - 1$. Then $\det(D(v|S]) \in \{-1, 1\}$ if (V, S) is a spanning tree of G, and $\det(D(v|S]) = 0$ otherwise.*

The Matrix–Tree Theorem may be "lifted" from a conclusion about the cardinality of the set of spanning trees of $G = (V, E)$ to a conclusion about the generating function for this set. To define this generating function, let $\mathbf{y} := \{y_e : e \in E\}$ be pairwise commuting indeterminates, and for $S \subseteq E$ let $\mathbf{y}^S := \prod_{e \in S} y_e$. We may identify a spanning subgraph of G with its edge–set, and define

$$T(G; \mathbf{y}) := \sum_{T \in \mathfrak{T}(G)} \mathbf{y}^T,$$

in which the sum is over the set $\mathfrak{T}(G)$ of all spanning trees of G.

Theorem 16.5 (The Weighted Matrix–Tree Theorem)**.** *Let $G = (V, E)$ be a connected graph with no loops, orient G arbitrarily, and let D be the corresponding signed incidence matrix. Let $Y := \mathrm{diag}(y_e : e \in E)$ be the E–by–E diagonal matrix of the indeterminates $\mathbf{y}$. Then, for any $v \in V$,*

$$T(G; \mathbf{y}) = \det(D(v|)YD^\dagger(|v)).$$

The proof is just as for Theorem 16.1, and is left as an exercise. The matrix $Q(G; \mathbf{y}) := DYD^\dagger$ is known as the *weighted Laplacian matrix* of G, and does not depend on the choice of orientation in the definition of D.

Theorem 16.5 can be generalized another step, as follows. Theorem 16.6 will be used in the next section.

Theorem 16.6 (The Principal Minors Matrix–Tree Theorem)**.** *Let $G = (V, E)$ be a connected graph with no loops, orient G arbitrarily, and let D be the corresponding signed incidence matrix. Let $Y := \mathrm{diag}(y_e : e \in E)$ be the E–by–E diagonal matrix of the indeterminates $\mathbf{y}$. For any subset $S \subseteq V$ of vertices,*

$$\det(Q(G; \mathbf{y})(S|S)) = \det(D(S|)YD^\dagger(|S)) = \sum_F \mathbf{y}^F,$$

in which the sum is over all spanning forests F of G for which each component of F contains exactly one vertex of S.

Proof. The key ideas are just the same as for the proof of Theorem 16.1. The main novelty consists in showing that for $S \subseteq V$ and $F \subseteq E$ such that $\#S + \#F = \#V$, the determinant $\det D(S|F]$ is ± 1 if and only if (V, F) is a spanning forest of G as in

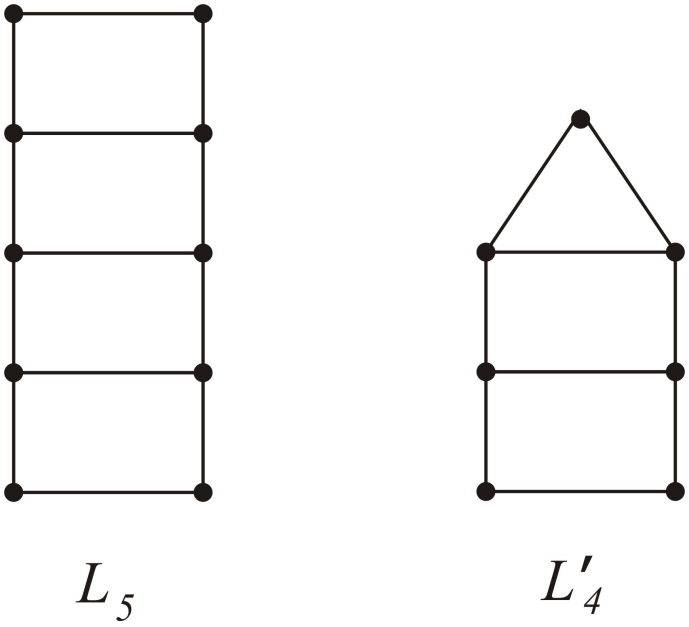


FIGURE 16.2. the graphs L_5 and L'_4.

the statement, and is zero otherwise. After this the Binet–Cauchy Formula finishes the proof as before. □

---

16. Exercises.

1(a) Prove that for a graph $G = (V, E)$ and $e \in E$,

$$T(G, \mathbf{y}) = T(G \setminus e; \mathbf{y}) + y_e T(G/e; \mathbf{y}).$$

(b) Prove that if G and H intersect in exactly one vertex then

$$T(G \cup H, \mathbf{y}) = T(G, \mathbf{y}) T(H, \mathbf{y}).$$

---

2. For $n \geq 1$, let L_n denote the *ladder graph* illustrated in Figure 16.1 with n "rungs". Let L'_n denote L_n with one of its end–rungs contracted.

