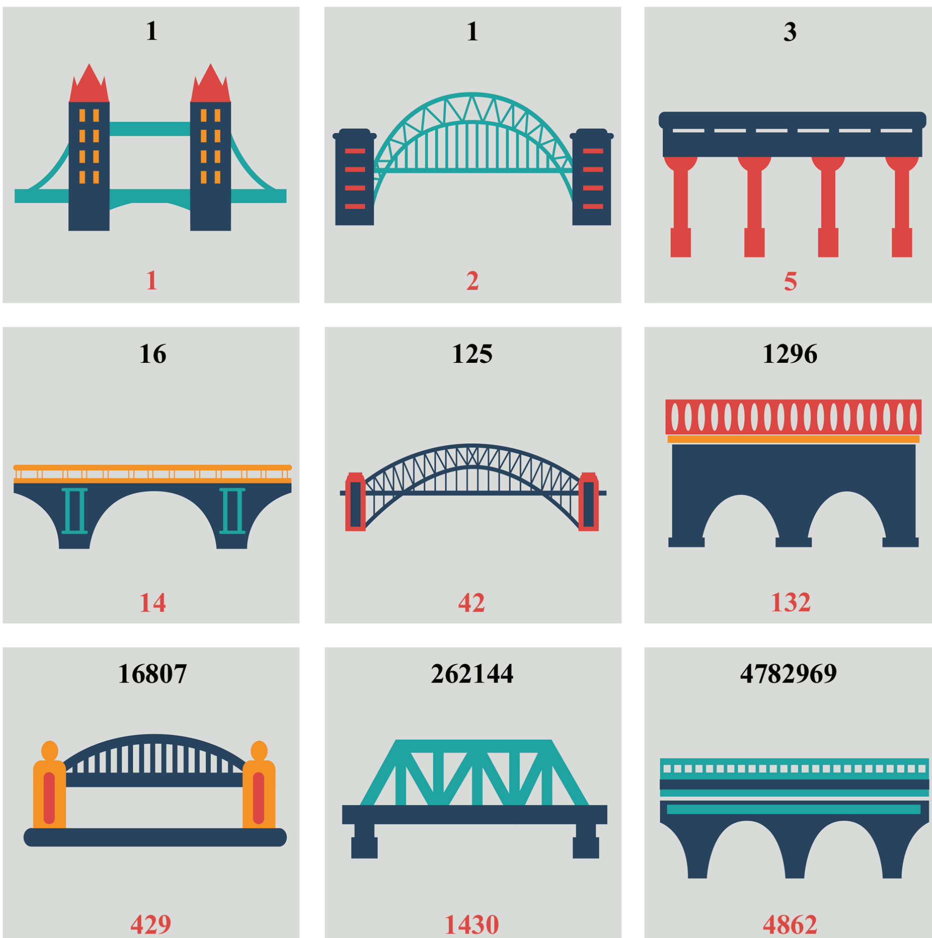
DISCRETE MATHEMATICS AND ITS APPLICATIONS

HANDBOOK OF ENUMERATIVE COMBINATORICS



Edited by
Miklós Bóna

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

HANDBOOK OF
ENUMERATIVE COMBINATORICS

DISCRETE MATHEMATICS AND ITS APPLICATIONS

R. B. J. T. Allenby and Alan Slomson, How to Count: An Introduction to Combinatorics, Third Edition

Craig P. Bauer, Secret History: The Story of Cryptology

Juergen Bierbrauer, Introduction to Coding Theory

Katalin Bimbó, Combinatory Logic: Pure, Applied and Typed

Katalin Bimbó, Proof Theory: Sequent Calculi and Related Formalisms

Donald Bindner and Martin Erickson, A Student's Guide to the Study, Practice, and Tools of Modern Mathematics

Francine Blanchet-Sadri, Algorithmic Combinatorics on Partial Words

Miklós Bóna, Combinatorics of Permutations, Second Edition

Miklós Bóna, Handbook of Enumerative Combinatorics

Jason I. Brown, Discrete Structures and Their Interactions

Richard A. Brualdi and Dragoš Cvetković, A Combinatorial Approach to Matrix Theory and Its Applications

Kun-Mao Chao and Bang Ye Wu, Spanning Trees and Optimization Problems

Charalambos A. Charalambides, Enumerative Combinatorics

Gary Chartrand and Ping Zhang, Chromatic Graph Theory

Henri Cohen, Gerhard Frey, et al., Handbook of Elliptic and Hyperelliptic Curve Cryptography

Charles J. Colbourn and Jeffrey H. Dinitz, Handbook of Combinatorial Designs, Second Edition

Abhijit Das, Computational Number Theory

Matthias Dehmer and Frank Emmert-Streib, Quantitative Graph Theory: Mathematical Foundations and Applications

Martin Erickson, Pearls of Discrete Mathematics

Martin Erickson and Anthony Vazzana, Introduction to Number Theory

Titles (continued)

Steven Furino, Ying Miao, and Jianxing Yin, Frames and Resolvable Designs: Uses, Constructions, and Existence

Mark S. Gockenbach, Finite-Dimensional Linear Algebra

Randy Goldberg and Lance Riek, A Practical Handbook of Speech Coders

Jacob E. Goodman and Joseph O'Rourke, Handbook of Discrete and Computational Geometry, Second Edition

Jonathan L. Gross, Combinatorial Methods with Computer Applications

Jonathan L. Gross and Jay Yellen, Graph Theory and Its Applications, Second Edition

Jonathan L. Gross, Jay Yellen, and Ping Zhang Handbook of Graph Theory, Second Edition

David S. Gunderson, Handbook of Mathematical Induction: Theory and Applications

Richard Hammack, Wilfried Imrich, and Sandi Klavžar, Handbook of Product Graphs, Second Edition

Darrel R. Hankerson, Greg A. Harris, and Peter D. Johnson, Introduction to Information Theory and Data Compression, Second Edition

Darel W. Hardy, Fred Richman, and Carol L. Walker, Applied Algebra: Codes, Ciphers, and Discrete Algorithms, Second Edition

Daryl D. Harms, Miroslav Kraetzl, Charles J. Colbourn, and John S. Devitt, Network Reliability: Experiments with a Symbolic Algebra Environment

Silvia Heubach and Toufik Mansour, Combinatorics of Compositions and Words

Leslie Hogben, Handbook of Linear Algebra, Second Edition

Derek F. Holt with Bettina Eick and Eamonn A. O'Brien, Handbook of Computational Group Theory

David M. Jackson and Terry I. Visentin, An Atlas of Smaller Maps in Orientable and Nonorientable Surfaces

Richard E. Klima, Neil P. Sigmon, and Ernest L. Stitzinger, Applications of Abstract Algebra with Maple™ and MATLAB®, Second Edition

Richard E. Klima and Neil P. Sigmon, Cryptology: Classical and Modern with Maplets

Patrick Knupp and Kambiz Salari, Verification of Computer Codes in Computational Science and Engineering

William Kocay and Donald L. Kreher, Graphs, Algorithms, and Optimization

Donald L. Kreher and Douglas R. Stinson, Combinatorial Algorithms: Generation Enumeration and Search

Hang T. Lau, A Java Library of Graph Algorithms and Optimization

C. C. Lindner and C. A. Rodger, Design Theory, Second Edition

San Ling, Huaxiong Wang, and Chaoping Xing, Algebraic Curves in Cryptography

Nicholas A. Loehr, Bijective Combinatorics

Toufik Mansour, Combinatorics of Set Partitions

Titles (continued)

Alasdair McAndrew, Introduction to Cryptography with Open-Source Software

Elliott Mendelson, Introduction to Mathematical Logic, Fifth Edition

Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone, Handbook of Applied Cryptography

Stig F. Mjølsnes, A Multidisciplinary Introduction to Information Security

Jason J. Molitierno, Applications of Combinatorial Matrix Theory to Laplacian Matrices of Graphs

Richard A. Mollin, Advanced Number Theory with Applications

Richard A. Mollin, Algebraic Number Theory, Second Edition

Richard A. Mollin, Codes: The Guide to Secrecy from Ancient to Modern Times

Richard A. Mollin, Fundamental Number Theory with Applications, Second Edition

Richard A. Mollin, An Introduction to Cryptography, Second Edition

Richard A. Mollin, Quadratics

Richard A. Mollin, RSA and Public-Key Cryptography

Carlos J. Moreno and Samuel S. Wagstaff, Jr., Sums of Squares of Integers

Gary L. Mullen and Daniel Panario, Handbook of Finite Fields

Goutam Paul and Subhamoy Maitra, RC4 Stream Cipher and Its Variants

Dingyi Pei, Authentication Codes and Combinatorial Designs

Kenneth H. Rosen, Handbook of Discrete and Combinatorial Mathematics

Douglas R. Shier and K.T. Wallenius, Applied Mathematical Modeling: A Multidisciplinary Approach

Alexander Stanoyevitch, Introduction to Cryptography with Mathematical Foundations and Computer Implementations

Jörn Steuding, Diophantine Analysis

Douglas R. Stinson, Cryptography: Theory and Practice, Third Edition

Roberto Tamassia, Handbook of Graph Drawing and Visualization

Roberto Togneri and Christopher J. deSilva, Fundamentals of Information Theory and Coding Design

W. D. Wallis, Introduction to Combinatorial Designs, Second Edition

W. D. Wallis and J. C. George, Introduction to Combinatorics

Jiacun Wang, Handbook of Finite State Based Models and Applications

Lawrence C. Washington, Elliptic Curves: Number Theory and Cryptography, Second Edition

DISCRETE MATHEMATICS AND ITS APPLICATIONS

HANDBOOK OF
ENUMERATIVE COMBINATORICS

Edited by

Miklós Bóna

University of Florida
Gainesville, Florida, USA

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20150220

International Standard Book Number-13: 978-1-4822-2086-5 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Foreword

When I first became seriously interested in enumerative combinatorics around 1967, the subject did not really exist *per se*. There were numerous results and methods, beginning with Euler (with some hints of even earlier work), scattered throughout the literature, but there was little systematic attempt to bring some order to this chaos. It is remarkable that enumerative combinatorics has progressed so far that we now need over 1000 pages just to present a basic overview of techniques and results. Numerous topics such as pattern avoidance and parking functions existed in only very rudimentary form 40–50 years ago but are now flourishing subjects in their own right. The seventeen authors of the present volume, in addition to being world leaders in the area of their contribution, are also superb writers. Their fifteen chapters are a combination of broad surveys of major areas and techniques, together with more specialized expositions of many of the most active research topics in enumerative combinatorics today.

A prominent reason why practitioners of enumerative combinatorics find it so appealing is its unexpected connections with other areas of mathematics. These connections have grown increasingly sophisticated over the years. It is no longer sufficient to know some rudimentary algebraic topology, say, to give a significant connection with enumerative combinatorics. The papers in this handbook do an exemplary job of explaining deep connections with such areas as complex analysis, probability theory, linear algebra, commutative algebra, representation theory, algebraic geometry, algebraic topology, and computer science. Just this list of topics gives an idea of the breadth and depth of modern enumerative combinatorics. Readers from neophytes to experts have much to look forward to when they peruse the riches that follow.

Richard Stanley
Cambridge, MA
December 2014

Preface

When designing a handbook of a large and rapidly developing fields like Enumerative Combinatorics, one faces several questions: What subjects to cover? How to organize the subjects? What audience to target?

We have decided to include both chapters that focus on *methods* of enumeration of various objects and chapters that focus on specific kinds of *objects* that need to be counted, by any method available. The chapters are organized so that we advance from the more general ones, namely enumeration methods, towards the more specialized ones that focus on the counting of specific objects. These objects become more and more specialized as we proceed.

As far as our preferred audience goes, we believe that each chapter can benefit at least three different kinds of readers as listed below.

- The experts, who are familiar with most of the information in the chapter, but are interested in its presentation, and some of the finer points.
- The "relative outsiders," that is, readers who have already seen a few results here and there, a proof or two here and there, but nothing systematic, and who are interested to see an organized treatment of the topic.
- The novices, who are new to the field, and have no background information past the first year of graduate school. These readers will hopefully see that the subject is interesting, accessible, and challenging.

We do hope that all three groups of our targeted readership will find the book as useful and enjoyable as we do.

Miklós Bóna
Gainesville, FL
February 2015

Acknowledgments

At the completion of this book, I want to say thanks to Robert Ross at CRC Press, who initiated the project and selected me to lead it. The most important part of this endeavor were certainly the authors, who agreed to the difficult and time-consuming task of contributing a chapter to the book, and fulfilled that task, sometimes on time. My gratitude is extended to the mathematicians who have reviewed the various chapters. Turning 15 chapters from 17 different authors into one consistent LaTeX file was often too difficult for me, so I am grateful to Shashi Kumar at cenveo.com, who often helped me with technical matters. My gratitude is extended to my colleague Vincent Vatter, who was my local expert.

A non-negligible part of my editing work was done while I was on vacation. I owe a lot to my brother Péter who made sure that I had high-speed internet access when most people around us could not even make a phone call. Last, but not least, I am thankful to my wife Linda, and my sons Miki, Benny, and Vinnie for putting up with me when I worked at the book at highly unexpected times.

Part I

Methods

Chapter 1

Algebraic and Geometric Methods in Enumerative Combinatorics

Federico Ardila
San Francisco State University, San Francisco, CA, USA; Universidad de los Andes, Bogotá, Colombia

CONTENTS

1.1 Introduction

Enumerative combinatorics is about counting. The typical question is to find the number of objects with a given set of properties.

However, enumerative combinatorics is not just about counting. In "real life," when we talk about *counting*, we imagine lining up a set of objects and counting them off: $1,2,3,\ldots$. However, families of combinatorial objects do not come to us in a natural linear order. To give a very simple example: We do not count the squares in an $m \times n$ rectangular grid linearly. Instead, we use the rectangular structure to understand that the number of squares is $m \cdot n$. Similarly, to enumerate a more complicated combinatorial set, we usually spend most of our efforts understanding the underlying structure of the individual objects, or of the set itself.

Many combinatorial objects of interest have a rich and interesting **algebraic** or **geometric** structure, which often becomes a very powerful tool toward their enumeration. In fact, there are many families of objects that we **only** know how to count using these tools. Our goal in this chapter is to highlight some key aspects of the rich interplay between algebra, discrete geometry, and combinatorics, with an eye toward enumeration.

About this chapter. Over the last fifty years, combinatorics has undergone a radical transformation. Not too long ago, combinatorics mostly consisted of ad hoc methods and clever solutions to problems that were fairly isolated from the rest of mathematics. It has since grown to be a central area of mathematics, largely thanks to the discovery of deep connections to other fields. Combinatorics has become an essential tool in many disciplines. Conversely, even though ingenious methods and clever new ideas still abound, there is now a powerful, extensive toolkit of algebraic, geometric, topological, and analytic techniques that can be applied to combinatorial problems.

It is impossible to give a meaningful summary of the many facets of algebraic and geometric combinatorics in a writeup of this length. I found it very difficult but necessary to omit several beautiful, important directions. In the spirit of a *Handbook of Enumerative Combinatorics*, my guiding principle was to focus on algebraic and geometric techniques that are useful toward the solution of enumerative problems. My main goal was to state clearly and concisely some of the most useful tools in algebraic and geometric enumeration, and to give many examples that quickly and concretely illustrate how to put these tools to use.

PART 1. ALGEBRAIC METHODS

The first part of this chapter focuses on algebraic methods in enumeration. In Section 1.2 we discuss the question, "What is a good answer to an enumerative problem." Generating functions are the most powerful tool to unify the different kinds of answers that interest us: explicit formulas, recurrences, asymptotic formulas, and generating functions. In Section 1.3 we develop the algebraic theory of generating functions. Various natural operations on combinatorial families of objects correspond to simple algebraic operations on their generating functions, and this allows us to count many families of interest. In Section 1.4 we show how many problems in combinatorics can be rephrased in terms of linear algebra, and reduced to the problem of computing determinants. Finally, Section 1.5 is devoted to the theory of posets. Many combinatorial sets have a natural poset structure, and this general theory is very helpful in enumerating such sets.

1.2 What is a good answer?

The main goal of enumerative combinatorics is to count the elements of a finite set. Most frequently, we encounter a family of sets $T_0, T_1, T_2, T_3, \ldots$ and we need to find the number $t_n = |T_n|$ for $n = 1, 2, \ldots$. What constitutes a good answer?

Some answers are obviously good. For example, the number of subsets of $\{1, 2, \ldots, n\}$ is 2^n, and it seems clear that this is the simplest possible answer to this question. Sometimes an answer "is so messy and long, and so full of factorials and sign alternations and whatnot, that we may feel that the disease was preferable to the cure" [218]. Usually, the situation is somewhere in between, and it takes some experience to recognize a good answer.

A combinatorial problem often has several kinds of answers. Which answer is better depends on what one is trying to accomplish. Perhaps this is best illustrated with an example. Let us count the number a_n of **domino tilings** of a $2 \times n$ rectangle into 2×1 rectangles. There are several different ways of answering this question.

Explicit formula 1. We first look for an explicit combinatorial formula for a_n. To do that, we play with a few examples, and quickly notice that these tilings are structurally very simple: They are just a sequence of 2×1 vertical tiles, and 2×2 blocks covered by two horizontal tiles. Therefore constructing a tiling is the same as writing n as an ordered sum of 1s and 2s. For example, the tilings of Figure 1.1 correspond, respectively, to $1+1+1+1$, $1+1+2$, $1+2+1$, $2+1+1, 2+2$. These sums are

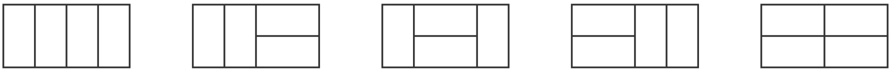


Figure 1.1
The five domino tilings of a 2×4 rectangle.

easy to count. If there are k summands equal to 2 there must be $n-2k$ summands equal to 1, and there are $\binom{n-2k+k}{k} = \binom{n-k}{k}$ ways of ordering the summands. Therefore

$$a_n = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n-k}{k} = \binom{n}{0} + \binom{n-1}{1} + \binom{n-2}{2} + \cdots. \tag{1.1}$$

This is a pretty good answer. It is certainly an explicit formula, and it may be used to compute a_n directly for small values of n. It does have two drawbacks. Aesthetically, it is certainly not as satisfactory as "2^n." In practice, it is also not as useful as it seems; after computing a few examples, we will soon notice that computing binomial coefficients is a non-trivial task. In fact there is a more efficient method of computing a_n.

Recurrence. Let $n \geq 2$. In a domino tiling, the leftmost column of a $2 \times n$ can be covered by a vertical domino or by two horizontal dominoes. If the leftmost domino is vertical, the rest of the dominoes tile a $2 \times (n-1)$ rectangle, so there are a_{n-1} such tilings. On the other hand, if the two leftmost dominoes are horizontal, the rest of the dominoes tile a $2 \times (n-2)$ rectangle, so there are a_{n-2} such tilings. We obtain the **recurrence relation**

$$a_0 = 1, \quad a_1 = 1, \quad a_n = a_{n-1} + a_{n-2} \text{ for } n \geq 2, \tag{1.2}$$

which allows us to compute each term in the sequence in terms of the previous ones. We see that $a_n = F_{n+1}$ is the $(n+1)$th **Fibonacci number**.

This recursive answer is not as nice as "2^n" either; it is not even an explicit formula for a_n. If we want to use it to compute a_n, we need to compute all the first n terms of the sequence $1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots$. However, we can compute those very quickly; we only need to perform $n-1$ additions. This is an extremely efficient method for computing a_n.

Explicit formula 2. There is a well-established method that turns linear recurrence relations with constant coefficients, such as (1.2), into explicit formulas. We will review it in Theorem 1.3.5. In this case, the method gives

$$a_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{n+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{n+1} \right). \tag{1.3}$$

This is clearly the simplest possible explicit formula for a_n; in that sense it is a great formula.

A drawback is that this formula is really not very useful if we want to compute the exact value of, say, a_{1000}. It is not even clear why (1.3) produces an integer, and to get it to produce the correct integer would require arithmetic calculations with extremely high precision.

An advantage is that, unlike (1.1), (1.3) tells us very precisely how a_n grows with n.