(a) Show that $\varkappa(L_1) = \varkappa(L'_1) = 1$ and for $n \geq 2$,

$$\begin{cases} \varkappa(L_n) &= \varkappa(L_{n-1}) + \varkappa(L'_n), \\ \varkappa(L'_n) &= 2\varkappa(L_{n-1}) + \varkappa(L'_{n-1}). \end{cases}$$

That is,

$$\begin{bmatrix} \varkappa(L_n) \\ \varkappa(L'_n) \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} \varkappa(L_{n-1}) \\ \varkappa(L'_{n-1} \end{bmatrix}.$$

(b) Write the matrix in part (a) in the form $U^{-1}\Lambda U$ with Λ a diagonal matrix.

(c) Use part (b) to solve the recurrence in part (a) to obtain formulas for $\varkappa(L_n)$ and $\varkappa(L'_n)$.

---

3. Prove Lemma 16.3.

---

4. Prove Lemma 16.4.

---

5. Prove Theorems 16.1 and 16.5.

---

6. Prove Theorem 16.6.

---

7. Let $G = (V, E)$ be a graph, number the vertices $V = \{v_1, v_2, \ldots, v_n\}$, and write down the weighted Laplacian matrix $Q = DYD^\dagger$ of G with respect to this ordering of the vertices. Show that for all $1 \leq i, j \leq n$,

$$\det(Q(v_i|v_j)) = (-1)^{i+j} T(G; \mathbf{y}).$$

---

8. Use Theorem 16.1 to prove that $\varkappa(K_n) = n^{n-2}$ for all $n \geq 1$.

---

9. The *characteristic polynomial* of a graph G is defined to be $\chi(G; x) := \det(xI - A)$, in which A is the adjacency matrix of G. (The roots of this polynomial are the eigenvalues of A.)

(a) Prove that if G is connected and k–regular, then k is an eigenvalue of A of multiplicity one, and the all–ones vector is an eigenvector of eigenvalue k.

(b) Prove that if G is connected and k–regular, then $\varkappa(G) = |\chi'(G; k)|$. [Hint: $\det(M)$ is the product of the eigenvalues of M.]

---

16. Endnotes.

Theorem 16.6 was first proved by Kirchhoff in 1847:

• G. Kirchhoff, *Über die Auflösung der Gleichungen, auf welche man bei der Untersuchungen der linearen Vertheilung galvanischer Ströme geführt wird,* Ann. Phys. Chem. **72** (1847), 497-508.

All the minors of Q can be interpreted as (signed) generating functions of sets of forests in G:

• S. Chaiken, *A combinatorial proof of the all minors matrix tree theorem,* SIAM J. Algebraic Discrete Methods **3** (1982).

If this combination of algebra and graph theory appeals to you, check out these two books:

• N.L. Biggs, "Algebraic Graph Theory. Second Edition," Cambridge U.P., Cambridge, 1993.

• C.D. Godsil and G.F. Royle, "Algebraic Graph Theory," Graduate Texts in Mathematics **207**, Springer–Verlag, New York, 2001.

17. Kirchhoff's Effective Admittance Formula.

The material of this section properly belongs to the subject matter of Electrical Engineering. That said, the central result – Kirchhoff's formula for the effective admittance of a (linear) electrical network – has direct enumerative content. I will give two proofs: the classical one based on Cramer's Rule, and one of my own devising which uses linear algebra in a quite different way (and remains closer to the underlying combinatorial structures throughout).

The situation is as follows. We are given a finite undirected graph $G = (V, E)$ (which may contain loops or multiple edges), and each edge $e \in E$ is assigned an electrical resistance r_e. Given two distinct vertices $a, b \in V$, we pass an electric current through the graph G by attaching the vertices a and b to the poles of an external current source. By measuring the difference in voltage (or electric potential) between the vertices a and b we can then determine the effective resistance of the network G connecting the terminals a and b. All of this is physically quite reasonable – the truly lovely thing is that the calculation can be made *a priori* and depends entirely on the enumerative combinatorics of the graph G.

First of all, it turns out that the formula is more naturally expressed in terms of conductance rather than resistance. (Conductance is merely the reciprocal of resistance.) Second, resistance and conductance are conventionally real–valued quantities, but Kirchhoff's formula remains valid for quantities taken from any given field. The field $\mathbb{F} := \mathbb{C}(s)$ is particularly important for circuits to which a time–varying source of current is applied. (The variable s is the Laplace transform of the time variable.) In this case, the analogue of resistance is referred to as *impedance*, and the analogue of conductance is referred to as *admittance*. The linear algebra really doesn't care what field of quantities we use, so most of this linguistic hairsplitting is moot. However, it is worth the effort to explain the terminology.

In order to derive Kirchhoff's Formula we need to specify the behaviour of an electrical network precisely. This is accomplished by Ohm's Law, Kirchhoff's Current Law, and Kirchhoff's Voltage Law. All three are physically intuitive and we do not dwell on their justifications.

Ohm's Law: In a wire e with ends v and w, the current j_e flowing through e from v to w is directly proportional to the difference in electric potential $\varphi(v) - \varphi(w)$ between the ends. The constant of proportionality is the admittance y_e of the wire e. That is, $j_e = y_e(\varphi(v) - \varphi(w))$.

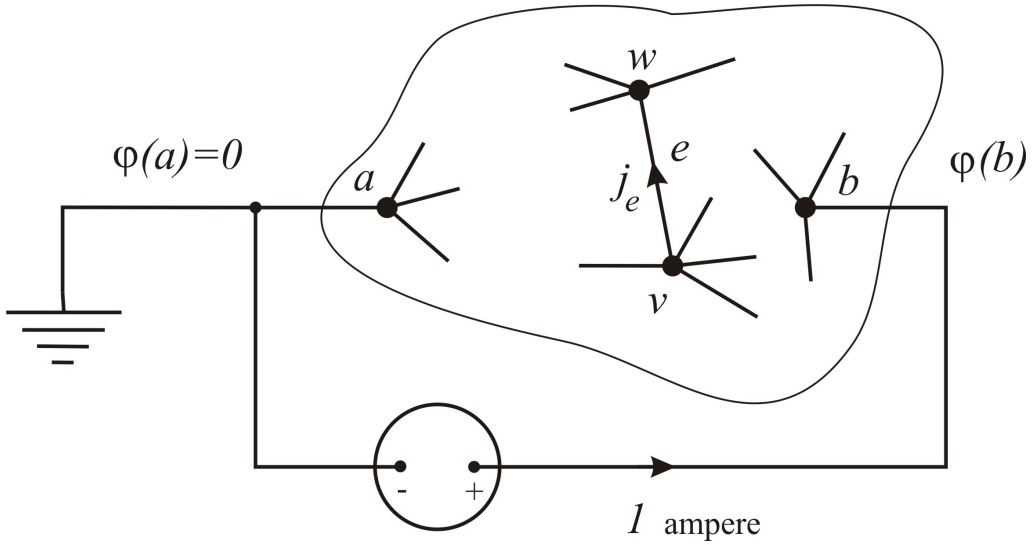


FIGURE 17.1. how to measure $\mathcal{Y}_{ab}(G;\mathbf{y})$.

Kirchhoff's Current Law: In an electrical network $G = (V, E, \mathbf{y})$, at every vertex v the amount of current flowing in equals the amount of current flowing out.

Kirchhoff's Voltage Law: In an electrical network $G = (V, E, \mathbf{y})$, there is a potential function $\varphi : V \to \mathbb{F}$ such that Ohm's Law is satisfied for every wire $e \in E$.

To measure the effective admittance of an electrical network $G = (V, E, \mathbf{y})$ between the vertices $a, b \in V$ we can do the following. Connect a and b to an external source of current and force one ampere of current through the network from b to a. Ground the vertex a so that its electric potential is $\varphi(a) = 0$. The electric potential $\varphi(b)$ is then inversely proportional to the effective admittance of G, by Ohm's Law. So a calculation of this quantity

$$\mathcal{Y}_{ab}(G;\mathbf{y}) := \frac{1}{\varphi(b)}$$

is what we seek. See Figure 17.1 for an illustration.