Asymptotic formula. It follows immediately from (1.3) that

$$a_n \sim c \cdot \varphi^n, \tag{1.4}$$

where $c = \frac{1+\sqrt{5}}{2\sqrt{5}}$ and $\varphi = \frac{1+\sqrt{5}}{2} \approx 1.6179\ldots$ is the golden ratio. This notation means that $\lim_{n\to\infty} a_n/(c \cdot \varphi^n) = 1$. In fact, since $|\frac{1-\sqrt{5}}{2}| < 1$, a_n is the closest integer to $c \cdot \varphi^n$.

Generating function. The last kind of answer we discuss is the generating function. This is perhaps the strangest kind of answer, but it is often the most powerful one.

Consider the infinite power series $A(x) = a_0 + a_1 x + a_2 x^2 + \cdots$. We call this the **generating function** of the sequence $a_0, a_1, a_2, \ldots$. * We now compute this power series: From (1.2) we obtain that $A(x) = 1 + x + \sum_{n\geq 2}(a_{n-1} + a_{n-2})x^n = 1 + x + x(A(x) - 1) + x^2 A(x)$, which implies

$$A(x) = a_0 + a_1 x + a_2 x^2 + \cdots = \frac{1}{1 - x - x^2}. \tag{1.5}$$

With a bit of theory and some practice, we will be able to write the equation (1.5) immediately, with no further computations (Example 18 in Section 1.3.2). To show this is an **excellent** answer, let us use it to derive all our other answers, and more.

- Generating functions help us obtain explicit formulas. For instance, rewriting

  $$A(x) = \frac{1}{1 - (x + x^2)} = \sum_{k\geq 0} (x + x^2)^k$$

  we recover (1.1). If, instead, we use the method of partial fractions, we get

  $$A(x) = \left(\frac{1/\sqrt{5}}{1 - \frac{1+\sqrt{5}}{2} x} \right) - \left(\frac{1/\sqrt{5}}{1 - \frac{1-\sqrt{5}}{2} x} \right)$$

  which brings us to our second explicit formula (1.3).

- Generating functions help us obtain recursive formulas. In this example, we simply compare the coefficients of x^n on both sides of the equation $A(x)(1 - x - x^2) = 1$, and we get the recurrence relation (1.2).

*For the moment, let us not worry about where this series converges. The issue of convergence can be easily avoided (as combinatorialists often do, in a way that will be explained in Section 1.3.1) or resolved and exploited to our advantage; let us postpone that discussion for the moment.

- Generating functions help us obtain asymptotic formulas. In this example, (1.5) leads to (1.3), which gives (1.4). In general, almost everything that we know about the rate of growth of combinatorial sequences comes from their generating functions, because analysis tells us that the asymptotic behavior of a_n is intimately tied to the singularities of the function $A(x)$.

- Generating functions help us enumerate our combinatorial objects in more detail, and understand some of their statistical properties. For instance, say we want to compute the number $a_{m,n}$ of domino tilings of a $2 \times n$ rectangle that use exactly m vertical tiles. Once we really understand (1.5) in Section 1.3.2, we will get the answer immediately:

$$\frac{1}{1 - vx - x^2} = \sum_{m,n \geq 0} a_{m,n} v^m x^n.$$

  Now suppose we wish to know what fraction of the tiles is vertical in a large random tiling. Among all the a_n domino tilings of the $2 \times n$ rectangle, there are $\sum_{m \geq 0} m a_{m,n}$ vertical dominoes. We compute

$$\sum_{n \geq 0} \left(\sum_{m \geq 0} m a_{m,n} \right) x^n = \left[\frac{\partial}{\partial v} \left(\frac{1}{1 - vx - x^2} \right) \right]_{v=1} = \frac{x}{(1 - x - x^2)^2}.$$

  Partial fractions then tell us that $\sum_{m \geq 0} m a_{m,n} \sim \frac{n}{5} \left(\frac{1+\sqrt{5}}{2} \right)^{n+1} \sim \frac{1}{\sqrt{5}} n a_n$. Hence the fraction of vertical tiles in a random domino tiling of a $2 \times n$ rectangle converges to $1/\sqrt{5}$ as $n \to \infty$.

So what is a good answer to an enumerative problem? Not surprisingly, there is no definitive answer to this question. When we count a family of combinatorial objects, we look for explicit formulas, recursive formulas, asymptotic formulas, and generating functions. They are all useful. Generating functions are the most powerful framework we have to relate these different kinds of answers and, ideally, find them all.

1.3 Generating functions

In combinatorics, one of the most useful ways of "determining" a sequence of numbers $a_0, a_1, a_2, \ldots$ is to compute its **ordinary generating function**

$$A(x) = \sum_{n \geq 0} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \cdots$$

or its **exponential generating function**

$$A_{\exp}(x) = \sum_{n \geq 0} a_n \frac{x^n}{n!} = a_0 + a_1 x + a_2 \frac{x^2}{2} + a_3 \frac{x^3}{6} + a_4 \frac{x^4}{24} + \cdots.$$

This simple idea is extremely powerful because some of the most common algebraic operations on ordinary and exponential generating functions correspond to some of the most common operations on combinatorial objects. This allows us to count many interesting families of objects; this is the content of Section 1.3.2 (for ordinary generating functions) and Section 1.3.3 (for exponential generating functions). In Section 1.3.4 we see how nice generating functions can be turned into explicit, recursive, and asymptotic formulas for the corresponding sequences.

Before we get to this interesting theory, we have to understand what we mean by power series. Section 1.3.1 provides a detailed discussion, which is probably best skipped the first time one encounters power series. In the meantime, let us summarize it in one paragraph:

There are two main attitudes toward power series in combinatorics: the analytic attitude and the algebraic attitude. To harness the full power of power series, one should really understand both. Chapter 2 of this *Handbook of Enumerative Combinatorics* is devoted to the analytic approach, which treats $A(x)$ as an honest analytic function of x, and uses analytic properties of $A(x)$ to derive combinatorial properties of a_n. In this chapter we follow the algebraic approach, which treats $A(x)$ as a formal algebraic expression, and manipulates it using the usual laws of algebra, without having to worry about any convergence issues.

1.3.1 The ring of formal power series

Enumerative combinatorics is full of intricate algebraic computations with power series, where justifying convergence is cumbersome, and usually unnecessary. In fact, many natural power series in combinatorics, such as $\sum_{n\geq 0} n!x^n$, only converge at 0, so analytic methods are not available to study them. For these reasons we often prefer to carry out our computations algebraically in terms of **formal** power series. We will see that even in this approach, analytic considerations are often useful.

In this section we review the definition and basic properties of the **ring of formal power series** $\mathbb{C}[[x]]$. For a more in-depth discussion, including the (mostly straightforward) proofs of the statements we make here, see [149].

Formal power series. A **formal power series** is an expression of the form

$$A(x) = a_0 + a_1x + a_2x^2 + \cdots, \qquad a_0, a_1, a_2, \ldots \in \mathbb{C}.$$

Formally, this series is just a sequence of complex numbers $a_0, a_1, a_2, \ldots$. We will see that it is convenient to denote it $A(x)$, but we do not consider it to be a function of x.

Let $\mathbb{C}[[x]]$ be the **ring of formal power series**, where the sum and the product of $A(x) = \sum_{n\geq 0} a_nx^n$ and $B(x) = \sum_{n\geq 0} b_nx^n$ are

$$A(x)+B(x) = \sum_{n\geq 0}(a_n+b_n)x^n, \qquad A(x)B(x) = \sum_{n\geq 0}\left(\sum_{k=0}^{n} a_kb_{n-k}\right)x^n.$$

It is implicit in the definition, but worth mentioning, that $\sum_{n\geq 0} a_n x^n = \sum_{n\geq 0} b_n x^n$ if and only if $a_n = b_n$ for all $n \geq 0$.

We define the **degree** of $A(x) = \sum_{n\geq 0} a_n x^n$ to be the **smallest** n such that $a_n \neq 0$. We also write

$$[x^n]A(x) := a_n, \qquad A(0) := [x^0]A(x) = a_0.$$

We also define formal power series inspired by series from analysis, such as

$$e^x := \sum_{n\geq 0} \frac{x^n}{n!}, \qquad -\log(1-x) := \sum_{n\geq 1} \frac{x^n}{n}, \qquad (1+x)^r := \sum_{n\geq 0} \binom{r}{n} x^n,$$

for any complex number r, where $\binom{r}{n} := r(r-1)\cdots(r-n+1)/n!$.

The ring $\mathbb{C}[[x]]$ is commutative with $0 = 0+0x+\cdots$ and $1 = 1+0x+\cdots$. It is an integral domain; that is, $A(x)B(x) = 0$ implies that $A(x) = 0$ or $B(x) = 0$. It is easy to describe the units:

$$\sum_{n\geq 0} a_n x^n \text{ is invertible} \quad \Longleftrightarrow \quad a_0 \neq 0.$$

For example, $\frac{1}{1-x} = 1+x+x^2+\cdots$ because $(1-x)(1+x+x^2+\cdots) = 1+0x+0x^2+\cdots$.

Convergence. When working in $\mathbb{C}[[x]]$, we will not consider convergence of sequences or series of complex numbers. In particular, we will never substitute a complex number x into a formal power series $A(x)$.

However, we do need a notion of convergence for sequences in $\mathbb{C}[[x]]$. We say that a sequence $A_0(x), A_1(x), A_2(x), \ldots$ of formal power series **converges** to $A(x) = \sum_{n\geq 0} a_n x^n$ if $\lim_{n\to\infty} \deg(A_n(x) - A(x)) = \infty$; that is, if for any $n \in \mathbb{N}$, the coefficient of x^n in $A_m(x)$ **equals** a_n for all sufficiently large m. This gives us a useful criterion for convergence of infinite sums and products in $\mathbb{C}[[x]]$:

$$\sum_{j=0}^{\infty} A_j(x) \text{ converges} \quad \Longleftrightarrow \quad \lim_{j\to\infty} \deg A_j(x) = \infty$$

$$\prod_{j=0}^{\infty} (1+A_j(x)) \text{ converges} \quad \Longleftrightarrow \quad \lim_{j\to\infty} \deg A_j(x) = \infty \qquad (A_j(0) = 0)$$

For example, the infinite sum $\sum_{n\geq 0} (x+1)^n/2^n$ does not converge in this topology. Notice that the coefficient of x^0 in this sum cannot be obtained through a finite computation; it would require interpreting the infinite sum $\sum_{n\geq 0} 1/2^n$. On the other hand, the following infinite sum converges:

$$\sum_{n\geq 0} \frac{1}{n!}\left(-\sum_{m\geq 1} \frac{x^m}{m}\right)^n = 1-x. \tag{1.6}$$

It is clear from the criterion above that this series converges; but why does it equal $1-x$?

Borrowing from analysis. In $\mathbb{C}[[x]]$, (1.6) is an algebraic identity which says that the coefficients of x^k in the left-hand side, for which we can give an ugly but finite formula, equal $1, -1, 0, 0, 0, \ldots$. If we were to follow a purist algebraic attitude, we would give an algebraic or combinatorial proof of this identity. This is probably possible, but intricate and rather dogmatic. A much simpler approach is to shift toward an analytic attitude, at least momentarily, and recognize that (1.6) is the Taylor series expansion of

$$e^{-\log(1-x)} = 1 - x$$

for $|x| < 1$. Then we can just invoke the following simple fact from analysis.

Theorem 1.3.1 *If two analytic functions are equal in an open neighborhood of* 0*, then their Taylor series at* 0 *are equal coefficient-by-coefficient; that is, they are equal as formal power series.*

Composition. The **composition** of two series $A(x) = \sum_{n\geq 0} a_n x^n$ and $B(x) = \sum_{n\geq 0} b_n x^n$ with $b_0 = 0$ is naturally defined to be

$$A(B(x)) = \sum_{n\geq 0} a_n \left(\sum_{m\geq 0} b_m x^m \right)^n.$$

Note that this sum converges if and only if $b_0 = 0$. Two very important special cases in combinatorics are the series $\frac{1}{1-B(x)}$ and $e^{B(x)}$.

"Calculus." We define the **derivative** of $A(x) = \sum_{n\geq 0} a_n x^n$ to be

$$A'(x) = \sum_{n\geq 0} (n+1) a_{n+1} x^n.$$

This formal derivative satisfies the usual laws of derivatives, such as

$$(A+B)' = A' + B', \qquad (AB)' = A'B + AB', \qquad [A(B(x))]' = A'(B(x))B'(x).$$

We can still solve differential equations formally. For example, if we know that $F'(x) = F(x)$ and $F(0) = 1$, then $(\log F(x))' = F'(x)/F(x) = 1$, which gives $\log F(x) = x$ and $F(x) = e^x$.

This concludes our discussion on the formal properties of power series. Now let us return to combinatorics.

1.3.2 Ordinary generating functions

Suppose we are interested in enumerating a family $\mathscr{A} = \mathscr{A}_0 \sqcup \mathscr{A}_1 \sqcup \mathscr{A}_2 \sqcup \cdots$ of combinatorial structures, where $\mathscr{A}_n$ is a finite set consisting of the objects of "size" n. Denote by $|a|$ the size of $a \in \mathscr{A}$. The **ordinary generating function** of $\mathscr{A}$ is

$$A(x) = \sum_{a\in\mathscr{A}} x^{|a|} = a_0 + a_1 x + a_2 x^2 + \cdots$$

where a_n is the number of elements of size n.

We are not interested in the philosophical question of determining what it means for $\mathscr{A}$ to be "combinatorial"; we are willing to call $\mathscr{A}$ a combinatorial structure as long as a_n is finite for all n. We consider two structures $\mathscr{A}$ and $\mathscr{B}$ combinatorially equivalent, and write $\mathscr{A} \cong \mathscr{B}$, if $A(x) = B(x)$.

More generally, we may consider a family $\mathscr{A}$ where each element a is given a weight $\mathrm{wt}(a)$, often a constant multiple of $x^{|a|}$, or a monomial in one or more variables $x_1, \dots, x_n$. Again, we require that there are finitely many objects of any given weight. Then we define the **weighted** ordinary generating function of $\mathscr{A}$ to be the formal power series

$$A_{\mathrm{wt}}(x_1, \dots, x_n) = \sum_{a \in \mathscr{A}} \mathrm{wt}(a).$$

Examples of combinatorial structures (with their respective size functions) are words on the alphabet $\{0,1\}$ (length), domino tilings of rectangles of height 2 (width), or Dyck paths (length). We may weight these objects by t^k where k is, respectively, the number of 1s, the number of vertical tiles, or the number of returns to the x axis.

1.3.2.1 Operations on combinatorial structures and their generating functions

There are a few simple but very powerful operations on combinatorial structures, all of which have nice counterparts at the level of ordinary generating functions. Many combinatorial objects of interest may be built up from basic structures using these operations.

Theorem 1.3.2 *Let $\mathscr{A}$ and $\mathscr{B}$ be combinatorial structures.*

1. *($\mathscr{C} = \mathscr{A} + \mathscr{B}$: Disjoint union) If a $\mathscr{C}$-structure of size n is obtained by choosing an $\mathscr{A}$-structure of size n **or** a $\mathscr{B}$-structure of size n, then*

   $$C(x) = A(x) + B(x).$$

   This result also holds for weighted structures if the weight of a $\mathscr{C}$-structure is the same as the weight of the respective $\mathscr{A}$- or $\mathscr{B}$-structure.

2. *($\mathscr{C} = \mathscr{A} \times \mathscr{B}$: Product) If a $\mathscr{C}$-structure of size n is obtained by choosing an $\mathscr{A}$-structure of size k **and** a $\mathscr{B}$-structure of size $n-k$ for some k, then*

   $$C(x) = A(x)B(x).$$

   This result also holds for weighted structures if the weight of a $\mathscr{C}$-structure is the product of the weights of the respective $\mathscr{A}$- and $\mathscr{B}$-structures.

3. *($\mathscr{C} = \mathsf{Seq}(\mathscr{B})$: Sequence) Assume $|\mathscr{B}_0| = 0$. If a $\mathscr{C}$-structure of size n is obtained by choosing a sequence of $\mathscr{B}$-structures of total size n, then*

   $$C(x) = \frac{1}{1 - B(x)}.$$

   This result also holds for weighted structures if the weight of a $\mathscr{C}$-structure is the product of the weights of the respective $\mathscr{B}$-structures.

4. *($\mathscr{C} = \mathscr{A} \circ \mathscr{B}$: Composition)* ***Compositional formula.*** *Assume that $|\mathscr{B}_0| = 0$. If a $\mathscr{C}$-structure of size n is obtained by choosing a sequence of (say, k) $\mathscr{B}$-structures of total size n and placing an $\mathscr{A}$-structure of size k on this sequence of $\mathscr{B}$-structures, then*
$$C(x) = A(B(x)).$$
This result also holds for weighted structures if the weight of a $\mathscr{C}$-structure is the product of the weights of the $\mathscr{A}$-structure on its blocks and the weights of the $\mathscr{B}$-structures on the individual blocks.

5. *($\mathscr{C} = \mathscr{A}^{-1}$: Inversion)* ***Lagrange inversion formula.***

   (a) ***Algebraic version.*** *If $A^{<-1>}(x)$ is the compositional inverse of $A(x)$ then*
$$n[x^n]A^{<-1>}(x) = [x^{n-1}]\left(\frac{x}{A(x)}\right)^n.$$

   (b) ***Combinatorial version.*** *Assume $|\mathscr{A}_0| = 0$, $|\mathscr{A}_1| = 1$, and let*
$$A(x) = x - a_2x^2 - a_3x^3 - a_4x^4 - \cdots$$
   where a_n is the number of $\mathscr{A}$-structures of size n for $n \geq 2$. *

Let an ***$\mathscr{A}$-decorated plane rooted tree*** *(or simply $\mathscr{A}$-tree) be a rooted tree T where every internal vertex v has an ordered set D_v of at least two "children," and each set D_v is given an $\mathscr{A}$-structure. The* **size** *of T is the number of leaves.*

Let $C(x)$ be the generating function for $\mathscr{A}$-decorated plane rooted trees. Then
$$C(x) = A^{<-1>}(x).$$
This result also holds for weighted structures if the weight of a tree is the product of the weights of the $\mathscr{A}$-structures at its vertices.

Proof. 1. is clear. The identity in 2. is equivalent to $c_n = \sum_k a_k b_{n-k}$, which corresponds to the given combinatorial description. Iterating 2., the generating function for k-sequences of $\mathscr{B}$-structures is $B(x)^k$, so in 3. we have $C(x) = \sum_k B(x)^k = 1/(1-B(x))$ and in 4. we have $C(x) = \sum_k a_k B(x)^k$. The weighted statements follow similarly.

5(b). Observe that, by the Compositional Formula, an $\mathscr{A} \circ \mathscr{C}$ structure is either (i) an $\mathscr{A}$-tree, or (ii) a sequence of $k \geq 2$ $\mathscr{A}$-trees $T_1, \ldots, T_k$ with an $\mathscr{A}$-structure on $\{T_1, \ldots, T_k\}$.

The structure in (ii) is equivalent to an $\mathscr{A}$-tree T, obtained by grafting $T_1, \ldots, T_k$ at a new root and placing the $\mathscr{A}$-structure on its offspring (which contributes a negative

*Here, to simplify matters, we introduced signs into $A(x)$. Instead we could let $A(x)$ be the ordinary generating function for $\mathscr{A}$-structures, but we would need to give each $\mathscr{A}$-tree the sign $(-1)^m$, where m is the number of internal vertices. Similarly, we could allow $a_1 \neq 1$ at the cost of some factors of a_1 on the $\mathscr{A}$-trees.

sign). This T also arises in (i) with a positive sign. These two appearances of T cancel each other out in $A(C(x))$, and the only surviving tree is the trivial tree $\bullet$ with one vertex, which only arises once and has weight x.

5(a). Let a **sprig** be a rooted plane tree consisting of a path $r = v_1 v_2 \cdots v_k = l$ starting at the root r and ending at the leaf l, and at least one leaf hanging from each v_i and to the right of v_{i+1} for $1 \le i \le k-1$. The trivial tree $\bullet$ is an allowable sprig with $k = 1$.

An $\mathscr{A}$**-sprig** is a sprig where the children of v_i are given an $\mathscr{A}$-structure for $1 \le i \le k-1$; its **size** is the number of leaves other than l minus 1. * The right panel of Figure 1.2 shows several $\mathscr{A}$-sprigs. An $\mathscr{A}$-sprig is equivalent to a sequence of $\mathscr{A}$-structures, with weights shifted by -1, so Theorem 1.3.2.3 tells us that

$$\frac{1}{A(x)} = \frac{1}{x} \cdot \frac{1}{1-(a_2x + a_3x^2 + \cdots)} = \sum_{n \ge -1} (\text{\# of } \mathscr{A}\text{-sprigs of size } n)\, x^n$$

Hence $[x^{n-1}](x/A(x))^n = [x^{-1}](1/A(x))^n$ is the number of sequences of n $\mathscr{A}$-sprigs of total size -1 by Theorem 1.3.2.2. We need to show that

$$n \cdot (\text{\# of } \mathscr{A}\text{-trees with } n \text{ leaves}) = (\text{\# of sequences of } n \ \mathscr{A}\text{-sprigs of total size } -1)$$

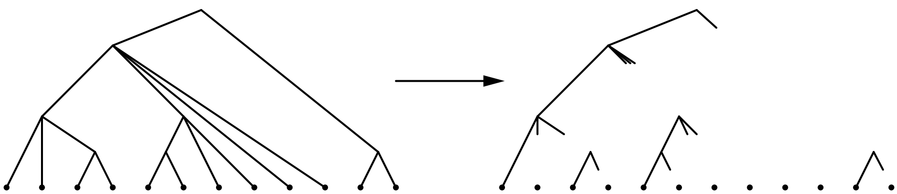


Figure 1.2
The map from $\mathscr{A}$-trees to sequences of $\mathscr{A}$-sprigs.

An $\mathscr{A}$-tree T can be trimmed into a sequence of n $\mathscr{A}$-sprigs $S_1, \ldots, S_n$ as follows. At each step, look at the leftmost leaf and the path P to its highest remaining ancestor. Remove P and all the branches hanging directly from P (which form an $\mathscr{A}$-sprig), but do not remove any other vertices. Repeat this until the tree is completely decomposed into $\mathscr{A}$-sprigs. The total size of these sprigs is -1. Figure 1.2 shows a tree of weight $x^{5+(-1)+0+(-1)+2+(-1)+(-1)+(-1)+(-1)+(-1)+0+(-1)} = x^{-1}$. Notice that all the partial sums of the sum $5+(-1)+0+(-1)+2+(-1)+(-1)+(-1)+(-1)+(-1)+0+(-1) = -1$ are non-negative.

Conversely, suppose we wish to recover the $\mathscr{A}$-tree corresponding to a sequence of sprigs $S_1, \ldots, S_n$ with $|S_1| + \cdots + |S_n| = -1$. We must reverse the process, adding

*We momentarily allow negative sizes, since the trivial $\mathscr{A}$-sprig $\bullet$ has size -1. Thus we need to compute with Laurent series, which are power series with finitely many negative exponents.

$S_1,\dots,S_n$ to T one at a time; at each step we must graft the new sprig at the leftmost free branch. Note that after grafting $S_1,\dots,S_k$ we are left with $1+|S_1|+\cdots+|S_k|$ free branches, so a sequence of sprigs corresponds to a tree if and only if the partial sums $|S_1|+\cdots+|S_k|$ are non-negative for $k=1,\dots,n-1$. Finally, it remains to observe that any sequence $a_1,\dots,a_n$ of integers adding to -1 has a unique cyclic shift $a_i,\dots,a_n,a_1,\dots,a_{i-1}$ whose partial sums are all non-negative. Therefore, out of the n cyclic shifts of $S_1,\dots,S_n$, exactly one of them corresponds to an $\mathscr{A}$-tree. The desired result follows. ■

The last step of the proof above is a special case of the **cycle lemma** of Dvoretsky and Motzkin [192, Lemma 5.3.7], which is worth stating explicitly. Suppose $a_1,\dots,a_n$ is a string of 1s and -1s with $a_1+\cdots+a_n=k>0$. Then there are exactly k cyclic shifts $a_i,a_{i+1},\dots,a_n,a_1,\dots,a_{i-1}$ whose partial sums are all non-negative.

1.3.2.2 Examples

Classical applications. With practice, these simple ideas give very easy solutions to many classical enumeration problems.

1. (Trivial classes) It is useful to introduce the trivial class $\circ$ having only one element of size 0, and the trivial class $\bullet$ having only one element of size 1. Their generating functions are 1 and x, respectively.

2. (Sequences) The slightly less trivial class $\mathsf{Seq}=\{\emptyset,\bullet,\bullet\bullet,\bullet\bullet\bullet,\dots\}=\mathsf{Seq}(\bullet)$ contains one set of each size. Its generating function is $\sum_n x^n=1/(1-x)$.

3. (Subsets and binomial coefficients) Let Subset consist of the pairs $([n],A)$ where n is a natural number and A is a subset of $[n]$. Let the size of that pair be n. A Subset-structure is equivalent to a word of length n in the alphabet $\{0,1\}$, so $\mathsf{Subset}\cong\mathsf{Seq}(\{0,1\})$ where $|0|=|1|=1$, and

$$\mathsf{Subset}(x)=\frac{1}{1-(x^1+x^1)}=\sum_{n\geq 0}2^nx^n.$$

We can use the extra variable y to keep track of the size of the subset A, by giving $([n],A)$ the weight $x^ny^{|A|}$. This corresponds to giving the letters 0 and 1 weights x and xy respectively, so we get the generating function

$$\mathsf{Subset}_{\mathrm{wt}}(x)=\frac{1}{1-(x+xy)}=\sum_{n\geq k\geq 0}\binom{n}{k}x^ny^k$$

for the **binomial coefficients** $\binom{n}{k}=\frac{n!}{k!(n-k)!}$, which count the k-subsets of $[n]$.

From this generating function, we can easily obtain the main results about binomial coefficients. Computing the coefficient of x^ny^k in $\left(\sum\binom{n}{k}x^ny^k\right)(1-x-xy)=1$ gives **Pascal's recurrence**

$$\binom{n}{k}=\binom{n-1}{k}+\binom{n-1}{k-1},\qquad n\geq k\geq 1$$

with initial values $\binom{n}{0} = \binom{n}{n} = 1$. Expanding $\mathsf{Subset}_{\mathrm{wt}}(x) = \frac{1}{1-x(1+y)} = \sum_{n\geq 0} x^n(1+y)^n$ gives the **Binomial theorem**

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

4. (Multinomial coefficients) Let $\mathsf{Words}^k \cong \mathsf{Seq}(\{1,\ldots,k\})$ consist of the words in the alphabet $\{1,2,\ldots,k\}$. The words of length n are in bijection with the ways of putting n numbered balls into k numbered boxes. The placements having a_i balls in box i, where $a_1+\cdots+a_k = n$, are enumerated by the **multinomial coefficient** $\binom{n}{a_1,\ldots,a_k} = \frac{n!}{a_1!\cdots a_k!}$.

   Giving the letter i weight x_i, we obtain the generating function

   $$\mathsf{Words}^k(x_1,\ldots,x_k) = \sum_{a_1,\ldots,a_k\geq 0} \binom{a_1+\cdots+a_k}{a_1,\ldots,a_k} x_1^{a_1}\cdots x_k^{a_k} = \frac{1}{1-x_1-\cdots-x_k}$$

   from which we obtain the recurrence

   $$\binom{n}{a_1,\ldots,a_k} = \binom{n-1}{a_1-1,a_2,\ldots,a_k} + \cdots + \binom{n-1}{a_1,\ldots a_{k-1},a_k-1}$$

   and the multinomial theorem

   $$(x_1+\cdots+x_k)^n = \sum_{\substack{a_1,\ldots,a_k\geq 0\\ a_1+\cdots+a_k=n}} \binom{n}{a_1,\ldots,a_k} x_1^{a_1}\cdots x_k^{a_k}.$$

5. (Compositions) A **composition** of n is a way of writing $n = a_1+\cdots+a_k$ as an **ordered** sum of positive integers $a_1,\ldots,a_k$. For example, 523212 is a composition of 15. A composition is just a sequence of natural numbers, so $\mathsf{Comp} \cong \mathsf{Seq}(\mathbb{N})$ where $|a| = a$. Therefore

   $$\mathsf{Comp}(x) = \frac{1}{1-(x+x^2+x^3+\cdots)} = \frac{1-x}{1-2x} = \sum_{n\geq 1} 2^{n-1}x^n.$$

   and there are 2^{n-1} compositions of n.

   If we give a composition of n with k summands the weight $x^n y^k$, the weighted generating function is

   $$\mathsf{Comp}_{\mathrm{wt}}(x) = \frac{1}{1-(xy+x^2y+x^3y+\cdots)} = \frac{1-x}{1-x(1+y)} = \sum_{n\geq 1} \binom{n-1}{k-1} x^n y^k$$

   so there are $\binom{n-1}{k-1}$ compositions of n with k summands.

6. (Compositions into restricted parts) Given a subset $A \subseteq \mathbb{N}$, an **A-composition** of n is a way of writing n as an ordered sum $n = a_1+\cdots+a_k$ where $a_1,\ldots,a_k \in$

A. The corresponding combinatorial structure is $A\text{-}\mathsf{Comp} \cong \mathsf{Seq}(A)$ where $|a| = a$, so

$$A\text{-}\mathsf{Comp}(x) = \frac{1}{1 - (\sum_{a \in A} x^a)}.$$

For example, the number of compositions of n into odd parts is the Fibonacci number F_{n-1}, because the corresponding generating function is

$$\mathsf{OddComp}(x) = \frac{1}{1 - (x + x^3 + x^5 + \cdots)} = \frac{1 - x^2}{1 - x - x^2} = 1 + \sum_{n \geq 1} F_{n-1} x^n.$$

7. (Multisubsets) Let $\mathsf{Multiset}^m$ be the collection of multisets consisting of possibly repeated elements of $[m]$. The size of a multiset is the number of elements, counted with repetition. For example, $\{1,2,2,2,3,5\}$ is a multisubset of $[7]$ of size 6. Then $\mathsf{Multiset}^m \cong \mathsf{Seq}(\{1\}) \times \cdots \times \mathsf{Seq}(\{m\})$, where $|i| = 1$ for $i = 1, \ldots, m$, so the corresponding generating function is

$$\mathsf{Multiset}^m(x) = \left(\frac{1}{1-x}\right)^m = \sum_{n \geq 0} \binom{-m}{n} (-x)^n,$$

and the number of multisubsets of $[m]$ of size n is $\left(\!\binom{m}{n}\!\right) := (-1)^n \binom{-m}{n} = \binom{m+n-1}{n}$.

8. (Partitions) A **partition** of n is a way of writing $n = a_1 + \cdots + a_k$ as an **unordered** sum of positive integers $a_1, \ldots, a_k$. We usually write the parts in weakly decreasing order. For example, 532221 is a partition of 15 into 6 parts. Let $\mathsf{Partition}$ be the family of partitions weighted by $x^n y^k$ where n is the sum of the parts and k is the number of parts. Then $\mathsf{Partition} \cong \mathsf{Seq}(\{1\}) \times \mathsf{Seq}(\{2\}) \times \cdots$, where $\mathrm{wt}(i) = x^i y$ for $i = 1, 2, \ldots$, so the corresponding generating function is

$$\mathsf{Partition}(x,y) = \left(\frac{1}{1-xy}\right)\left(\frac{1}{1-x^2y}\right)\left(\frac{1}{1-x^3y}\right)\cdots$$

There is no simple explicit formula for the number $p(n)$ of partitions of n, although there is a very elegant and efficient recursive formula. Setting $y = -1$ in the previous identity, and invoking Euler's pentagonal theorem [2]

$$\prod_{n \geq 0} (1 - x^n) = 1 + \sum_{j \geq 1} (-1)^j \left(x^{j(3j-1)/2} + x^{j(3j+1)/2}\right) \tag{1.7}$$

we obtain

$$p(n) = p(n-1) + p(n-2) - p(n-5) - p(n-7) + p(n-12) + p(n-15) - \cdots$$

where $1, 2, 5, 7, 12, 15, 22, 26, \ldots$ are the pentagonal numbers.

9. (Partitions into distinct parts) Let $\mathsf{DistPartition}$ be the family of partitions into distinct parts, weighted by $x^n y^k$ where n is the sum of the parts and k is the number of parts. Then $\mathsf{DistPartition} \cong \{1,\overline{1}\} \times \{2,\overline{2}\} \times \cdots$, where $\mathrm{wt}(i) = x^i y$ and $\mathrm{wt}(\overline{i}) = 1$ for $i = 1, 2, \ldots$, so the corresponding generating function is
$$\mathsf{DistPartition}(x,y) = (1+xy)(1+x^2y)(1+x^3y)\cdots$$

10. (Partitions into restricted parts) It is clear how to adapt the previous generating functions to partitions where the parts are restricted. For example, the identity
$$\frac{1}{1-x} = (1+x)(1+x^2)(1+x^4)(1+x^8)(1+x^{16})\cdots$$
expresses the fact that every positive integer can be written uniquely in binary notation, as a sum of distinct powers of 2. The identity
$$(1+x)(1+x^2)(1+x^3)(1+x^4)\cdots = \frac{1}{1-x}\cdot\frac{1}{1-x^3}\cdot\frac{1}{1-x^5}\cdot\frac{1}{1-x^7}\cdots,$$
which may be proved by writing $1+x^k = (1-x^{2k})/(1-x^k)$, expresses that the number of partitions of n into distinct parts equals the number of partitions of n into odd parts.

11. (Partitions with restrictions on the size and the number of parts) Let $p_{\leq k}(n)$ be the number of partitions of n into at most k parts. This is also the number of partitions of n into parts of size at most k. To see this, represent a partition $n = a_1 + \cdots + a_j$ as a left-justified array of squares, where the ith row has a_i squares. Each partition λ has a conjugate partition λ' obtained by exchanging the rows and the columns of the Ferrers diagram. Figure 1.3 shows the Ferrers diagram of 431 and its conjugate partition 3221. It is clear that λ has at most k parts if and only if λ' has parts of size at most k.

    From the previous discussion it is clear that
$$\sum_{n\geq 0} p_{\leq k}(n)x^n = \frac{1}{1-x}\cdot\frac{1}{1-x^2}\cdots\frac{1}{1-x^k}.$$

    Now let $p_{\leq j,\leq k}(n)$ be the number of partitions of n into at most j parts of size at most k. Then
$$\sum_{n\geq 0} p_{\leq j,\leq k}(n)x^n = \frac{(1-x)(1-x^2)\cdots(1-x^{j+k})}{(1-x)(1-x^2)\cdots(1-x^j)\cdot(1-x)(1-x^2)\cdots(1-x^k)}.$$

    This is easily proved by induction, using that $p_{\leq j,\leq k}(n) = p_{\leq j,\leq k-1}(n) + p_{\leq j,\leq k}(n-k)$.

12. (Even and odd partitions) Setting $y = -1$ into the generating function for partitions into distinct parts of Example 9, we get
$$\sum_{n\geq 0} (\mathsf{edp}(n) - \mathsf{odp}(n))x^n = \prod_{j\geq 1}(1-x^j) = 1 - x - x^2 + x^5 + x^7 - x^{12} - x^{15} + \cdots$$

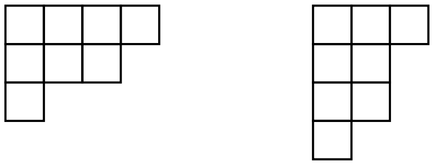


Figure 1.3
The Ferrers diagrams of the conjugate partitions 431 and 3221.

where $\mathsf{edp}(n)$ (respectively, $\mathsf{odp}(n)$) counts the partitions of n into an even (respectively, odd) number of distinct parts. Euler's pentagonal formula (1.7) says that $\mathsf{edp}(n) - \mathsf{odp}(n)$ equals 0 for all n except for the pentagonal numbers, for which it equals 1 or -1.

There are similar results for partitions into distinct parts coming from a given set S.

- When S is the set of Fibonacci numbers, the coefficients of the generating function

$$\prod_{n\geq 1}(1-x^{F_n}) = 1 - x - x^2 + x^4 + x^7 - x^8 + x^{11} - x^{12} - x^{13} + x^{14} + \cdots$$

  are also equal to $0, 1$, or -1. [5, 168]
- This is also true for any "k-Fibonacci sequence" $S = \{a_1, a_2, \ldots\}$ given by $a_n = a_{n-1} + \cdots + a_{n-k}$ for $n > k$ and $a_j > a_{j-1} + \cdots + a_1$ for $1 \leq j \leq k$. [67]
- The result also holds trivially for $S = \{2^j : j \in \mathbb{N}\}$ since there is a unique partition of any n into distinct powers of 2.

These three results seem qualitatively different from (and increasingly less surprising than) Euler's result, as these sequences S grow much faster than $\{1, 2, 3, \ldots\}$, and S-partitions are sparser. Can more be said about the sets S of positive integers for which the coefficients of $\prod_{n\in S}(1-x^n)$ are all $1, 0$ or -1?

13. (Set partitions) A **set partition** of a set S is an unordered collection of pairwise disjoint sets $S_1, \ldots, S_k$ whose union is S. The family of set partitions with k parts is $\mathsf{SetPartition}^k \cong \bullet \times \mathsf{Seq}(\{1\}) \times \bullet \times \mathsf{Seq}(\{1,2\}) \times \cdots \times \bullet \times \mathsf{Seq}(\{1,2,\ldots,k\})$, where the singleton $\bullet$ and all numbers i have size 1. To see this, we regard a word such as $w = \bullet 11 \bullet 1221 \bullet 31$ as an instruction manual to build a set partition $S_1, \ldots, S_k$. The jth symbol w_j tells us where to put the number j: if w_j is a number h, we add j to the part S_h; if w_j is the ith $\bullet$, then we add j to a new part S_i. The sample word above leads to the partition $\{1,2,3,5,8,11\}, \{4,6,7\}, \{9,10\}$. This process is easily reversible. It follows

that
$$\sum_{n\geq 0} S(n,k)x^n = \frac{x}{1-x}\cdot\frac{x}{1-2x}\cdot\cdots\cdot\frac{x}{1-kx},$$
where $S(n,k)$ is the number of set partitions of $[n]$ into k parts. These numbers are called the **Stirling numbers of the second kind**.

The equation $(1-kx)\sum_{n\geq 0} S(n,k)x^n = x\sum_{n\geq 0} S(n,k-1)x^n$ gives the recurrence
$$S(n,k) = kS(n-1,k) + S(n-1,k-1), \qquad 1 \leq k \leq n,$$
with initial values $S(n,0) = S(n,n) = 1$. Note the great similarity with Pascal's recurrence.

14. (Catalan structures) It is often said that if you encounter a new family of mathematical objects, and you have to guess how many objects of size n there are, you should guess "the Catalan number $C_n = \frac{1}{n+1}\binom{2n}{n}$." The Catalan family has more than 200 incarnations in combinatorics and other fields [185, 192]; let us see three important ones.

    (a) (Plane binary trees) A **plane binary tree** is a rooted tree where every internal vertex has a left and a right child. Let PBTree be the family of plane binary trees, where a tree with n internal vertices (and necessarily $n+1$ leaves) has size n. A tree is either the trivial tree $\circ$ of size 0, or the grafting of a left subtree and a right subtree at the root $\bullet$, so $\mathsf{PBTree} \cong \circ + (\mathsf{PBTree} \times \bullet \times \mathsf{PBTree})$. It follows that the generating function for plane binary trees satisfies
    $$T(x) = 1 + T(x)xT(x).$$
    We may use the quadratic formula * and the binomial theorem to get
    $$T(x) = \frac{1-\sqrt{1-4x}}{2x} = \sum_{n\geq 0}\frac{1}{n+1}\binom{2n}{n}x^n.$$
    It follows that the number of plane binary trees with n internal vertices (and $n+1$ leaves) is the **Catalan number** $C_n = \frac{1}{n+1}\binom{2n}{n}$.

    (b) (Triangulations) A **triangulation** of a convex polygon is a subdivision into triangles using only the diagonals of P. A triangulation of an $(n+2)$-gon has n triangles; we say it has size n. If we fix an edge e of P, then a triangulation of P is obtained by choosing the triangle T that will cover e, and then choosing a triangulation of the two polygons to the left and to the right of T. Therefore $\mathsf{Triang} \cong \circ + (\mathsf{Triang} \times \bullet \times \mathsf{Triang})$ and the number of triangulations of an $(n+2)$-gon is also the Catalan number C_n.