Finally, to express Kirchhoff's Formula we recall the generating functions for spanning trees introduced in Section 16. That is, for a graph $G = (V, E)$ and indeterminates $\mathbf{y} := \{y_e : \ e \in E\}$ we defined

$$T(G;\mathbf{y}) := \sum_{T \in \mathcal{T}(G)} \mathbf{y}^T,$$

in which the sum is over the set $\mathcal{T}(G)$ of all spanning trees of G. One more piece of notation is required: for a graph $G = (V, E)$ and vertices $a, b \in V$, we let G/ab

denote the graph obtained by merging the two vertices a and b together into a single vertex.

Theorem 17.1 (Kirchhoff's Formula)**.** *Let $G = (V, E, \mathbf{y})$ be an electrical network, and let $a, b \in V$. The effective admittance between a and b in G is*

$$\mathcal{Y}_{ab}(G; \mathbf{y}) = \frac{T(G; \mathbf{y})}{T(G/ab; \mathbf{y})}.$$

First Proof of Kirchhoff's Formula. We begin by translating the physical "laws" above into linear algebra. To do this we must fix an arbitrary orientation of $G = (V, E)$ by drawing an arrow on each edge of G in one of the two possible directions. (The choice of orientation does not affect the formula, but the orientation is needed in order to write down the equations corresponding to Ohm's Law.) Next we consider the V–by–E incidence matrix D of G with respect to this orientation. This was defined in Section 16 for a graph with no loops. We extend this definition to allow loops by saying that if $e \in E$ is a loop then the corresponding column of D is entirely zero. Let $\mathbf{j} := \{j_e : \ e \in E\}$ be the E–indexed column vector of currents, let $Y := \mathrm{diag}(y_e : \ e \in E)$ be the diagonal matrix of admittances, and let $\boldsymbol{\varphi} := \{\varphi(v) : \ v \in V\}$ be the V–indexed column vector of voltages, normalized so that $\varphi(a) = 0$. We may restate the physical "laws" as follows:

Ohm's Law: $\mathbf{j} = -YD^\dagger\boldsymbol{\varphi}$. This is the statement of Ohm's Law for every wire in the network simultaneously.

Kirchhoff's Current Law: $D\mathbf{j} = \boldsymbol{\delta}_a - \boldsymbol{\delta}_b$. Here $\boldsymbol{\delta}_v$ is the V–indexed column vector given by

$$(\boldsymbol{\delta}_v)_w := \begin{cases} 1 & \text{if } w = v, \\ 0 & \text{if } w \neq v. \end{cases}$$

The reason that the RHS is not zero is that one ampere of current is being supplied to b externally and removed from a externally. The currents internal to the network G must compensate for this external driving current.

Kirchhoff's Voltage Law: A solution $\boldsymbol{\varphi}$ to Ohm's Law exists.

Combining these equations, our task is now to solve the system

$$DYD^\dagger\boldsymbol{\varphi} = \boldsymbol{\delta}_b - \boldsymbol{\delta}_a$$

for $\boldsymbol{\varphi}$. More precisely, we only need to determine the value $\varphi(b)$. Every column of D sums to zero, as does the RHS. Therefore, this system of linear equations is redundant and we can strike out any one of them. Since we have normalized $\varphi(a) = 0$, the a–th column of $D^\dagger$ will not contribute at all to the product $D^\dagger\boldsymbol{\varphi}$.

Thus, let $D_a := D(a|)$ be the matrix obtained from D by deleting row a. We seek a solution to

$$D_a Y D_a^\dagger \boldsymbol{\varphi} = \boldsymbol{\delta}_b,$$

in which $\boldsymbol{\varphi}$ and $\boldsymbol{\delta}_b$ are now column vectors indexed by $V \setminus \{a\}$.