*Since this is the first time we are using the quadratic formula, let us do it carefully. Rewrite the equation as $(1-2xT(x))^2 = 1-4x$, or $(1-2xT(x)-\sqrt{1-4x})(1-2xT(x)+\sqrt{1-4x}) = 0$. Since $\mathbb{C}[[x]]$ is an integral domain, one of the factors must be 0. From the constant coefficients we see that it must be the first factor.

(c) (Dyck paths) A **Dyck path** P of length n is a path from $(0,0)$ to $(2n,0)$ that uses the steps $(1,1)$ and $(1,-1)$ and never goes below the x-axis. Say P is **irreducible** if it touches the x-axis exactly twice, at the beginning and at the end. Let $D(x)$ and $I(x)$ be the generating functions for Dyck paths and irreducible Dyck paths.

A Dyck path is equivalent to a sequence of **irreducible** Dyck paths. Also, an irreducible path of length n is the same as a Dyck path of length $n-1$ with an additional initial and final step. Therefore

$$D(x) = \frac{1}{1-I(x)}, \qquad I(x) = xD(x)$$

from which it follows that $D(x) = \frac{1-\sqrt{1-4x}}{2x}$ as well, and the number of Dyck paths of length n is also the Catalan number.

Generatingfunctionology gives us fairly easy algebraic proofs that these three families are enumerated by the Catalan numbers. Once we have discovered this fact, the temptation to search for nice bijections is hard to resist.

Our algebraic analysis suggests a bijection ϕ from (b) to (a). The families of plane binary trees and triangulations grow under the same recursive recipe, and so we can let the bijection grow with them, mapping a triangulation $T \times \bullet \times T'$ to the tree $\phi(T) \times \bullet \times \phi(T')$. A non-recursive description of the bijection is the following. Consider a triangulation T of the polygon P, and fix an edge e. Put a vertex inside each triangle of T, and a vertex outside P next to each edge other than e. Then connect each pair of vertices separated by an edge. Finally, root the resulting tree at the vertex adjacent to e. This bijection is illustrated in Figure 1.4.

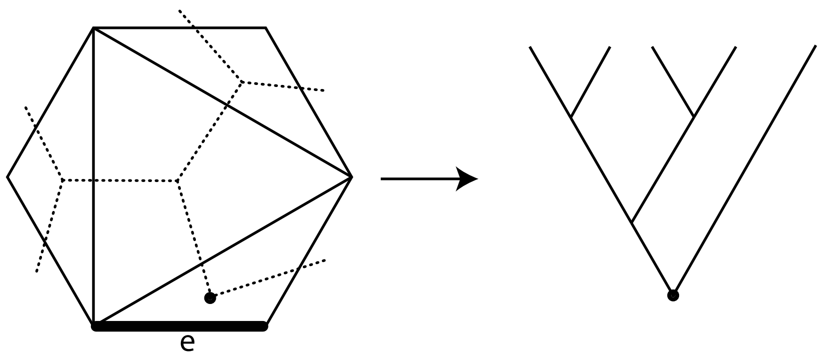


Figure 1.4
The bijection from triangulations to plane binary trees.

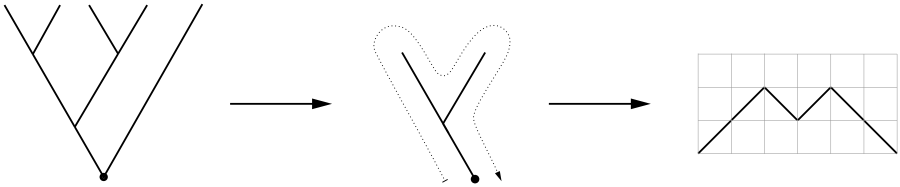


Figure 1.5
The bijection from plane binary trees to Dyck paths.

A bijection from (a) to (c) is less obvious from our algebraic computations, but is still not difficult to obtain. Given a plane binary tree T of size n, prune all the leaves to get a tree T' with n vertices. Now walk around the periphery of the tree, starting on the left side from the root, and continuing until we traverse the whole tree. Record the walk in a Dyck path $D(T)$: every time we walk up (respectively, down) a branch we take a step up (respectively, down) in $D(T)$. One easily checks that this is a bijection. See Figure 1.5 for an illustration.

Even if it may be familiar, it is striking that two different (and straightforward) algebraic computations show us that two families of objects that look quite different are in fact equivalent combinatorially. Although a simple, elegant bijection can often explain the connection between two families more transparently, the algebraic approach is sometimes simpler, and better at discovering such connections.

15. (k-Catalan structures) Let PTree_k be the class of **plane k-ary trees**, where every vertex that is not a leaf has k ordered children; let the size of such a tree be its number of leaves. In the sense of Theorem 1.3.2.5(a), this is precisely an $\mathscr{A}$-tree, where $\mathscr{A} = \{\bullet, \bullet^k\}$ consisting of one structure of size 1 and one of size k. Therefore $\mathsf{PTree}_k = (x - x^k)^{<-1>}$. Lagrange inversion then gives

$$m[x^m]A^{<-1>}(x) = [x^{m-1}]\left(\frac{1}{1-x^{k-1}}\right)^m = [x^{m-1}]\sum_{n\geq 0}\binom{m+n-1}{n}x^{(k-1)n}.$$

It follows that a plane k-ary tree must have $m = (k-1)n+1$ leaves for some integer n, and the number of such trees is the **k-Catalan number**

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

This is an alternative way to compute the ordinary Catalan numbers $C_n = C_n^2$.

The k-Catalan number C_n^k also has many different interpretations [100]; we mention two more. It counts the subdivisions of an $(n(k-1)+2)$-gon P into (necessarily n) $(k+1)$-gons using diagonals of P, and the paths from $(0,0)$ to $(n,(k-1)n)$ with steps $(0,1)$ and $(1,0)$ that never rise above the line $y = (k-1)x$.

Other applications. Let us now discuss a few other interesting applications that illustrate the power of Theorem 1.3.2.

16. (Motzkin paths) The **Motzkin number** M_n is the number of paths from $(0,0)$ to $(n,0)$ using the steps $(1,1)$, $(1,-1)$, and $(1,0)$, which never go below the x-axis. Imitating our argument for Dyck paths, we obtain a formula for the generating function

$$M(x) = \frac{1}{1-(x+xM(x)x)} \qquad \Longrightarrow \qquad M(x) = \frac{1-x-\sqrt{1-2x-3x^2}}{2x^2}.$$

The quadratic equation $x^2M^2+(x-1)M+1=0$ gives rise to the quadratic recurrence $M_n = M_{n-1}+\sum_i M_iM_{n-2-i}$. The fact that $M(x)$ satisfies a polynomial equation leads to a more efficient recurrence

$$(n+2)M_n = (2n+1)M_{n-1}+(3n-3)M_{n-2}.$$

We will see this in Section 1.3.4.2.

17. (Schröder paths) The (large) **Schröder number** r_n is the number of paths from $(0,0)$ to $(2n,0)$ using steps $NE=(1,1), SE=(1,-1)$, and $E=(2,0)$ which stays above the x-axis. Their generating function satisfies $R(x) = 1/(1-x-xR(x))$, and therefore

$$R(x) = \frac{1-x-\sqrt{1-6x+x^2}}{2x}.$$

Theorem 1.3.2.3 is useful when we are counting combinatorial objects that "factor" uniquely into an ordered "product" of "irreducible" objects. It tells us that we can count all objects if and only if we can count the irreducible ones. We have already used that idea several times; let us see it in action in some other examples.

18. (Domino tilings of rectangles) In Section 1.2 we let a_n be the number of domino tilings of a $2\times n$ rectangle. Such a tiling is uniquely a sequence of blocks, where each block is either a vertical domino (of width 1) or two horizontal dominoes (of width 2). This truly explains the formula:

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

Similarly, if $a_{m,n}$ is the number of domino tilings of a $2\times n$ rectangle using v vertical tiles, we immediately obtain

$$\sum_{m,n\geq 0} a_{m,n}v^mx^n = \frac{1}{1-(vx+x^2)}.$$

Sometimes the enumeration of irreducible structures is not immediate, but still tractable.

19. (Monomer-dimer tilings of rectangles) Let $T(2,n)$ be the number of tilings of a $2 \times n$ rectangle with dominoes and unit squares. Say a tiling is irreducible if it does not contain an internal vertical line from top to bottom. Then $\mathsf{Tilings} \cong \mathsf{Seq}(\mathsf{IrredTilings})$. It now takes some thought to recognize the irreducible tilings in Figure 1.6.

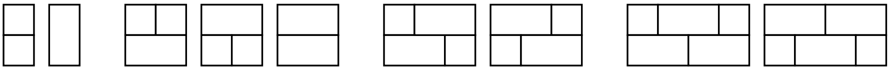


Figure 1.6
The irreducible tilings of $2 \times n$ rectangles into dominoes and unit squares.

There are three irreducible tilings of length 2, and two of every other length greater than or equal to 1. Therefore

$$\sum_{n \geq 0} T(2,n)x^n = \frac{1}{1-(2x+3x^2+2x^3+2x^4+\cdots)} = \frac{1-x}{1-3x-x^2+x^3}.$$

We will see in Theorem 1.3.5.2 that this gives $T(2,n) \sim c \cdot \alpha^n$ where $\alpha \approx 3.214\ldots$ is the inverse of the smallest positive root of the denominator.

Sometimes the enumeration of **all** objects is easier than the enumeration of the irreducible ones. In that case we can use Theorem 1.3.2.3 in the opposite direction.

20. (Irreducible permutations) A permutation π of $[n]$ is **irreducible** if it does not factor as a permutation of $\{1,\ldots,m\}$ and a permutation of $\{m+1,\ldots,n\}$ for $1 \leq m < n$; that is, if $\pi([m]) \neq [m]$ for all $1 \leq m < n$. Clearly every permutation factors uniquely into irreducibles, so

$$\sum_{n \geq 0} n!x^n = \frac{1}{1-\mathsf{IrredPerm}(x)}.$$

This gives the series for $\mathsf{IrredPerm}$.

There are many interesting situations where it is possible, but not at all trivial, to decompose the objects that interest us into simpler structures. To a combinatorialist this is good news, the techniques of this section are useful tools, but are not enough. There is no shortage of interesting work to do. Here is a great example.

21. (Domino towers) [91, 31, 221] A **domino tower** is a stack of horizontal 2×1 bricks in a brickwork pattern, so that no brick is directly above another brick, such that the bricks on the bottom level are contiguous, and every higher brick is (half) supported on at least one brick in the row below it. Let the size of a domino tower be the number of bricks. See Figure 1.7 for an illustration.

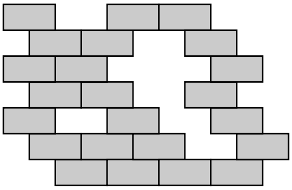


Figure 1.7
A domino tower of 19 bricks.

Remarkably, there are 3^{n-1} domino towers consisting of n bricks. Equally remarkably, no simple bijection is known. The nicest argument known is as follows.

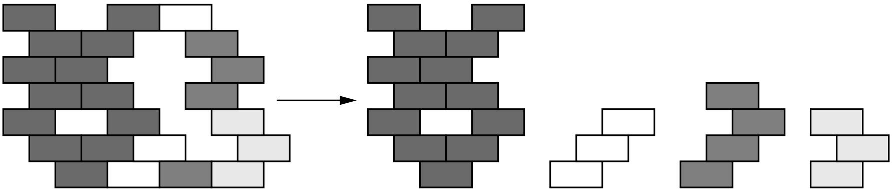


Figure 1.8
The decomposition of a domino tower into a pyramid and three half-pyramids.

We decompose a domino tower x into smaller pieces, as illustrated in Figure 1.8. Each new piece is obtained by pushing up the leftmost remaining brick in the bottom row, dragging with it all the bricks encountered along the way. The first piece p will be a **pyramid**, which we define to be a domino tower with only one brick in the bottom row. All subsequent pieces $h_1, \ldots, h_k$ are **half-pyramids**, which are pyramids containing no bricks to the left of the bottom brick. This decomposition is reversible. To recover x, we drop $h_k, h_{k-1}, \ldots, h_1, p$ from the top in that order; each piece is dropped in its correct horizontal position, and some of its bricks may get stuck on the previous pieces. This shows that the corresponding combinatorial classes satisfy $X \cong P \times \mathsf{Seq}(H)$.

Similarly, we may decompose a pyramid p into half-pyramids, as shown in Figure 1.9. Each new half-pyramid is obtained by pushing up the leftmost remaining brick (which is not necessarily in the bottom row), dragging with it all the bricks that it encounters along the way. This shows that $P \cong \mathsf{Seq}_{\geq 1}(H) := H + (H \times H) + \cdots$.

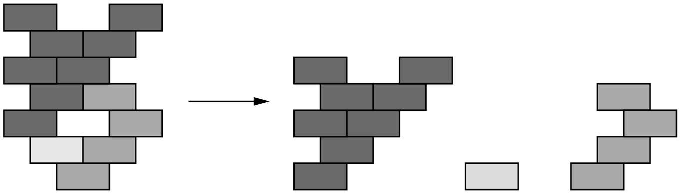


Figure 1.9
A pyramid and its decomposition into half-pyramids.

Finally consider a half-pyramid h; there are two cases. If there are other bricks on the same horizontal position as the bottom brick, consider the lowest such brick, and push it up, dragging with it all the bricks it encounters along the way, obtaining a half-pyramid h_1. Now remove the bottom brick; what remains is a half-pyramid h_2. This is shown in Figure 1.10. As before, we can recover h from h_1 and h_2. On the other hand, if there are no bricks above the bottom brick, removing the bottom brick leaves either a half-pyramid or the empty set. Therefore $H \cong (H \times \bullet \times H) + (\bullet \times H) + \bullet$.

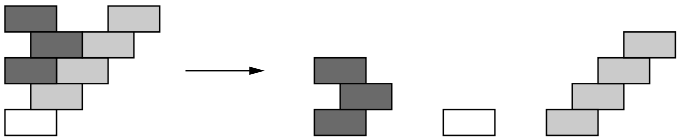


Figure 1.10
A (non-half-pyramid) pyramid and its decomposition into two half-pyramids and a bottom brick.

The above relations correspond to the following identities for the corresponding generating functions:

$$X = \frac{P}{1-H}, \qquad P = \frac{H}{1-H}, \qquad H = xH^2 + xH + x.$$

Surprisingly cleanly, we obtain $X(x) = x/(1-3x) = \sum_{n\geq 1} 3^{n-1}x^n$. This proves that there are 3^{n-1} domino towers of size n.

Although we do not need this here, it is worth noting that half-pyramids are enumerated by Motzkin numbers; their generating functions are related by $H(x) = xM(x)$.

1.3.3 Exponential generating functions

Ordinary generating functions are usually not well suited for counting combinatorial objects with a labeled ground set. In such situations, exponential generating functions are a more effective tool.

Consider a family $\mathscr{A} = \mathscr{A}_0 \sqcup \mathscr{A}_1 \sqcup A_2 \sqcup \cdots$ of labeled combinatorial structures, where $\mathscr{A}_n$ consists of the structures that we can place on the ground set $[n] = \{1,\ldots,n\}$ (or, equivalently, on any other labeled ground set of size n). If $a \in \mathscr{A}_n$, we let $|a| = n$ be the **size** of a. We also let a_n be the number of elements of size n. The **exponential generating function** of $\mathscr{A}$ is

$$A(x) = \sum_{a\in\mathscr{A}} \frac{x^{|a|}}{|a|!} = a_0\frac{x^0}{0!} + a_1\frac{x^1}{1!} + a_2\frac{x^2}{2!} + a_3\frac{x^3}{3!} + \cdots.$$

We may again assign a weight $\mathrm{wt}(a)$ to each object a, usually a monomial in variables $x_1,\ldots,x_n$, and consider the **weighted exponential generating function** of $\mathscr{A}$ to be the formal power series

$$A_{\mathrm{wt}}(x_1,\ldots,x_n,x) = \sum_{a\in\mathscr{A}} \mathrm{wt}(a)\frac{x^{|a|}}{|a|!}.$$

Examples of combinatorial structures (with their respective size functions) are permutations (number of elements), graphs (number of vertices), or set partitions (size of the set). We may weight these objects by t^k where k is, respectively, the number or cycles, the number of edges, or the number of parts.

1.3.3.1 Operations on labeled structures and their exponential generating functions

Again, there are some simple operations on *labeled* combinatorial structures, which correspond to simple algebraic operations on the exponential generating functions. Starting with a few simple structures, these operations are sufficient to generate many interesting combinatorial structures. This will allow us to compute the exponential generating functions for those structures.

Theorem 1.3.3 *Let $\mathscr{A}$ and $\mathscr{B}$ be labeled combinatorial structures.*

1. *($\mathscr{C} = \mathscr{A} + \mathscr{B}$: Disjoint union) If a $\mathscr{C}$-structure on a finite set S is obtained by choosing an $\mathscr{A}$-structure on S **or** a $\mathscr{B}$-structure on S, then*

$$C(x) = A(x) + B(x).$$

This result also holds for weighted structures if the weight of a $\mathscr{C}$-structure is the same as the weight of the respective $\mathscr{A}$- or $\mathscr{B}$-structure.

2. *($\mathscr{C} = \mathscr{A} * \mathscr{B}$: Labeled Product) If a $\mathscr{C}$-structure on a finite set S is obtained by partitioning S into disjoint sets S_1 and S_2 and putting an $\mathscr{A}$-structure on S_1 **and** a $\mathscr{B}$-structure on S_2, then*

$$C(x) = A(x)B(x).$$

This result also holds for weighted structures if the weight of a $\mathscr{C}$-structure is the product of the weights of the respective $\mathscr{A}$- and $\mathscr{B}$-structures.

3. *($\mathscr{C} = \mathsf{Seq}_*(\mathscr{B})$: Labeled Sequence) If a $\mathscr{C}$-structure on a finite set S is obtained by choosing an **ordered** partition of S into a sequence of blocks and putting a $\mathscr{B}$-structure on each block, then*
$$C(x) = \frac{1}{1-B(x)}.$$
This result also holds for weighted structures if the weight of a $\mathscr{C}$-structure is the product of the weights of the respective $\mathscr{B}$-structures.

4. *($\mathscr{C} = \mathsf{Set}(\mathscr{B})$: Set) **Exponential formula.** If a $\mathscr{C}$-structure on a finite set S is obtained by choosing an **unordered** partition of S into a set of blocks and putting a $\mathscr{B}$-structure on each block, then*
$$C(x) = e^{B(x)}.$$
This result also holds for weighted structures if the weight of a $\mathscr{C}$-structure is the product of the weights of the respective $\mathscr{B}$-structures.
In particular, if $c_k(n)$ is the number of $\mathscr{C}$-structures of an n-set that decompose into k components ($\mathscr{B}$-structures), we have
$$\sum_{n,k,\geq 0} c_k(n)\frac{x^n}{n!}y^k = e^{yB(x)} = C(x)^y$$

5. *($\mathscr{C} = \mathscr{A} \circ \mathscr{B}$: Composition) **Compositional formula.** If a $\mathscr{C}$-structure on a finite set S is obtained by choosing an* unordered *partition of S into a set of blocks, putting a $\mathscr{B}$-structure on each block, and putting an $\mathscr{A}$-structure on the set of blocks, then*
$$C(x) = A(B(x)).$$
This result also holds for weighted structures if the weight of a $\mathscr{C}$-structure is the product of the weights of the $\mathscr{A}$-structure on its set of blocks and the weights of the $\mathscr{B}$-structures on the individual blocks.

Again, Theorem 1.3.4.4 is useful when we are counting combinatorial objects that “decompose” uniquely as a set of “indecomposable” objects. It tells us that we can count all objects if and only if we can count the indecomposable ones. Amazingly, we also obtain for free the finer enumeration of the objects by their number of components.

Proof. 1. is clear. The identity in 2. is equivalent to $c_n = \sum_k \binom{n}{k} a_k b_{n-k}$, which corresponds to the given combinatorial description. Iterating 2., we see that the exponential generating functions for k-sequences of $\mathscr{B}$-structures is $B(x)^k$, and hence the one for k-sets of $\mathscr{B}$-structures is $B(x)^k/k!$. This readily implies 3, 4, and 5. The weighted statements follow similarly. ∎

The following statements are perhaps less fundamental, but also useful.

Theorem 1.3.4 *Let $\mathscr{A}$ be a labeled combinatorial structure.*

1. *($\mathscr{C} = \mathscr{A}_+$: Shifting) If a $\mathscr{C}$-structure on S is obtained by adding a new element t to S and choosing an $\mathscr{A}$-structure on $S \cup \{t\}$, then*

$$C(x) = A'(x).$$

2. *($\mathscr{C} = \mathscr{A}_\bullet$: Rooting) If a $\mathscr{C}$-structure on S is a* rooted $\mathscr{A}$-structure*, obtained by choosing an $\mathscr{A}$-structure on S and an element of S called the* **root***, then*

$$C(x) = xA(x).$$

3. *(Sieving by parity of size) If the $\mathscr{C}$-structures are precisely the $\mathscr{A}$-structures of even size,*

$$C(x) = \frac{A(x) + A(-x)}{2}.$$

4. *(Sieving by parity of components) Suppose $\mathscr{A}$-structures decompose uniquely into components, so $\mathscr{A} = \mathsf{Set}(\mathscr{B})$ for some $\mathscr{B}$. If the $\mathscr{C}$-structures are the $\mathscr{A}$-structures having only components of even size,*

$$C(x) = \sqrt{A(x)A(-x)}.$$

5. *(Sieving by parity of number of components) Suppose $\mathscr{A}$-structures decompose uniquely into components, so $\mathscr{A} = \mathsf{Set}(\mathscr{C})$ for some $\mathscr{C}$. If the $\mathscr{C}$-structures are precisely the $\mathscr{A}$-structures having an even number of components,*

$$C(x) = \frac{1}{2}\left(A(x) + \frac{1}{A(x)}\right).$$

Similar sieving formulas hold modulo k for any $k \in \mathbb{N}$.

Proof. We have $c_n = a_{n+1}$ in 1., $c_n = na_n$ in 2., and $c_n = \frac{1}{2}(a_n + (-1)^n a_n)$ in 3., from which the generating function formulas follow. Combining 3. with the Exponential Formula we obtain 4. and 5.

Similarly we see that the generating function for $\mathscr{A}$-structures whose size is a multiple of k is $\frac{1}{k}\left(A(x) + A(\omega x) + \cdots + A(\omega^{k-1}x)\right)$ where ω is a primitive kth root of unity. If we wish to count elements of size i mod k, we use 1. to shift this generating function i times. ■

1.3.3.2 Examples

Classical applications. Once again, these simple ideas give very easy solutions to many classical enumeration problems.

1. (Trivial classes) Again we consider the trivial classes $\circ$ with only one element of size 0, and $\bullet$ with only one element of size 1. Their exponential generating functions are 1 and x, respectively.

2. (Sets) A slightly less trivial class of Set contains one set of each size. We also let $\mathsf{Set}_{\geq 1}$ denote the class of non-empty sets, with generating function $e^x - 1$. The exponential generating functions are

$$\mathsf{Set}(x) = e^x, \qquad \mathsf{Set}_{\geq 1}(x) = e^x - 1.$$

3. (Set partitions) In Section 1.3.2.2 we found the ordinary generating function for Stirling numbers $S(n,k)$ for a given k; but in fact it is easier to use exponential generating functions. Simply notice that $\mathsf{SetPartition} \cong \mathsf{Set}(\mathsf{Set}_{\geq 1})$, and the Weighted Exponential Formula then gives

$$\mathsf{SetPartition}(x,y) = \sum_{n,k \geq 0} S(n,k) \frac{x^n}{n!} y^k = e^{y(e^x - 1)}.$$

4. (Permutations) Let Perm_n consist of the $n!$ permutations of $[n]$. A permutation is a labeled sequence of singletons, so $\mathsf{Perm} = \mathsf{Seq}_*(\bullet)$, and the generating function for permutations is

$$\mathsf{Perm}(x) = \sum_{n \geq 0} n! \frac{x^n}{n!} = \frac{1}{1-x}.$$

5. (Cycles) Let Cycle_n consist of the cyclic orders of $[n]$. These are the ways of arranging $1, \ldots, n$ around a circle, where two orders are the same if they differ by a rotation of the circle. There is an n-to-1 mapping from permutations to cyclic orders obtained by wrapping a permutation around a circle, so

$$\mathsf{Cycle}(x) = \sum_n (n-1)! x^n / n! = -\log(1-x).$$

   There is a more indirect argument that will be useful to us later. Recall that a permutation π can be written uniquely as a (commutative) product of disjoint cycles of the form $(i, \pi(i), \pi^2(i), \ldots, \pi^{k-1}(i))$, where k is the smallest index such that $\pi^k(i) = i$. For instance, the permutation 835629741 can be written in cycle notation as $(18469)(235)(7)$. Then $\mathsf{Perm} = \mathsf{Set}(\mathsf{Cycle})$ so $1/(1-x) = e^{\mathsf{Cycle}(x)}$.

6. (Permutations by number of cycles) The (signless) **Stirling number of the first kind** $c(n,k)$ is the number of permutations of n having k cycles. The Weighted Exponential Formula gives

$$\sum_{n,k \geq 0} c(n,k) \frac{x^n}{n!} y^k = e^{y \mathsf{Cycle}(x)} = \left(\frac{1}{1-x} \right)^y = \sum_{n \geq 0} y(y+1) \cdots (y+n-1) \frac{x^n}{n!}$$

   so the Stirling numbers $c(n,k)$ of the first kind are the coefficients of the polynomial $y(y+1) \cdots (y+n-1)$.

Other applications. The applications of these techniques are countless; let us consider a few more applications, old and recent.

7. (Permutations by cycle type) The **type** of a permutation $\pi \in S_n$ is $\mathrm{type}(w) = (c_1, \ldots, c_n)$ where c_i is the number of cycles of length i. For indeterminates $\mathbf{t} = (t_1, \ldots, t_n)$, let $\mathbf{t}^{\mathrm{type}(w)} = t_1^{c_1} \cdots t_n^{c_n}$. The **cycle indicator** of the symmetric group S_n is $Z_n = \frac{1}{n!}\sum_{w \in S_n} \mathbf{t}^{\mathrm{type}(w)}$. The Weighted Exponential Formula immediately gives
$$\sum_{n \geq 0} Z_n x^n = e^{t_1 x + t_2 x^2/2 + t_3 x^3/3 + \cdots}.$$
Let us discuss two special cases of interest.

8. (Derangements) A **derangement** of $[n]$ is a permutation such that $\pi(i) \neq i$ for all $i \in [n]$. Equivalently, a derangement is a permutation with no cycles of length 1. It follows that $\mathsf{Derangement} = \mathsf{Set}(\mathsf{Cycle}_{\geq 2})$, so the number d_n of derangements of $[n]$ is given by
$$\mathsf{Derangement}(x) = \sum_{n \geq 0} d_n \frac{x^n}{n!} = e^{-\log(1-x) - x} = e^{-x} + xe^{-x} + x^2 e^{-x} + \cdots,$$
which leads to the explicit formula
$$d_n = n!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots \pm \frac{1}{n!}\right) \sim \frac{n!}{e}.$$

9. (Involutions) An **involution** of $[n]$ is a permutation w such that w^2 is the identity. Equivalently, an involution is a permutation with cycles of length 1 and 2, so the number i_n of involutions of $[n]$ is given by
$$\mathsf{Inv}(x) = \sum_{n \geq 0} i_n \frac{x^n}{n!} = e^{x + \frac{x^2}{2}}.$$
Note that $\mathsf{Inv}'(x) = (x+1)\mathsf{Inv}(x)$, which gives $i_n = i_{n-1} + (n-1)i_{n-2}$. In Section 1.3.4.2 we will explain the theory of D-finite power series, which turns differential equations for power series into recurrences for the corresponding sequences.

10. (Trees) A **tree** is a connected graph with no cycles. Consider a "birooted" tree (T, a, b) on $[n]$ with two (possibly equal) root vertices a and b. Regard the unique path $a = v_0, v_1, \ldots, v_k = b$ as a "spine" for T; the rest of the tree consists of rooted trees hanging from the v_is; direct their edges toward the spine. Now regard $v_1 \ldots v_k$ as a permutation in one-line notation, and rewrite it in cycle notation, while continuing to hang the rooted trees from the respective v_i's. This transforms (T, a, b) into a directed graph consisting of a disjoint collection of cycles with trees directed toward them. Every vertex has outdegree 1, so this defines a function $f : [n] \to [n]$. A moment's thought will convince us that this is a bijection. Therefore there are n^n birooted trees on $[n]$, and hence there are n^{n-2} trees on $[n]$. See Figure 1.11 for an illustration.

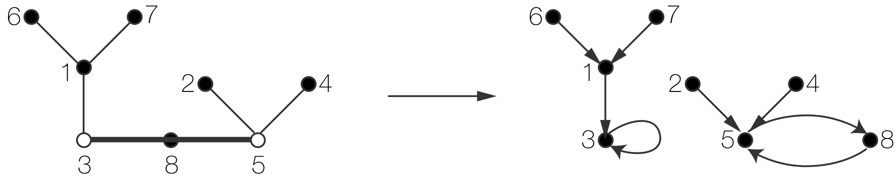


Figure 1.11
A tree on $[8]$ birooted at $a = 3$ and $b = 5$, and the corresponding function $f : [8] \to [8]$.

11. (Trees, revisited) Let us count trees in a different way. Let a **rooted tree** be a tree with a special vertex called the root, and a **planted forest** be a graph with no cycles where each connected component has a root. Let t_n, r_n, f_n and $T(x), R(x), F(x)$ be the sequences and exponential generating functions enumerating trees, rooted trees, and planted forests, respectively.

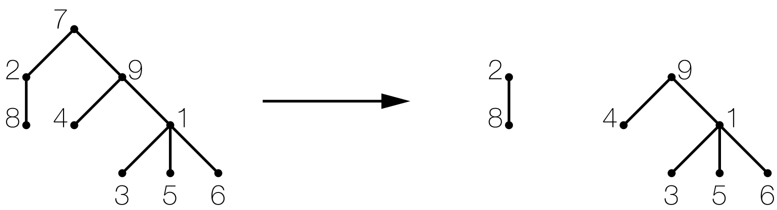


Figure 1.12
A rooted tree seen as a root attached to the roots of a planted forest.

Planted forests are vertex-disjoint unions of rooted trees, so $F(x) = e^{R(x)}$. Also, as illustrated in Figure 1.12, a rooted tree T consists of a root attached to the roots of a planted forest, so $R(x) = xF(x)$. It follows that $x = R(x)e^{-R(x)}$, so

$$R(x) = (xe^{-x})^{<-1>}.$$

Lagrange inversion gives $n \cdot \frac{r_n}{n!} = [x^{n-1}]e^{nx} = \frac{n^{n-1}}{(n-1)!}$, so

$$r_n = n^{n-1}, \qquad f_n = (n+1)^{n-1}, \qquad t_n = n^{n-2}.$$

We state a finer enumeration; see [192, Theorem 5.3.4] for a proof. The **degree sequence** of a rooted forest on $[n]$ is $(\deg 1, \ldots, \deg n)$ where $\deg i$ is the number of children of i. For example the degree sequence of the rooted tree in Figure 1.12 is $(3,1,0,0,0,0,2,0,2)$. Then the number of planted forests with a given degree sequence $(d_1, \ldots, d_n)$ and (necessarily) $k = n - (d_1 + \cdots + d_n)$

components is

$$\binom{n-1}{k-1}\binom{n-k}{d_1,\ldots,d_n}.$$

The number of forests on $[n]$ is given by a more complicated alternating sum; see [197].

12. (Permutations, revisited) Here is an unnecessarily complicated way of proving there are $n!$ permutations of $[n]$. A permutation π of $[n+1]$ decomposes uniquely as a concatenation $\pi = L(n+1)R$ for permutations L and R of two complementary subsets of $[n]$. Therefore $\mathsf{Shift}(\mathsf{Perm}) = (\mathsf{Perm}) * (\mathsf{Perm})$, and the generating function $P(x)$ for permutations satisfies $P'(x) = P(x)^2$ with $P(0) = 1$. Solving this differential equation gives $P(x) = \frac{1}{1-x} = \sum_{n\geq 0} n!\frac{x^n}{n!}$.

13. (Alternating permutations) The previous argument was gratuitous for permutations, but it will now help us to enumerate the class Alt of **alternating** permutations w, which satisfy $w_1 < w_2 > w_3 < w_4 > \cdots$. The **Euler numbers** are $E_n = |\mathsf{Alt}_n|$; let $E(x)$ be their exponential generating function. We will need the class RevAlt of permutations w with $w_1 > w_2 < w_3 > w_4 < \cdots$. The map $w = w_1 \ldots w_n \mapsto w' = (n+1-w_1)\ldots(n+1-w_n)$ on permutations of $[n]$ shows that $\mathsf{Alt} \cong \mathsf{RevAlt}$.

    Now consider alternating permutations L and R of two complementary subsets of $[n]$. For $n \geq 1$, exactly one of the permutations $L(n+1)R$ and $L'(n+1)R$ is alternating or reverse alternating, and every such permutation arises uniquely in that way. For $n = 0$ both are alternating. Therefore $\mathsf{Shift}(\mathsf{Alt} + \mathsf{RevAlt}) = \mathsf{Alt} * \mathsf{Alt} + \circ$, so $2E'(x) = E(x)^2 + 1$ with $E(0) = 1$. Solving this differential equation we get

$$E(x) = \sum_{n\geq 0} E_n \frac{x^n}{n!} = \sec x + \tan x.$$

    Therefore $\sec x$ and $\tan x$ enumerate the alternating permutations of even and odd length, respectively. The Euler numbers are also called secant and tangent numbers. This surprising connection allows us to give combinatorial interpretations of various trigonometric identities, such as $1 + \tan^2 x = \sec^2 x$.

14. (Graphs) Let $g(v)$ and $g_{\text{conn}}(v)$ be the number of simple graphs and connected graphs on $[v]$, respectively. (A graph is simple if it contains no loops and no multiple edges.) The Exponential Formula tells us that their exponential generating functions are related by $G(x) = e^{G_{\text{conn}}(x)}$. In this case it is hard to count the connected graphs directly, but it is easy to count all graphs: to choose a graph we just have to decide whether each edge is present or not, so $g(v) = 2^{\binom{v}{2}}$. This gives us

$$\sum_{v\geq 0} g_{\text{conn}}(v)\frac{x^v}{v!} = \log\left(\sum_{v\geq 0} 2^{\binom{v}{2}}\frac{x^v}{v!}\right).$$

    We may easily adjust this computation to account for edges and components.

There are $\binom{v(v-1)/2}{e}$ graphs on $[v]$ with e edges; say $g(v,c,e)$ of them have c components, and give them weight $y^c z^e$. Then

$$\sum_{v,c,e \geq 0} g(v,c,e)\frac{x^v}{v!}y^c z^e = \left(\sum_{v,e\geq 0} \binom{\binom{v}{2}}{e}\frac{x^v}{v!}z^e\right)^y = F(x,1+z)^y$$

where

$$F(\alpha,\beta) = \sum_{n\geq 0} \frac{\alpha^n \beta^{\binom{n}{2}}}{n!}$$

is the **deformed exponential function** of [179].

15. (Signed graphs) A **signed graph** G is a set of vertices, with at most one "positive" edge and one "negative" edge connecting each pair of vertices. We say G is *connected* if and only if its underlying graph $\overline{G}$ (ignoring signs) is connected. A *cycle* in G corresponds to a cycle of $\overline{G}$; we call it *balanced* if it contains an even number of negative edges, and *unbalanced* otherwise. We say that G is *balanced* if all its cycles are balanced. Let $s(v,c_+,c_-,e)$ be the number of signed graphs with v vertices, e edges, c_+ balanced components, and c_- unbalanced components; we will need the generating function

$$S(x,y_+,y_-,z) = \sum_{G \text{ signed graph}} s(v,c_+,c_-,e)\,\frac{x^v}{v!}y_+^{c_+}\,y_-^{c_-}\,z^e$$

in order to carry out a computation in Section 1.8.9; we follow [11].

Let $S(x,y_+,y_-,z)$, $B(x,y_+,z)$, $C_+(x,z)$, and $C_-(x,z)$ be the generating functions for signed, balanced, connected balanced, and connected unbalanced graphs, respectively. The Weighted Exponential Formula gives

$$B = e^{y_+C_+}, \qquad S = e^{y_+C_+ + y_-C_-},$$

so if we can compute C_+ and C_- we will obtain B and S. In turn, these equations give

$$C_+(x,z) = \frac{1}{2}\log B(x,2,z), \qquad C_+(x,z)+C_-(x,z) = \log S(x,1,1,z),$$

and we now compute the right-hand side of these two equations. (In the first equation, we set $t_+ = 2$ because, surprisingly, $B(x,2,z)$ is easier to compute than $B(x,1,z)$.) One is easy:

$$S(x,1,1,z) = \sum_{e,v\geq 0} \binom{v(v-1)}{e}\frac{x^v}{v!}z^e = F(x,(1+z)^2).$$

For the other one, we count balanced signed graphs by relating them with **marked graphs**, which are simple graphs with a sign $+$ or $-$ on each **vertex**. [95] A marked graph M gives rise to a balanced signed graph G by assigning to each edge the product of its vertex labels. Furthermore, if G has c

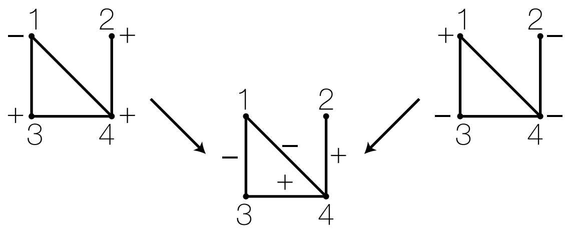


Figure 1.13
The two marked graphs that give rise to one balanced signed graph.

components, then it arises from precisely 2^c different marked graphs, obtained from M by choosing some connected components and changing their signs. This correspondence is illustrated in Figure 1.13. It follows that $B(x,2y,z) = \sum_{B \text{ balanced}} 2^c\, b(v,c,e) \frac{x^v}{v!} y^c z^e = \sum_{M \text{ marked}} m(v,c,e) \frac{x^v}{v!} y^c z^e$ is the generating function for marked graphs, and hence $B(x,2,z)$ may be computed easily:

$$B(x,2,z) = \sum_{e,v} \binom{\binom{v}{2}}{e} 2^v \frac{x^v}{v!} z^e = F(2x, 1+z).$$

Putting these equations together yields

$$S(x,y_+,y_-,z) = F(2x,1+z)^{(y_+-y_-)/2} F(x,(1+z)^2)^{y_-}.$$

1.3.4 Nice families of generating functions

In this section we discuss three nice properties that a generating function can have: being rational, algebraic, or D-finite. Each one of these properties gives rise to useful properties for the corresponding sequence of coefficients.

1.3.4.1 Rational generating functions

Many sequences in combinatorics and other fields satisfy three equivalent properties: They satisfy a recursive formula with constant coefficients, they are given by an explicit formula in terms of polynomials and exponentials, and their generating functions are rational. We understand these sequences very well. The following theorem tells us how to translate any one of these formulas into the others.

Theorem 1.3.5 [194, Theorem 4.1.1] *Let $a_0, a_1, a_2, \ldots$ be a sequence of complex numbers and let $A(x) = \sum_{n \geq 0} a_n x^n$ be its ordinary generating function. Let $q(x) = 1 + c_1 x + \cdots + c_d x^d = (1-r_1x)^{d_1} \cdots (1-r_kx)^{d_k}$ be a complex polynomial of degree d. The following are equivalent:*

1. *The sequence satisfies the linear recurrence with constant coefficients*

$$a_n + c_1 a_{n-1} + \cdots + c_d a_{n-d} = 0 \qquad (n \geq d).$$

2. *There exist polynomials* $f_1(x), \ldots, f_k(x)$ *with* $\deg f_i(x) < d_i$ *for* $1 \leq i \leq n$ *such that*
$$a_n = f_1(n)\, r_1^n + \cdots + f_k(n)\, r_k^n.$$

3. *There exists a polynomial* $p(x)$ *with* $\deg p(x) < d$ *such that* $A(x) = p(x)/q(x)$.