Since we only want the value of $\varphi(b)$, Cramer's Rule is the perfect technique to use. By Theorem 16.5,

$$\det(D_a Y D_a^\dagger) = T(G; \mathbf{y}),$$

and this is nonzero for a generic choice of admittances $\mathbf{y}$, so that the system is invertible. Replacing column b of $D_a Y D_a^\dagger$ by $\boldsymbol{\delta}_b$, we obtain a matrix M with $M_{bb} = 1$ being the only nonzero entry in column b. The Laplace expansion of $\det(M)$ along column b then shows that

$$\det(M) = \det(D(ab|))Y D^\dagger(|ab)).$$

By Theorem 16.6 this determinant is the generating function for spanning forests of G which have exactly two components, one containing a and one containing b. These forests of G correspond bijectively with spanning trees of G/ab, so that $\det(M) = T(G/ab; \mathbf{y})$. Cramer's Rule thus implies that

$$\varphi(b) = \frac{T(G/ab; \mathbf{y})}{T(G; \mathbf{y})}.$$

Since the effective admittance was defined to be $\mathcal{Y}_{ab}(G; \mathbf{y}) := 1/\varphi(b)$, this completes the proof. □

The second proof uses linear algebra in a way quite different from the first proof, and it avoids Cramer's Rule and the Matrix–Tree Theorems by staying closer to the underlying graph theory. At first sight it seems to involve a strange way of thinking, but this kind of strategy is often very useful in combinatorial enumeration problems.

Second Proof of Kirchhoff's Formula. We work with the field $\mathbb{K} := \mathbb{F}(\mathbf{y}) := \mathbb{C}(s, y_e : e \in E)$ of rational functions in the indeterminates s and $\mathbf{y} = \{y_e : e \in E\}$ with complex coefficients. Ohm's Law and Kirchhoff's Laws imply that all the currents $\{j_e : e \in E\}$ and potentials $\{\varphi(v) : v \in V\}$ are in the field $\mathbb{K}$, justifying the calculations to follow.

Recall that $\mathcal{T}(G)$ is the set of spanning trees of G. Let $\mathbb{K}\mathcal{T}(G)$ denote the vector space over $\mathbb{K}$ which has as a basis $\{[T] : T \in \mathcal{T}(G)\}$. That is, a vector in $\mathbb{K}\mathcal{T}(G)$ is a formal linear combination

$$\sum_{T \in \mathcal{T}(G)} c_T [T]$$

in which the coefficients c_T are in $\mathbb{K}$. The vector space $\mathbb{K}\mathcal{T}(G/ab)$ is defined similarly. We are going to define a linear transformation $L : \mathbb{K}\mathcal{T}(G) \to \mathbb{K}\mathcal{T}(G/ab)$, the

properties of which will allow us to prove Kirchhoff's Formula. In order to define L we must fix an arbitrary orientation for each edge $e \in E$, just as in the first proof. For each $e \in E$, let head(e) denote the vertex into which e points, and let tail(e) denote the vertex out of which e points.

It suffices to define the action of L on each basis vector $[T]$ for $T \in \mathcal{T}(G)$. In the spanning tree T of G, there is a unique path $P(T)$ which begins at b and ends at a. For each edge $e \in P(T)$, say that the sign of (T, e) is $\text{sgn}(T, e) := +1$ if e is directed from b to a along $P(T)$, and that $\text{sgn}(T, e) := -1$ of e is directed from a to b along $P(T)$. We define $L([T])$ by the formula

$$L([T]) := \sum_{e \in P(T)} \text{sgn}(T, e) j_e [T \smallsetminus e],$$

in which $\mathbf{j} = \{j_e : \ e \in E\}$ is the vector of currents as in the first proof. Notice that as a set of edges, $T \smallsetminus e$ is a spanning tree of G/ab for every $e \in P(T)$. This action of L is extended linearly to all vectors in $\mathbb{K}\mathcal{T}(G)$.

Now consider the vector $\mathbf{g} := \sum_{T \in \mathcal{T}(G)} [T]$ in $\mathbb{K}\mathcal{T}(G)$ – that is, the formal sum of all spanning trees of G. Then $L(\mathbf{g})$ is some vector in $\mathbb{K}\mathcal{T}(G/ab)$, so that

$$L(\mathbf{g}) = \sum_{Z \in \mathcal{T}(G/ab)} c_Z [Z]$$

for some coefficients $c_Z \in \mathbb{K}$. By the definition of L, we see that

$$c_Z = \sum_{T \in \mathcal{T}(G):\ T \smallsetminus e = Z} \text{sgn}(T, e) j_e.$$