Notice that Theorem 1.3.5.2 gives us the asymptotic growth of a_n immediately. Let us provide more explicit recipes.

$(1 \Rightarrow 2)$ Extract the inverses r_i of the roots of $q(x) = 1 + c_1x + \cdots + c_dx^d$ and their multiplicities d_i. The $d_1 + \cdots + d_k = d$ coefficients of the f_is are the unknowns in the system of d linear equations $a_n = f_1(n)\, r_1^n + \cdots + f_k(n) r_k^n$ $(n = 0, 1, \ldots, d-1)$, which has a unique solution.

$(1 \Rightarrow 3)$ Read off $q(x) = 1 + c_1x + \cdots + c_dx^d$ from the recurrence; the coefficients of $p(x)$ are $[x^k]p(x) = a_k + c_1a_{k-1} + \cdots + c_da_{k-d}$ for $0 \leq k < d$, where $a_i = 0$ for $i < 0$.

$(2 \Rightarrow 1)$ Compute the c_is using $q(x) = \prod_i^k (1 - r_ix)^{\deg f_i + 1}$.

$(2 \Rightarrow 3)$ Let $q(x) = \prod_i (1 - r_ix)^{\deg f_i + 1}$, and compute the first k terms of $p(x) = A(x)q(x)$; the others are 0.

$(3 \Rightarrow 1)$ Extract the c_is from the denominator $q(x)$.

$(3 \Rightarrow 2)$ Compute the partial fraction decomposition $p(x)/q(x) = \sum_{i=1}^k p_i(x)/(1 - r_ix)^{d_i}$ where $\deg p_i(x) < d_i$ and use $(1 - r_ix)^{-d_i} = \sum_n \binom{d_i + n - 1}{d_i - 1} r_i^n x^n$ to extract $a_n = [x^n]p(x)/q(x)$.

Characterizing polynomials. As a special case of Theorem 1.3.5, we obtain a useful characterization of sequences given by a polynomial. The **difference operator** Δ acts on sequences, sending the sequence $\{a_n : n \in \mathbb{N}\}$ to the sequence $\{\Delta a_n : n \in \mathbb{N}\}$ where $\Delta a_n = a_{n+1} - a_n$.

Theorem 1.3.6 [194, Theorem 4.1.1] *Let* $a_0, a_1, a_2, \ldots$ *be a sequence of complex numbers and let* $A(x) = \sum_{n \geq 0} a_nx^n$ *be its ordinary generating function. Let* d *be a positive integer. The following are equivalent:*

1. *We have* $\Delta^{d+1}a_n = 0$ *for all* $n \in \mathbb{N}$.

2. *There exists a polynomial* $f(x)$ *with* $\deg f \leq d$ *such that* $a_n = f(n)$ *for all* $n \in \mathbb{N}$.

3. *There exists a polynomial* $p(x)$ *with* $\deg p(x) \leq d$ *such that* $A(x) = p(x)/(1-x)^{d+1}$.

We have already seen some combinatorial polynomials and generating functions whose denominator is a power of $1 - x$; we will see many more examples in the following sections.

1.3.4.2 Algebraic and D-finite generating functions

When the generating function $A(x) = \sum_n a_n x^n$ we are studying is not rational, the next natural question to ask is whether $A(x)$ is algebraic. If it is, then just as in the rational case, the sequence a_n still satisfies a linear recurrence, although now the coefficients are polynomial in n. This general phenomenon is best explained by introducing the wider family of "D-finite" (also known as "differentially finite" or "holonomic") power series. Let us discuss a quick example before we proceed to the general theory.

We saw that the ordinary generating function for the Motzkin numbers satisfies the quadratic equation

$$x^2M^2 + (x-1)M + 1 = 0 \tag{1.8}$$

which gives rise to the quadratic recurrence $M_n = M_{n-1} + \sum_i M_i M_{n-2-i}$ with $M_0 = 1$. This is not a bad recurrence, but we can find a better one. Differentiating (1.8) we can express M' in terms of M. Our likely first attempt leads us to $M' = -(2xM^2 + M)/(2x^2M + x - 1)$, which is not terribly enlightening. However, using (1.8) and a bit of purposeful algebraic manipulation, we can rewrite this as a linear equation with polynomial coefficients:

$$(x - 2x^2 - 3x^3)M' + (2 - 3x - 3x^2)M - 2 = 0.$$

Extracting the coefficient of x^n we obtain the much more efficient recurrence relation

$$(n+2)M_n - (2n+1)M_{n-1} - (3n-3)M_{n-2} = 0. \qquad (n \geq 2)$$

We now explain the theoretical framework behind this example.

Rational, algebraic, and D-finite series. Consider a formal power series $A(x)$ over the complex numbers. We make the following definitions.

$A(x)$ is **rational**	There exist polynomials $p(x)$ and $q(x) \neq 0$ such that $$q(x)A(x) = p(x).$$
$A(x)$ is **algebraic**	There exist polynomials $p_0(x), \ldots, p_d(x)$ such that $$p_0(x) + p_1(x)A(x) + p_2(x)A(x)^2 + \cdots + p_d(x)A(x)^d = 0.$$
$A(x)$ is **D-finite**	There exist polynomials $q_0(x), \ldots, q_d(x), q(x)$ such that $$q_0(x)A(x) + q_1(x)A'(x) + q_2(x)A''(x) + \cdots + q_d(x)A^{(d)}(x) = q(x).$$

Now consider the corresponding sequence $a_0, a_1, a_2 \ldots$ and make the following definitions.

$\{a_0, a_1, \ldots\}$ is **c-recursive**	There are constants $c_0, \ldots, c_d \in \mathbb{C}$ such that for all $n \geq d$ $$c_0 a_n + c_1 a_{n-1} + \cdots + c_d a_{n-d} = 0$$
$\{a_0, a_1, \ldots\}$ is **P-recursive**	There are complex polynomials $c_0(x), \ldots, c_d(x)$ such that for all $n \geq d$ $$c_0(n) a_n + c_1(n) a_{n-1} + \cdots + c_d(n) a_{n-d} = 0$$

These families contain most (but certainly not all) series and sequences that we encounter in combinatorics. They are related as follows.

Theorem 1.3.7 *Let $A(x) = a_0 + a_1 x + a_2 x^2 + \cdots$ be a formal power series. The following implications hold.*

$$\begin{array}{ccccc} A(x) \textit{ is rational} & \Longrightarrow & A(x) \textit{ is algebraic} & \Longrightarrow & A(x) \textit{ is D-finite} \\ \Updownarrow & & & & \Updownarrow \\ \{a_0, a_1, \ldots\} \textit{ is c-recursive} & & \Longrightarrow & & \{a_0, a_1, \ldots\} \textit{ is P-recursive} \end{array}$$

Proof. We already discussed the correspondence between rational series and c-recursive functions, and rational series are trivially algebraic. Let us prove the remaining statements.

(Algebraic $\Rightarrow$ D-finite) Suppose $A(x)$ satisfies an algebraic equation of degree d. Then A is algebraic over the field $\mathbb{C}(x)$, and the field extension $\mathbb{C}(x, A)$ is a vector space over $\mathbb{C}(x)$ having dimension of at most d.

Taking the derivative of the polynomial equation satisfied by A, we get an expression for A' as a rational function of A and x. Taking derivatives repeatedly, we find that all derivatives of A are in $\mathbb{C}(x, A)$. It follows that $1, A, A', A'', \ldots, A^{(d)}$ are linearly dependent over $\mathbb{C}(x)$, and a linear relation between them is a certificate for the D-finiteness of A.

(P-recursive $\Leftrightarrow$ D-finite) If $q_0(x)A(x) + q_1(x)A'(x) + \cdots + q_d(x)A^{(d)}(x) = q(x)$, comparing the coefficients of x^n gives a P-recursion for the a_is. In the other direction, given a P-recursion for the a_is of the form $c_0(n)a_n + \cdots + c_d(n)a_{n-d} = 0$, it is easy to obtain the corresponding differential equation after writing $c_i(x)$ in terms of the basis $\{(x+i)_k : k \in \mathbb{N}\}$ of $\mathbb{C}[x]$, where $(y)_k = y(y-1)\cdots(y-k+1)$. ■

The converses are not true. For instance, $\sqrt{1+x}$ is algebraic but not rational, and e^x and $\log(1-x)$ are D-finite but not algebraic.

Corollary 1.3.8 *The ordinary generating function $\sum_n a_n x^n$ is D-finite if and only if the exponential generating function $\sum_n a_n \frac{x^n}{n!}$ is D-finite.*

Proof. This follows from the observation that $\{a_n : n \in \mathbb{N}\}$ is P-recursive if and only if $\{a_n/n! : n \in \mathbb{N}\}$ is P-recursive. ■

A few examples. Before we discuss general tools, we collect some examples. We will prove all of the following statements later in this section.

The power series for subsets, Fibonacci numbers, and Stirling numbers are rational:

$$\sum_{n\geq 0} 2^n x^n = \frac{1}{1-2x}, \quad \sum_{n\geq 0} F_n x^n = \frac{x}{1-x-x^2}, \quad \sum_{n\geq k} S(n,k)x^n = \frac{x}{1-x}\cdot\frac{x}{1-2x}\cdots\frac{x}{1-kx}.$$

The "diagonal binomial," k-Catalan, and Motzkin series are algebraic but not rational:

$$\sum_{n\geq 0}\binom{2n}{n}x^n = \frac{1}{\sqrt{1-4x}}, \qquad \sum_{n\geq 0}\frac{1}{(k-1)n+1}\binom{kn}{n}x^n,$$

and

$$\sum_{n\geq 0} M_n x^n = \frac{1-x-\sqrt{1-2x-3x^2}}{2x^2}.$$

The following series are D-finite but not algebraic:

$$e^x, \quad \log(1+x), \quad \sin x, \quad \cos x, \quad \arctan x, \quad \sum_{n\geq 0}\binom{2n}{n}^2 x^n, \quad \sum_{n\geq 0}\binom{3n}{n,n,n}x^n$$

The following series are not D-finite:

$$\sqrt{1+\log(1+x^2)}, \quad \sec x, \quad \tan x, \quad \sum_{n\geq 0} p(n)x^n = \prod_{k\geq 0}\frac{1}{1-x^k}.$$

Recognizing algebraic and D-finite series. It is not always obvious whether a given power series is algebraic or D-finite, but there are some tools available. Fortunately, algebraic functions behave well under a few operations, and D-finite functions behave even better. This explains why these families contain most examples arising in combinatorics.

The following table summarizes the properties of formal power series that are preserved under various key operations. For example, the fifth entry on the bottom row says that if $A(x)$ and $B(x)$ are D-finite, then the composition $A(B(x))$ is not necessarily D-finite.

	cA	$A+B$	AB	$1/A$	$A\circ B$	$A\star B$	A'	$\int A$	$A^{\langle -1\rangle}$
rational	Y	Y	Y	Y	Y	Y	Y	N	N
algebraic	Y	Y	Y	Y	Y	N	Y	N	Y
D-finite	Y	Y	Y	N	N	Y	Y	Y	N

Here $A \star B(x) := \sum_{n\geq 0} a_n b_n x^n$ denotes the **Hadamard product** of $A(x)$ and $B(x)$, $\int A(x) := \sum_{n\geq 1} \frac{a_{n-1}}{n} x^n$ is the **formal integral** of $A(x)$, and $A^{\langle -1\rangle}(x)$ is the **compositional inverse** of $A(x)$. In the fourth column we are assuming that $A(0) \neq 0$ so that $1/A(x)$ is well-defined, in the fifth column we are assuming that $B(0) = 0$ so that $A(B(x))$ is well-defined, and in the last column we are assuming that $A(0) = 0$ and $A'(0) \neq 0$, so that $A^{\langle -1\rangle}(x)$ is well-defined.

For proofs of the "Yes" entries, see [183], [192], and [79]. For the "No" entries, we momentarily assume the statements of the previous subsection. Then we have the following counterexamples:

- $\cos x$ is D-finite but $1/\cos x = \sec x$ is not.
- $\sqrt{1+x}$ and $\log(1+x^2)$ are D-finite but their composition $\sqrt{1+\log(1+x^2)}$ is not.
- $A(x) = \sum_{n\geq 0} \binom{2n}{n} x^n$ is algebraic but $A \star A(x) = \sum_{n\geq 0} \binom{2n}{n}^2 x^n$ is not.
- $1/(1+x)$ is rational and algebraic but its integral $\log(1+x)$ is neither.
- $x + x^2$ is rational but its compositional inverse $(-1+\sqrt{1+4x})/2$ is not.
- $\arctan x$ is D-finite but its compositional inverse $\tan x$ is not.

Some of these negative results have weaker positive counterparts:

- If $A(x)$ is algebraic and $B(x)$ is rational, then $A(x) \star B(x)$ is algebraic.
- If $A(x)$ is D-finite and $A(0) \neq 0$, $1/A(x)$ is D-finite if and only if $A'(x)/A(x)$ is algebraic.
- If $A(x)$ is D-finite and $B(x)$ is algebraic with $B(0) = 0$, then $A(B(x))$ is D-finite.

See [192, Proposition 6.1.11], [96], and [192, Theorem 6.4.10] for the respective proofs.

The following result is also useful.

Theorem 1.3.9 [192, Section 6.3] *Consider a multivariate formal power series $F(x_1,\ldots,x_d)$ that is rational in $x_1,\ldots,x_d$ and its* ***diagonal****:*

$$F(x_1,\ldots,x_d) = \sum_{n_1,\ldots,n_d \geq 0} a_{n_1,\ldots,n_d} x_1^{n_1} \cdots x_d^{n_d}, \qquad \operatorname{diag} F(x) = \sum_{n\geq 0} a_{n,\ldots,n} x^n.$$

1. *If $d = 2$, then* diag $F(x)$ *is algebraic.*
2. *If $d > 2$, then* diag $F(x)$ *is D-finite but not necessarily algebraic.*

Now we are ready to prove our positive claims about the series at the beginning of this section. The first three expressions are visibly rational, and the diagonal binomial and Motzkin series are visibly algebraic. We proved that the k-Catalan series is algebraic in Section 1.3.2.2. The functions e^x, $\log(1+x)$, $\sin x$, $\cos x$, $\arctan x$ satisfy the differential equations $y' = y$, $(1+x)y' = 1$, $y'' = -y$, $y'' = -y$, $(1+x^2)y' = 1$, respectively. The series $\sum_{n\geq 0}\binom{2n}{n}^2 x^n$ is the Hadamard product of $(1-4x)^{-1/2}$ with itself, and hence D-finite. Finally, $\sum_{n\geq 0}\binom{3n}{n,n,n}x^n$ is the diagonal of the rational function $\frac{1}{1-x-y-z} = \sum_{a,b,c\geq 0}\binom{a+b+c}{a,b,c}x^a y^b z^c$, and hence D-finite.

Proving the negative claims requires more effort and, often, a bit of analytic machinery. We briefly outline some key results.

Recognizing series that are not algebraic. There are a few methods available to prove that a series is **not** algebraic. The simplest algebraic and analytic criteria are the following.

Theorem 1.3.10 (Eisenstein's theorem [155]) *If a series $A(x) = \sum_{n\geq 0} a_n x^n$ with rational coefficients is algebraic, then there exists a positive integer m such that $a_n m^n$ is an integer for all $n > 0$.*

This shows that $e^x, \log(1+x), \sin x, \cos x$, and $\arctan x$ are not algebraic.

Theorem 1.3.11 [111] *If the coefficients of an algebraic power series $A(x) = \sum_{n\geq 0} a_n x^n$ satisfy $a_n \sim c n^r \alpha^n$ for nonzero $c, \alpha \in \mathbb{C}$ and $r < 0$, then r cannot be a negative integer.*

Stirling's approximation $n! \sim \sqrt{2\pi n}\,(n/e)^n$ gives $\binom{2n}{n}^2 \sim c\cdot 16^n/n$ and $\binom{3n}{n,n,n} \sim c\cdot 27^n/n$, so the corresponding series are not algebraic.

Another useful analytic criterion is that an algebraic series $A(x)$ must have a Newton-Puiseux expansion at any of its singularities. See [79, Theorem VII.7] and [78] for details.

Recognizing series that are not D-finite. The most effective methods to show that a function is **not** D-finite are analytic.

Theorem 1.3.12 [97, Theorem 9.1] *Suppose that $A(x)$ is analytic at $x = 0$, and it is D-finite, satisfying the equation $q_0(x)A(x) + q_1(x)A'(x) + \cdots + q_d(x)A^{(d)}(x) = q(x)$ with $q_d(x) \neq 0$. Then $A(x)$ can be extended to an analytic function in any simply connected region of the complex plane not containing the (finitely many) zeroes of $q_d(x)$.*

Since $\sec x$ and $\tan x$ have a pole at every odd multiple of π, they are not D-finite. Similarly, $\sum_n p(n)x^n = \prod_{k=1}^{\infty}\frac{1}{1-x^k}$ is not D-finite because it has the circle $|x| = 1$ as a natural boundary of analyticity.

There are other powerful analytic criteria to prove a series is not D-finite. See [78, Theorem VII.7] for details and further examples.

Sometimes it is possible to give *ad hoc* proofs that series are D-finite. For instance, consider $y = \sqrt{1+\log(1+x^2)}$. By induction, for any $k \in \mathbb{N}$ there exist polynomials $r_1(x), \ldots, r_k(x)$ such that $y^{(k)} = r_1/y + r_2/y^3 + \cdots + r_k/y^{2k-1}$. An equation of the form $\sum_{i=0}^{d} q_i(x) y^{(i)} = q(x)$ would then give rise to a polynomial equation satisfied by y. This would also make $y^2 - 1 = \log(1+x^2)$ algebraic; but this contradicts Theorem 1.3.10.

1.4 Linear algebra methods

There are several important theorems in enumerative combinatorics that express a combinatorial quantity in terms of a determinant. Of course, evaluating a determinant is not always straightforward, but there is a wide array of tools at our disposal.

The goal of Section 1.4.1 is to reduce many combinatorial problems to "just computing a determinant"; examples include walks in a graph, spanning trees, Eulerian cycles, matchings, and routings. In particular, we discuss the transfer matrix method, which allows us to encode many combinatorial objects as walks in graphs, so that these linear algebraic tools apply. These problems lead us to many beautiful, mysterious, and highly non-trivial determinantal evaluations. We will postpone the proofs of the evaluations until Section 1.4.2, which is an exposition of some of the main techniques in the subtle science of computing combinatorial determinants.

1.4.1 Determinants in combinatorics

1.4.1.1 Preliminaries: Graph matrices

An **undirected graph**, or simply a **graph** $G = (V,E)$ consists of a set V of vertices and a set E of edges $\{u,v\}$ where $u,v \in V$ and $u \neq v$. In an undirected graph, we write uv for the edge $\{u,v\}$. The **degree** of a vertex is the number of edges incident to it. A **walk** is a set of edges of the form $v_1v_2, v_2v_3, \ldots, v_{k-1}v_k$. This walk is **closed** if $v_k = v_1$.

A **directed graph** or **digraph** $G = (V,E)$ consists of a set V of vertices and a set E of oriented edges (u,v) where $u,v \in V$ and $u \neq v$. In an undirected graph, we write uv for the directed edge (u,v). The **outdegree** (respectively, **indegree**) of a vertex is the number of edges coming out of it (respectively, coming into it). A **walk** is a set of directed edges of the form $v_1v_2, v_2v_3, \ldots, v_{k-1}v_k$. This walk is **closed** if $v_k = v_1$.

We will see in this section that many graph theory problems can be solved using tools from linear algebra. There are several matrices associated to graphs that play a crucial role; we review them here.

Directed graphs. Let $G = (V, E)$ be a directed graph.

- The **adjacency matrix** $A = A(G)$ is the $V \times V$ matrix whose entries are

$$a_{uv} = \text{number of edges from } u \text{ to } v.$$

- The **incidence matrix** $M = M(G)$ is the $V \times E$ matrix with

$$m_{ve} = \begin{cases} 1 & \text{if } v \text{ is the final vertex of edge } e, \\ -1 & \text{if } v \text{ is the initial vertex of edge } e, \\ 0 & \text{otherwise.} \end{cases}$$

- The **directed Laplacian matrix** $\overrightarrow{L} = \overrightarrow{L}(G)$ is the $V \times V$ matrix whose entries are

$$\overrightarrow{l}_{uv} = \begin{cases} -(\text{number of edges from } u \text{ to } v) & \text{if } u \neq v, \\ \text{outdeg}(u) & \text{if } u = v. \end{cases}$$

Undirected graphs. Let $G = (V, E)$ be an undirected graph.

- The (undirected) **adjacency matrix** $A = A(G)$ is the $V \times V$ matrix whose entries are

$$a_{uv} = \text{number of edges connecting } u \text{ and } v.$$

  This is the directed adjacency matrix of the directed graph on V containing edges $u \to v$ and $v \to u$ for every edge uv of G.

- The (undirected) **Laplacian matrix** $L = L(G)$ is the $V \times V$ matrix with entries

$$l_{uv} = \begin{cases} -(\text{number of edges connecting } u \text{ and } v) & \text{if } u \neq v \\ \deg u & \text{if } u = v \end{cases}$$

  If M is the incidence matrix of any orientation of the edges of G, then $L = MM^T$.

1.4.1.2 Walks: the transfer matrix method

Counting walks in a graph is a fundamental problem, which (often in disguise) includes many important enumerative problems. The transfer matrix method addresses this problem by expressing the number of walks in a graph G in terms of its adjacency matrix $A(G)$, and then uses linear algebra to count those walks.

Directed or undirected graphs. The transfer matrix method is based on the following simple, powerful observation, which applies to directed and undirected graphs:

Theorem 1.4.1 *Let $G = (V,E)$ be a graph and let $A = A(G)$ be the $V \times V$ adjacency matrix of G, where a_{uv} is the number of edges from u to v. Then*

$$(A^n)_{uv} = \text{ number of walks of length } n \text{ in } G \text{ from } u \text{ to } v.$$

Proof. Observe that

$$(A^n)_{uv} = \sum_{w_1,\ldots,w_{n-1}\in V} a_{uw_1}a_{w_1w_2}\cdots a_{w_{n-1}v}$$

and there are $a_{uw_1}a_{w_1w_2}\cdots a_{w_{n-1}v}$ walks of length n from u to v visiting vertices $u, w_1, \ldots, w_{n-1}, v$ in that order. ■

Corollary 1.4.2 *The generating function $\sum_{n\geq 0}(A^n)_{uv}x^n$ for walks of length n from u to v in G is a rational function.*

Proof. Using Cramer's formula, we have

$$\sum_{n\geq 0}(A^n)_{uv}x^n = ((I-xA)^{-1})_{uv} = (-1)^{u+v}\frac{\det(I-xA : v,u)}{\det(I-xA)}$$

where $(M : v,u)$ is the cofactor of M obtained by removing row v and column u. ■

Corollary 1.4.3 *If $C_G(n)$ is the number of closed walks of length n in G, then*

$$C_G(n) = \lambda_1^n + \cdots + \lambda_k^n, \qquad \sum_{n\geq 1} C_G(n)x^n = \frac{-xQ'(x)}{Q(x)}$$

where $\lambda_1,\ldots,\lambda_k$ are the eigenvalues of adjacency matrix A and $Q(x) = \det(I-xA)$.

Proof. Theorem 1.4.1 implies that $C_G(n) = \operatorname{tr}(A^n) = \lambda_1^n + \cdots + \lambda_k^n$. The second equation then follows from $Q(x) = (1-\lambda_1 x)\cdots(1-\lambda_k x)$. ■

In view of Theorem 1.4.1, we want to be able to compute powers of the adjacency matrix A. As we learn in linear algebra, this is very easy to do if we are able to diagonalize A. This is not always possible, but we can do it when A is undirected.

Undirected graphs. When our graph G is undirected, the adjacency matrix $A(G)$ is symmetric, and hence diagonalizable.

Theorem 1.4.4 *Let $G = (V,E)$ be an* ***undirected*** *graph and let $\lambda_1,\ldots,\lambda_k$ be the eigenvalues of the adjacency matrix $A = A(G)$. Then for any vertices u and v there exist constants $c_1,\ldots,c_k$ such that*

$$\text{number of walks of length } n \text{ from } u \text{ to } v = c_1\lambda_1^n + \cdots + c_k\lambda_k^n.$$

Proof. The key fact is that a real symmetric $k \times k$ matrix A has k real orthonormal eigenvectors $q_1, \ldots, q_k$ with real eigenvalues $\lambda_1, \ldots, \lambda_k$. Equivalently, the $k \times k$ matrix Q with columns $q_1, \ldots, q_k$ is orthogonal (so $Q^T = Q^{-1}$) and diagonalizes A:

$$Q^{-1}AQ = D = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$$

where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$ is the diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_k$. The result then follows from $A^n = QD^nQ^{-1} = Q\,\mathrm{diag}(\lambda_1^n, \ldots, \lambda_k^n)\,Q^T$, with $c_t = q_{it}q_{jt}$. ■

Applications. Many families of combinatorial objects can be enumerated by first recasting the objects as walks in a "transfer graph" and then applying the transfer matrix method. We illustrate this technique with a few examples.

1. (Colored necklaces) Let $f(n,k)$ be the number of ways of coloring the beads of a necklace of length n with k colors so that no two adjacent beads have the same color. (Different rotations and reflections of a coloring are considered different.) There are several ways to compute this number, but a very efficient one is to notice that such a coloring is a graph walk in disguise. If we label the beads $1, \ldots, n$ in clockwise order and let a_i be the color of the ith bead, then the coloring corresponds to the closed walk $a_1, a_2, \ldots, a_n, a_1$ in the complete graph K_n. The adjacency graph of K_n is $A = J - I$ where J is the matrix all of whose entries equal 1, and I is the identity. Since J has rank 1, it has $n-1$ eigenvalues equal to 0. Since the trace is n, the last eigenvalue is n. It follows that the eigenvalues of $A = J - I$ are $-1, -1, \ldots, -1, n-1$. Then Corollary 1.4.3 tells us that

   $$f(n,k) = (n-1)^k + (n-1)(-1)^k.$$

   It is possible to give a bijective proof of this formula, but this algebraic proof is much simpler.

2. (Words with forbidden subwords, 1) Let h_n be the number of words of length n in the alphabet $\{a, b\}$ that do not contain aa as a consecutive subword. This is the same as a walk of length $n-1$ in the transfer graph with vertices a and b and edges $a \to b$, $b \to a$ and $b \to b$. The absence of the edge $a \to a$ guarantees that these walks produce only the valid words we wish to count. The adjacency matrix and its powers are

   $$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \qquad A^n = \begin{pmatrix} F_{n-1} & F_n \\ F_n & F_{n+1} \end{pmatrix},$$

   where the Fibonacci numbers $F_0, F_1, \ldots$ are defined recursively by $F_0 = 0, F_1 = 1$, and $F_k = F_{k-1} + F_{k-2}$ for $k \geq 2$.

   Since h_n is the sum of the entries of A^{n-1}, we get that $h_n = F_{n+2}$, and $g_n \sim c \cdot \alpha^n$ where $\alpha = \frac{1}{2}(1 + \sqrt{5}) \approx 1.6179\ldots$ is the golden ratio. Of course there are easier proofs of this fact, but this approach works for any problem of enumerating words in a given alphabet with given forbidden consecutive subwords. Let us study a slightly more intricate example, which should make it clear how to proceed in general.

3. (Words with forbidden subwords, 2) Let g_n be the number of cyclic words of length n in the alphabet $\{a,b\}$ that do not contain *aa* or *abba* as a consecutive subword. We wish to model these words as walks in a directed graph. At first this may seem impossible because, as we construct the word sequentially, the validity of a new letter depends on more than just the previous letter. However, a simple trick resolves this difficulty: We can introduce more memory into the vertices of the transfer graph. In this case, since the validity of a new letter depends on the previous three letters, we let the vertices of the transfer graph be aba, abb, bab, bba, bbb (the allowable "windows" of length 3) and put an edge $wxy \to xyz$ in the graph if the window wxy is allowed to precede the window xyz; that is, if $wxyz$ is an allowed subword. The result is the graph of Figure 1.14, whose adjacency matrix A satisfies

$$\det(I-xA) = \det\begin{pmatrix} 1 & 0 & -x & 0 & 0 \\ 0 & 1 & 0 & 0 & -x \\ -x & -x & 1 & 0 & 0 \\ 0 & 0 & -x & 1 & 0 \\ 0 & 0 & 0 & -x & 1-x \end{pmatrix} = -x^4+x^3-x^2-x+1.$$

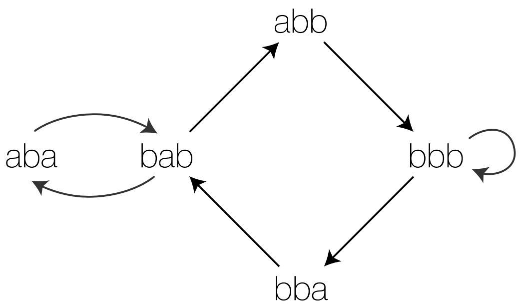


Figure 1.14
The transfer graph for words on the alphabet $\{a,b\}$ avoiding *aa* and *abba* as consecutive subwords.

The valid cyclic words of length n correspond to the closed walks of length n in the transfer graph, so Corollary 1.4.3 tells us that the generating function for g_n is

$$\begin{aligned}\sum_{n\geq 0} g_n x^n &= \frac{x+2x^2-3x^3+4x^4}{1-x-x^2+x^3-x^4}\\ &= x+3x^2+x^3+7x^4+6x^5+15x^6+15x^7+31x^8+37x^9+\cdots.\end{aligned}$$

Theorem 1.3.5.2 then tells us that $g_n \approx c\cdot\alpha^n$ where $\alpha\approx 1.5129$ is the inverse of the smallest positive root of $1-x-x^2+x^3-x^4=0$. The values of g_1,g_2,g_3 may surprise us. Note that the generating function does something counterintuitive: it does not count the words *a* (because *aa* is forbidden), *aba* (because *aa* is forbidden), or *abb* (because *abba* is forbidden).

This example serves as a word of caution: When we use the transfer matrix method to enumerate "cyclic" objects using Corollary 1.4.3, the initial values of the generating function may not be the ones we expect. In a particular problem of interest, it will be straightforward to adjust those values accordingly.

To illustrate the wide applicability of this method, we conclude this section with a problem where the transfer graph is less apparent.

4. (Monomer-dimer problem) An important open problem in statistical mechanics is the **monomer-dimer problem** of computing the number of tilings $T(m,n)$ of an $m\times n$ rectangle into dominoes (2×1 rectangles) and unit squares. Equivalently, $T(m,n)$ is the number of partial matchings of an $m\times n$ grid, where each node is matched to at most one of its neighbors.

   There is experimental evidence, but no proof, that $T(n,n)\sim c\cdot\alpha^{n^2}$ where $\alpha\approx$ 1.9402... is a constant for which no exact expression is known. The transfer-matrix method is able to solve this problem for any fixed value of m, proving that the generating function $\sum_{n\geq 0} T(m,n)x^n$ is rational. We carry this out for $m=3$.

   Let $t(n)$ be the number of tilings of a $3\times n$ rectangle into dominoes and unit squares. As with words, we can build our tilings sequentially from left to right by covering the first column, then the second column, and so on. The tiles that we can place on a new column depend only on the tiles from the previous column that are sticking out, and this can be modeled by a transfer graph.

   More specifically, let T be a tiling of a $3\times n$ rectangle. We define $n+1$ triples $v_0,\ldots,v_n$ which record how T interacts with the $n+1$ vertical grid lines of the rectangle. The ith grid line consists of three unit segments, and each coordinate of v_i is 0 or 1 depending on whether these three segments are edges of the tiling or not. For example, Figure 1.15 corresponds to the triples $111,110,011,101,010,111$.

   The choice of v_i is restricted only by v_{i-1}. The only restriction is that v_{i-1} and v_i cannot both have a 0 in the same position, because this would force us to put

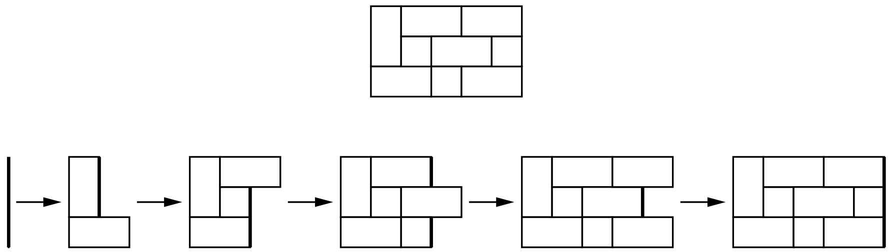


Figure 1.15
A tiling of a 3×5 rectangle into dominoes and unit squares.

two overlapping horizontal dominoes in T. These compatibility conditions are recorded in the transfer graph of Figure 1.16. When $v_{i-1} = v_i = 111$, there are three ways of covering column i. If v_{i-1} and v_i share two 1s in consecutive positions, there are two ways. In all other cases, there is a unique way. It follows that the tilings of a $3 \times n$ rectangle are in bijection with the walks of length n from 111 to 111 in the transfer graph.

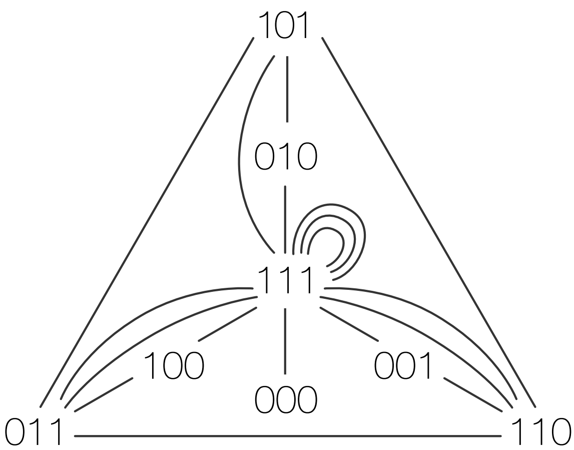


Figure 1.16
The transfer graph for tilings of $3 \times n$ rectangles into dominoes and unit squares.

Since the adjacency matrix is

$$A = \begin{array}{c} \\ 000 \\ 001 \\ 110 \\ 010 \\ 101 \\ 001 \\ 110 \\ 111 \end{array} \begin{array}{c} \begin{array}{cccccccc} 000 & 001 & 110 & 010 & 101 & 001 & 110 & 111 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 2 \\ 1 & 1 & 2 & 1 & 1 & 1 & 2 & 3 \end{pmatrix} \end{array}$$

Theorem 1.4.1 tells us that

$$\begin{aligned} \sum_{n\geq 0} t(n)x^n &= \frac{\det(I - xA : 111, 111)}{\det(I - xA)} \\ &= \frac{(1+x-x^2)(1-2x-x^2)}{(1+x)(1-5x-9x^2+9x^3+x^4-x^5)} \\ &= 1+3x+22x^2+131x^3+823x^4+5096x^5+31687x^6+\cdots \end{aligned}$$

By Theorem 1.3.5.2, $t_n \sim c \cdot \alpha^n$ where $\alpha \approx 6.21207\ldots$ is the inverse of the smallest positive root of the denominator $(1+x)(1-5x-9x^2+9x^3+x^4-x^5)$.

1.4.1.3 Spanning trees: the matrix-tree theorem

In this section we discuss two results: Kirkhoff's determinantal formula for the number of spanning trees of a graph, and Tutte's generalization to oriented spanning trees of directed graphs.

Undirected matrix-tree theorem. Let $G = (V, E)$ be a connected graph with no loops. A **spanning tree** T of G is a collection of edges such that for any two vertices u and v, T contains a unique path between u and v. If G has n vertices, then

- T contains no cycles,
- T spans G; that is, there is a path from u to v in T for any vertices $u \neq v$, and
- T has $n-1$ edges.

Furthermore, any two of these properties imply that T is a spanning tree. Our goal in this section is to compute the number $c(G)$ of spanning trees of G.

Orient the edges of G arbitrarily. Recall that the incidence matrix M of G is the $V \times E$ matrix whose eth column is $\mathbf{e}_v - \mathbf{e}_u$ if $e = u \to v$, where $\mathbf{e}_i$ is the ith basis vector. The Laplacian $L = MM^T$ has entries

$$l_{uv} = \begin{cases} -(\text{number of edges connecting } u \text{ and } v) & \text{if } u \neq v \\ \deg u & \text{if } u = v \end{cases}$$

Note that $L(G)$ is singular because all its row sums are 0. A **principal cofactor** $L_v(G)$ is obtained from $L(G)$ by removing the vth row and vth column for some vertex v.

Theorem 1.4.5 (Kirkhoff's Matrix-Tree Theorem) *The number $c(G)$ of spanning trees of a connected graph G is*

$$c(G) = \det L_v(G) = \frac{1}{n}\lambda_1 \cdots \lambda_{n-1},$$

where $L_v(G)$ is any principal cofactor of the Laplacian $L(G)$, and $\lambda_1, \ldots, \lambda_{n-1}, \lambda_n = 0$ are the eigenvalues of $L(G)$.

Proof. We use the Binet-Cauchy formula, which states that if A and B are $m \times n$ and $n \times m$ matrices, respectively, with $m < n$, then

$$\det AB = \sum_{S \subseteq [n]\,:\,|S|=m} \det A[S] \det B[S]$$

where $A[S]$ (respectively, $B[S]$) is the $n \times n$ matrix obtained by considering only the columns of A (respectively, the rows of B) indexed by S.

We also use the following observation: If M_v is the "reduced" adjacency matrix M with the vth row removed, and S is a set of $n-1$ edges of E, then

$$\det M_v[S] = \begin{cases} \pm 1 & \text{if } S \text{ is a spanning tree,} \\ 0 & \text{otherwise.} \end{cases}$$

This observation is easily proved: If S is not a spanning tree, then it contains a cycle C, which gives a linear dependence among the columns indexed by the edges of C. Otherwise, if S is a spanning tree, think of v as its root, and "prune" it by repeatedly removing a leaf $v_i \neq v$ and its only incident edge e_i for $1 \leq i \leq n-1$. Then if we list the rows and columns of $M[S]$ in the orders $v_1, \ldots, v_{n-1}$ and $e_1, \ldots, e_{n-1}$, respectively, the matrix will be lower triangular with 1s and -1s in the diagonal.

Combining these two equations, we obtain the first statement:

$$\det L_v(G) = \sum_{S \subseteq [n]\,:\,|S|=m} \det M[S] \det M^T[S] = \sum_{S \subseteq [n]\,:\,|S|=m} \det M[S]^2 = c(G).$$

To prove the second one, observe that the coefficient of $-x^1$ in the characteristic polynomial $\det(L - xI) = (\lambda_1 - x) \cdots (\lambda_{n-1} - x)(0 - x)$ is the sum of the n principal cofactors, which are all equal to $c(G)$. ■

The matrix-tree theorem is a very powerful tool for computing the number of spanning trees of a graph. Let us state a few examples.

The **complete graph** K_n has n vertices and an edge joining each pair of vertices. The **complete bipartite graph** $K_{m,n}$ has m "top" vertices and n "bottom" vertices, and mn edges joining each top vertex to each bottom vertex. The **hyperoctahedral graph** $\Diamond_n$ has vertices $\{1, 1', 2, 2', \ldots, n, n'\}$ and its only missing edges are ii' for $1 \leq$

$i \leq n$. The n**-cube graph** C_n has vertices $(\varepsilon_1, \ldots, \varepsilon_n)$ where $\varepsilon_i \in \{0,1\}$, and an edge connecting any two vertices that differ in exactly one coordinate. The n**-dimensional grid of size** m, denoted mC_n, has vertices $(\varepsilon_1, \ldots, \varepsilon_n)$ where $\varepsilon_i \in \{1, \ldots, m\}$, and an edge connecting any two vertices that differ in exactly one coordinate i, where they differ by 1.

Theorem 1.4.6 *The number of spanning trees of some interesting graphs are as follows.*

1. *(Complete graph)* $c(K_n) = n^{n-2}$
2. *(Complete bipartite graph)* $c(K_{m,n}) = m^{n-1}n^{m-1}$
3. *(Hyperoctahedral graph)* $c(\Diamond_n) = 2^{2n-2}(n-1)^n n^{n-2}$
4. *(n-cube)* $c(C_n) = 2^{2^n-n-1}\prod_{k=1}^n k^{\binom{n}{k}}$
5. *(n-dimensional grid of size m)* $c(mC_n) = m^{m^n-n-1}\prod_{k=1}^n k^{\binom{n}{k}(m-1)^k}$

We will see proofs of the first and third example in Section 1.4.2. For the others, and many additional examples, see [55].