Regarding Z as a forest in G with two components A containing a and B containing b, a spanning tree $T \in \mathcal{T}(G)$ is such that $T \smallsetminus e = Z$ if and only if $T = Z \cup \{e\}$ and e is an edge of G which has one end in A and the other end in B. Let $C(Z)$ be this set of edges of G, so that

$$c_Z = \sum_{e \in C(Z)} \text{sgn}(Z \cup \{e\}, e) j_e.$$

Notice that the signs $\text{sgn}(Z \cup \{e\}, e)$ are such that this sum is the total current flowing from B to A through the wires in the set $C(Z)$. Applying Kirchhoff's Current Law to every vertex of B, and remembering that one ampere of current is entering $b \in B$ from an external source, we see that the total current flowing from B to A through $C(Z)$ is also one ampere. That is, $c_Z = 1$ for every $Z \in \mathcal{T}(G/ab)$, so that

$$L(\mathbf{g}) = \sum_{Z \in \mathcal{T}(G/ab)} [Z].$$

This is an equation between vectors in the vector space $\mathbb{K}\mathcal{T}(G/ab)$.

Next, we define a linear functional $\alpha : \mathbb{K}\mathcal{T}(G/ab) \to \mathbb{K}$ as follows: for $Z \in \mathcal{T}(G/ab)$, let

$$\alpha([Z]) := \mathbf{y}^Z$$

and extend this linearly to all of $\mathbb{K}\mathcal{T}(G/ab)$. We examine the result of applying this linear functional to both sides of the vector equation derived in the previous paragraph. On the RHS we obtain

$$\alpha\left(\sum_{Z\in\mathcal{T}(G/ab)} [Z]\right) = \sum_{Z\in\mathcal{T}(G/ab)} \mathbf{y}^Z = T(G/ab;\mathbf{y}),$$

which is looking good. For the LHS, let's first consider some $T \in \mathcal{T}(G)$ and compute $\alpha(L([T]))$. That is, using Ohm's Law,

$$\begin{aligned}
\alpha(L([T])) &= \sum_{e\in P(T)} \operatorname{sgn}(T,e) j_e \mathbf{y}^{T\smallsetminus e} \\
&= \sum_{e\in P(T)} \operatorname{sgn}(T,e) y_e \left(\varphi(\operatorname{tail}(e)) - \varphi(\operatorname{head}(e))\right) \mathbf{y}^{T\smallsetminus e} \\
&= \mathbf{y}^T \sum_{e\in P(T)} \operatorname{sgn}(T,e) \left(\varphi(\operatorname{tail}(e)) - \varphi(\operatorname{head}(e))\right) \\
&= \mathbf{y}^T \left(\varphi(b) - \varphi(a)\right) = \varphi(b)\mathbf{y}^T,
\end{aligned}$$

since $\varphi(a) = 0$. The penultimate equality follows because the signs $\operatorname{sgn}(T,e)$ are such that each vertex v on $P(T)$ other than a or b will contribute both $+\varphi(v)$ and $-\varphi(v)$ to the summation, so the summation of differences "telescopes". Consequently, we have

$$\alpha(L(\mathbf{g})) = \sum_{T\in\mathcal{T}(G)} \varphi(b)\mathbf{y}^T = \varphi(b)T(G;\mathbf{y}).$$

Comparing this with the RHS we obtain the equation of rational functions

$$T(G/ab;\mathbf{y}) = \varphi(b)T(G;\mathbf{y}).$$

Finally, we conclude that

$$\mathcal{Y}_{ab}(G;\mathbf{y}) = \frac{1}{\varphi(b)} = \frac{T(G;\mathbf{y})}{T(G/ab;\mathbf{y})},$$

which completes the proof. □

17. Exercises.

1. Let L_n be the ladder graph defined in Exercise 16.2. Let a and b be two vertices of L_n which are adjacent accross the k–th rung of L_n. If each edge is given unit admittance, compute the effective admittance of L_n from a to b.

2. Let a and b be any two adjacent vertices of the ladder graph L_n. If each edge is given unit admittance, compute the effective admittance of L_n from a to b.

3. Let $G = (V, E)$ and $H = (W, F)$ be vertex–disjoint graphs. Let $a, b \in V$ and $a', b' \in W$. The *parallel connection* of the networks $(G; a, b)$ and $(H; a', b')$ is the graph $L := (G \cup H)/\{aa', bb'\}$ with terminals $a = a'$ and $b = b'$. Show that

$$\mathcal{Y}_{ab}(L; \mathbf{y}) = \mathcal{Y}_{ab}(G; \mathbf{y}) + \mathcal{Y}_{a'b'}(H; \mathbf{y}).$$

4. Let $G = (V, E)$ and $H = (W, F)$ be vertex–disjoint graphs. Let $a, b \in V$ and $a', b' \in W$. The *series connection* of the networks $(G; a, b)$ and $(H; a', b')$ is the graph $J := (G \cup H)/ba'$ with terminals a and b'. Show that

$$\frac{1}{\mathcal{Y}_{ab'}(J; \mathbf{y})} = \frac{1}{\mathcal{Y}_{ab}(G; \mathbf{y})} + \frac{1}{\mathcal{Y}_{a'b'}(H; \mathbf{y})}.$$

5(a) Use Exercises 3 and 4 to solve Exercise 1.

5(b) Use Exercises 3 and 4 to solve Exercise 2.

6. Let $G = (V, E)$ be a graph, orient G arbitrarily, and let D be the corresponding signed incidence matrix.

(a) Show that for any $v \in V$ and complex vector $\mathbf{z}$ indexed by $V \setminus v$:

$$\mathbf{z}^\dagger D_v Y D_v^\dagger \mathbf{z} = \sum_{e \in E} y_e |(D_v^\dagger \mathbf{z})_e|^2.$$

(Here $D_v := D(v|)$ and $D_v^\dagger$ denotes the conjugate transpose of D_v, etc.)

(b) Deduce that if $\mathrm{Re}(y_e) > 0$ for all $e \in E$ then $T(G, \mathbf{y}) \neq 0$.

(c) Deduce that if $\mathrm{Re}(y_e) > 0$ for all $e \in E$ then $\mathrm{Re}(\mathcal{Y}_{ab}(G, \mathbf{y})) \geq 0$ for all $a, b \in V$. (Hint: consider the graph $\widehat{G} := G \cup \{\widehat{e}\}$ in which $\widehat{e}$ is a new edge with ends a and

b.) This result has a physical interpretation – if every wire in a network dissipates energy, then the whole network dissipates energy.

---

17. Endnotes.

Theorem 17.1 was first proved by Kirchhoff in 1847:

• G. Kirchhoff, *Über die Auflösung der Gleichungen, auf welche man bei der Untersuchungen der linearen Vertheilung galvanischer Ströme geführt wird,* Ann. Phys. Chem. **72** (1847), 497–508.

There is a great deal of interesting mathematics relating to the physical behaviour of electrical networks – graph theory, differential equations, Laplace transformations, and complex analysis all play a part. (For example, in the situation of Exercise 17.6(c), the Open Mapping Theorem of complex analysis implies that either $\mathcal{Y}_{ab}(G, \mathbf{y})$ is a pure imaginary constant or else if $\mathrm{Re}(y_e) > 0$ for all $e \in E$ then $\mathrm{Re}(\mathcal{Y}_{ab}(G, \mathbf{y})) > 0$ for all $a, b \in V$. The first case can not occur since $\mathcal{Y}_{ab}(G; \mathbf{y})$ is a rational function of degree one, so the inequality in the conclusion is actually strict.)

Many textbooks have been written on electrical network theory – not surprisingly, most are intended for electrical engineers. Here are two books with particularly good treatment of the mathematics:

• N. Balabanian and T.A. Bickart, "Electrical Network Theory," Wiley, New York, 1969.

• I. Vágó, "Graph theory : Application to the Calculation of Electrical Networks," Elsevier, Amsterdam, 1985.

Perhaps somewhat surprisingly, there are deep connections between the properties of electrical networks and the properties of random walks on graphs. A good introduction to this subject is

• P.G. Doyle and J.L. Snell, "Random Walks and Electric Networks," Math. Assoc. of America, Washington, DC, 1984.

Perhaps even more surprisingly, electrical network theory can be used to solve the problem of dissecting a square into smaller squares, no two of which are the

same size!

• R.L. Brooks, C.A.B. Smith, A.H. Stone, and W.T. Tutte, *The dissection of rectangles into squares*, Duke Math. J. **7**, (1940). 312–340.

This is one of the most famous papers in graph theory. It is reprinted in

• "Selected Papers of W.T. Tutte," (McCarthy, Stanton, eds.), Charles Babbage Research Center, Winnipeg, 1979.