Directed matrix-tree theorem. Now let $G = (V,E)$ be a **directed** graph containing no loops. An **oriented spanning tree rooted at** v is a collection of edges T such that for any vertex u there is a unique path from u to v. The underlying unoriented graph $\underline{T}$ is a spanning tree of the unoriented graph $\underline{G}$. Let $c(G,v)$ be the number of spanning trees rooted at G.

Recall that the **directed Laplacian matrix** $\overrightarrow{L}$ has entries

$$\overrightarrow{l}_{uv} = \begin{cases} -(\text{number of edges from } u \text{ to } v) & \text{if } u \neq v \\ \text{outdeg } u & \text{if } u = v \end{cases}$$

Now the matrix $\overrightarrow{L}(G)$ is not necessarily symmetric, but it is still singular.

Theorem 1.4.7 (Tutte's Directed Matrix-Tree Theorem) *Let G be a directed graph and v be a vertex. The number $c(G,v)$ of oriented spanning trees rooted at v is*

$$c(G,v) = \det \overrightarrow{L}_v(G)$$

*where $\overrightarrow{L}_v(G)$ is obtained from $L(G)$ by removing the vth row and column. Furthermore, if G is **balanced**, so indeg v = outdeg v for all vertices v, then*

$$c(G,v) = \frac{1}{n}\lambda_1 \cdots \lambda_{n-1}$$

where $\lambda_1, \ldots, \lambda_{n-1}, \lambda_n = 0$ are the eigenvalues of $L(G)$.

Proof. Proceed by induction. Consider a vertex $w \neq v$ and an edge e starting at w. Let $G' = G - e$ be obtained from G by removing e. If e is the only edge starting at w, then every spanning tree must use it, and we have

$$c(G,v) = c(G',v) = \det \overrightarrow{L}_v(G') = \det \overrightarrow{L}_v(G).$$

Otherwise, let G'' be obtained from G by removing all edges starting at w other than e. There are $c(G',v)$ oriented spanning trees rooted at v that do not contain e, and $c(G'',v)$ that do contain e, so we have

$$c(G,v) = c(G',v) + c(G'',v) = \det \overrightarrow{L}_v(G') + \det \overrightarrow{L}_v(G'') = \det \overrightarrow{L}_v(G)$$

where the last equality holds since determinants are multilinear.

We postpone the proof of the second statement to the next section, where it will be an immediate consequence of Theorem 1.4.9. ■

1.4.1.4 Eulerian cycles: the BEST theorem

One of the earliest combinatorial questions is the problem of the Seven Bridges of Königsberg. In the early 1700s, the Prussian city of Königsberg was separated by the Pregel river into four regions, connected to each other by seven bridges. In the map of Figure 1.17 we have labeled the regions N, S, E, and I; there are two bridges between N and I, two between S and I, and three bridges connecting E to each of N, I, and S. The problem was to find a walk through the city that crossed each bridge exactly once. Euler proved in 1735 that it was impossible to find such a walk; this is considered to be the first paper in graph theory.

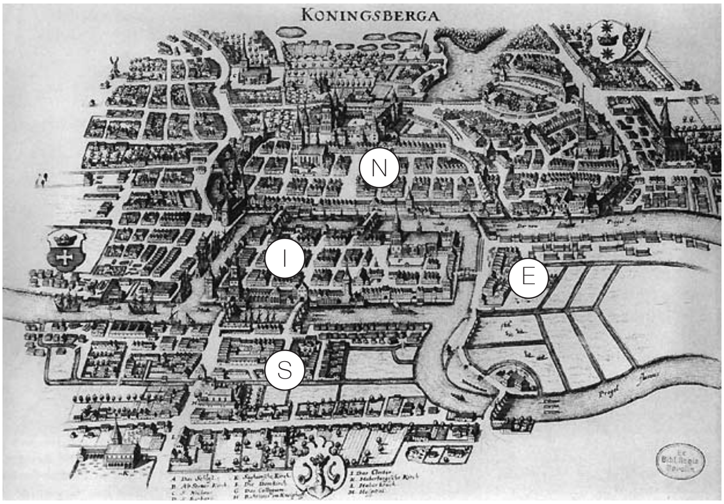


Figure 1.17
The seven bridges of Königsberg. Public domain map by Merian-Erben, 1652.

Euler's argument is simple, and relies on the fact that every region of Königsberg is adjacent to an odd number of bridges. Suppose there existed such a walk, starting at region A and ending at region B. Now consider a region C other than A and B. Then our path would enter and leave C the same number of times; but then it would not use all the bridges adjacent to C, because there is an odd number of such bridges.

In modern terminology, each region of the city is represented by a vertex, and each bridge is represented by an edge connecting two vertices. We will be more interested in the directed case, where every edge has an assigned direction. An **Eulerian path** is a path in the graph that visits every edge exactly once. If the path starts and ends at the same vertex, then it is called an **Eulerian cycle**. We say G is an **Eulerian graph** if it has an Eulerian cycle.

Theorem 1.4.8 *A directed graph is Eulerian if and only if it is connected and every vertex v satisfies $indeg(v) = outdeg(v)$.*

Proof. If a graph has an Eulerian cycle C, then C enters and leaves each vertex v the same number of times. Therefore $\mathrm{indeg}(v) = \mathrm{outdeg}(v)$.

To prove the converse, let us start by arbitrarily "walking around G until we get stuck." More specifically, we start at any vertex v_0, and at each step, we exit the current vertex by walking along any outgoing edge we have not used yet. If there is no available outgoing edge, we stop.

Whenever we enter a vertex $v \neq v_0$, we will also be able to exit it since $\mathrm{indeg}(v) = \mathrm{outdeg}(v)$; so the walk can only get stuck at v_0. Hence the resulting walk C is a cycle. If C uses all edges of the graph, we are done. If not, then since G is connected we can find a vertex v' of C with an unused outgoing edge, and we use this edge to start walking around the graph $G - C$ until we get stuck, necessarily at v'. The result will be a cycle C'. Starting at v' we can traverse C and then C', thus obtaining a cycle $C \cup C'$ that is longer than C. Repeating this procedure, we will eventually construct an Eulerian cycle. ■

There is a remarkable formula for the number of Eulerian cycles, due to de **B**ruijn, van Ardenne-**E**hrenfest, **S**mith, and **T**utte.

Theorem 1.4.9 (BEST Theorem) *If G is an Eulerian directed graph, then the number of Eulerian cycles of G is*

$$c(G,v) \cdot \prod_{w \in V} (outdeg(w) - 1)!$$

for any vertex v, where $c(G,v)$ is the number of oriented spanning trees rooted at v.

Proof. We fix an edge e starting at v, and let each Eulerian cycle start at e. For each vertex w let E_w be the set of outgoing edges from w.

Consider an Eulerian cycle C. For each vertex $w \neq v$, let e_w be the last outgoing edge from w that C visits, and let π_w (respectively, π_v) be the ordered set $E_w - e_w$ (respectively, $E_v - e$) of the other outgoing edges from w (respectively, v), listed in the order that C traverses them. It is easy to see that $T = \{e_w : w \neq v\}$ is an oriented spanning tree rooted at v.

Conversely, an oriented tree T and permutations $\{\pi_w : w \in V\}$ serve as directions to tour G. We start with edge e. Each time we arrive at vertex w, we exit it by using the first unused edge according to π_w. If we have used all the edges $E_w - e_w$ of π_w, then we use $e_w \in T$. It is not hard to check that this is a bijection. This completes the proof. ■

Corollary 1.4.10 *In an Eulerian directed graph, the number of oriented spanning trees rooted at v is the same for all vertices v; it equals*

$$c(G,v) = \frac{1}{n}\lambda_1 \cdots \lambda_{n-1}$$

where $\lambda_1, \ldots, \lambda_{n-1}, \lambda_n = 0$ *are the eigenvalues of* $\overrightarrow{L}(G)$.

Proof. The BEST theorem implies that $c(G,v)$ is independent of v, and then the argument in the proof of Theorem 1.4.5 applies to give the desired formula. ■

The BEST theorem can be used beautifully to enumerate a very classical, and highly nontrivial, family of objects. A k-ary **de Bruijn sequence** of order n is a cyclic word W of length k^n in the alphabet $\{1, \ldots, k\}$ such that the k^n consecutive subwords of W of length n are the k^n distinct words of length n. For example, the 2-ary deBruijn sequences of order 3 are 11121222 and 22212111; these "memory wheels" were described in Sanskrit poetry several centuries ago [112]. Their existence and enumeration was proved by Flye Saint-Marie in 1894 for $k = 2$ and by van Aardenne-Ehrenfest and de Bruijn in 1951 in general.

Theorem 1.4.11 [80, 58] *The number of k-ary de Bruijn sequences of order n is* $(k!)^{k^{n-1}}/k^n$.

Proof. Consider the **de Bruijn graph** whose vertices are the k^{n-1} sequences of length $n-1$ in the alphabet $\{1, \ldots, k\}$, and where there is an edge from $a_1a_2\ldots a_{n-1}$ to the word $a_2a_3\ldots a_n$ for all $a_1, \ldots, a_n$. It is natural to label this edge $a_1a_2\ldots a_n$. It then becomes apparent that k-ary de Bruijn sequences are in bijection with the Eulerian cycles of the de Bruijn graph. Since $\text{indeg}(v) = \text{outdeg}(v) = k$ for all vertices v, this graph is indeed Eulerian, and we proceed to count its Eulerian cycles. Notice that for any vertices u and v there is a unique path of length n from u to v. Therefore the $k^{n-1} \times k^{n-1}$ adjacency matrix A satisfies $A^n = J$, where J is the matrix whose entries are all equal to 1. We already saw that the eigenvalues of J are $0, \ldots, 0, k^n$. Since the trace of A is k, the eigenvalues of A must be $0, \ldots, 0, k$. Therefore the Laplacian $L = kI - A$ has eigenvalues $k, \ldots, k, 0$. It follows from Corollary 1.4.10 that the de Bruijn graph has $c(G,v) = \frac{1}{k^{n-1}}k^{(k^{n-1}-1)} = k^{k^{n-1}-n}$ oriented spanning trees rooted at any vertex v, and

$$c(G,v) \cdot \prod_{w \in V} (\text{outdeg}(w) - 1)! = k^{k^{n-1}-n} \cdot (k-1)!^{k^{n-1}}$$

Eulerian cycles, as desired. ■

1.4.1.5 Perfect matchings: the Pfaffian method

A **perfect matching** of a graph $G = (V,E)$ is a set M of edges such that every vertex of G is on exactly one edge from M. We are interested in computing the number $m(G)$ of perfect matchings of a graph G. We cannot expect to be able to do this in general; in fact, even for bipartite graphs G, the problem of computing $m(G)$ is #P-complete. However, for many graphs of interest, including all planar graphs, there is a beautiful technique that produces a determinantal formula for $m(G)$.

Determinants and Pfaffians. Let A be a **skew-symmetric** matrix of size $2m \times 2m$, so $A^T = -A$. The Pfaffian is a polynomial encoding the matchings of the complete graph K_{2m}. A perfect matching M of the complete graph K_{2m} is a partition M of $[2m]$ into disjoint pairs $\{i_1, j_1\}, \ldots, \{i_m, j_m\}$, where $i_k < j_k$ for $1 \leq k \leq m$. Draw the points $1, \ldots, 2m$ in order on a line and connect each i_k to j_k by a semicircle above the line. Let $\mathrm{cr}(M)$ be the number of crossings in this drawing, and let $\mathrm{sign}(M) = (-1)^{\mathrm{cr}(M)}$. Let $a_M = a_{i_1 j_1} \cdots a_{i_m j_m}$. The **Pfaffian** of A is

$$\mathrm{Pf}(A) = \sum_M \mathrm{sign}(M) a_M$$

summing over all perfect matchings M of the complete graph K_m.

Theorem 1.4.12 *If A is a skew-symmetric matrix, so $A^T = -A$, then*

$$\det(A) = \mathrm{Pf}(A)^2.$$

Sketch of Proof. The first step is to show that the skew symmetry of A causes many cancellations in the determinant, and

$$\det A = \sum_{\pi \in ECS_n} \mathrm{sign}(\pi) a_\pi$$

where $ECS_n \subset S_n$ is the set of permutations of $[n]$ having only cycles of even length. Then, to prove that this equals $(\sum_M \mathrm{sign}(M) a_M)^2$, we need a bijection between ordered pairs (M_1, M_2) of matchings and permutations π in ECS_n such that $a_{M_1} a_{M_2} = a_\pi$ and $\mathrm{sign}(M_1)\mathrm{sign}(M_2) = \mathrm{sign}(\pi)$. We now describe such a bijection.

Draw the matchings M_1 and M_2 above and below the points $1, \ldots, n$ on a line, respectively. Let π be the permutation given by the cycles of the resulting graph, where each cycle is oriented following the direction of M_1 at its smallest element. This is illustrated in Figure 1.18. It is clear that $a_{M_1} a_{M_2} = a_\pi$, while some care is required to show that $\mathrm{sign}(M_1)\mathrm{sign}(M_2) = \mathrm{sign}(\pi)$. For details, see [2]. ■

Counting perfect matchings via Pfaffians. Suppose we wish to compute the number $m(G)$ of perfect matchings of a graph $G = (V,E)$ with no loops. After choosing an orientation of the edges, we define the $V \times V$ **signed adjacency matrix** $S(G)$ whose entries are

$$s_{ij} = \begin{cases} 1 & \text{if } i \to j \text{ is an edge of } G \\ -1 & \text{if } j \to i \text{ is an edge of } G \\ 0 & \text{otherwise.} \end{cases}$$

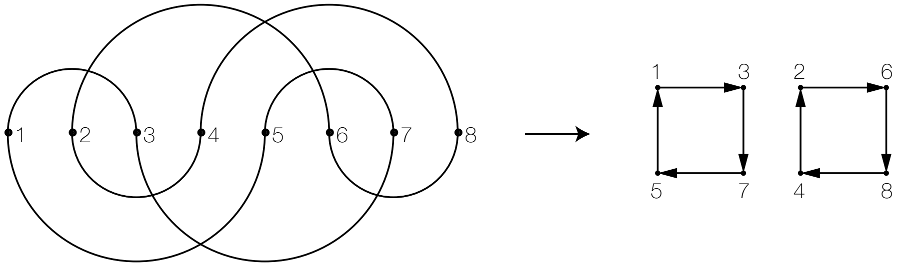


Figure 1.18
The pair of matchings $\{1,3\},\{2,6\},\{4,8\},\{5,7\}$ and $\{1,5\},\{2,4\},\{3,7\},\{6,8\}$ gives the permutation $(1375)(2684)$ with $(-a_{13}a_{26}a_{48}a_{57})(-a_{15}a_{24}a_{37}a_{68}) = a_{13}a_{37}a_{75}a_{51}a_{26}a_{68}a_{84}a_{42}$.

Then, for $\{i_1 j_1 \dots i_m j_m\} = \{1,\dots,2m\}$, $s_M = s_{i_1 j_1} \cdots s_{i_m j_m}$ is nonzero if and only if $\{i_1, j_1\},\dots,\{i_m, j_m\}$ is a perfect matching of G.

We say that our edge orientation is **Pfaffian** if all the perfect matchings of G have the same sign. At the moment there is no efficient test to determine whether a graph admits a Pfaffian orientation. There is a simple combinatorial restatement: An orientation is Pfaffian if and only if every even cycle C for which $G \backslash V(C)$ has a perfect matching has an odd number of edges in each direction.

Fortunately, we have the following result of Kasteleyn [115, 131]:

Every planar graph has a Pfaffian orientation.

This is very desirable, because Theorem 1.4.12 implies the following:

For a Pfaffian orientation of G, $m(G) = \sqrt{\det S(G)}$.

Therefore the number of matchings of a planar graph is reduced to the evaluation of a combinatorial determinant. We will see in Section 1.4.2 that there are many techniques at our disposal to carry out this evaluation.

Let us illustrate this method with an important example, due to Kasteleyn [115] and Temperley–Fisher [198].

Theorem 1.4.13 *The number $m(R_{a,b})$ of matchings of the $a \times b$ rectangular grid $R_{a,b}$ (where we assume b is even) is*

$$m(R_{a,b}) = 4^{\lfloor a/2 \rfloor (b/2)} \prod_{j=1}^{\lfloor a/2 \rfloor} \prod_{k=1}^{b/2} \left(\cos^2 \frac{\pi j}{a+1} + \cos^2 \frac{\pi k}{b+1} \right) \sim c \cdot e^{\frac{G}{\pi} ab} \sim c \cdot 1.3385^{ab},$$

where $G = 1 - \frac{1}{9} + \frac{1}{25} - \frac{1}{49} + \cdots$ is Catalan's constant.

Clearly this is also the number of domino tilings of an $a \times b$ rectangle.

Sketch of Proof. Orient all columns of $R_{a,b}$ going up, and let the rows alternate between going right or left, assigning the same direction to all edges of the same row. The resulting orientation is Pfaffian because every square has an odd number of edges in each direction. The adjacency matrix S satisfies $m(R_{a,b}) = \sqrt{\det S}$. To compute this determinant, it is slightly easier * to consider the following $mn \times mn$ matrix B:

$$b_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are horizontal neighbors} \\ i & \text{if } i \text{ and } j \text{ are vertical neighbors} \\ 0 & \text{otherwise.} \end{cases}$$

We can obtain B from the S by scaling the rows and columns by suitable powers of i, so we still have $m(R_{a,b}) = \sqrt{|\det B|}$. We will prove the product rule for this determinant in Section 1.4.2.

We then use this product formula to give an asymptotic formula for $m(R_{a,b})$. Note that $\log m(R_{a,b})/ab$ may be regarded as a Riemann sum; as $m,n \to \infty$ it converges to

$$c = \frac{1}{\pi^2} \int_0^{\pi/2} \int_0^{\pi/2} \log(4\cos^2 x + 4\cos^2 y)\, dx\, dy = \frac{G}{\pi}$$

where G is Catalan's constant. Therefore

$$m(R_{a,b}) \approx e^{\frac{G}{\pi}ab} \approx 1.3385^{ab}.$$

Loosely speaking, this means that in a matching of the rectangular grid there are about 1.3385 degrees of freedom per vertex. ∎

Obviously, this beautiful formula is not an efficient method of computing the exact value of $m(R_{a,b})$ for particular values of a and b; it is not even clear why it gives an integer! There are alternative determinantal formulas for this quantity that are more tractable; see for example [2, Section 10.1].

1.4.1.6 Routings: the Lindström–Gessel–Viennot lemma

Let G be a directed graph with no directed cycles, which has a weight $\mathrm{wt}(e)$ on each edge e. We are most often interested in the unweighted case, where all weights are 1. Let $S = \{s_1, \ldots, s_n\}$ and $T = \{t_1, \ldots, t_n\}$ be two (not necessarily disjoint) sets of vertices, which we call sources and sinks, respectively. A **routing** from S to T is a set of paths $P_1, \ldots, P_n$ from the n sources $s_1, \ldots, s_n$ to the n sinks $t_1, \ldots, t_n$ such that no two paths share a vertex. Let π be the permutation of $[n]$ such that P_i starts at source s_i and ends at sink $t_{\pi(i)}$, and define $\mathrm{sign}(R) = \mathrm{sign}(\pi)$.

Let the weight of a path or a routing be the product of the weights of the edges it contains. Consider the $n \times n$ **path matrix** Q whose (i,j) entry is

$$q_{ij} = \sum_{P \text{ path from } s_i \text{ to } t_j} \mathrm{wt}(P).$$

*In fact, this is the matrix that Kasteleyn uses in his computation.

Theorem 1.4.14 (Lindström–Gessel–Viennot lemma) *Let G be a directed acyclic graph with edge weights, and let $S = \{s_1, \dots, s_n\}$ and $T = \{t_1, \dots, t_n\}$ be sets of vertices in G. Then the determinant of the $n \times n$ path matrix Q is*

$$\det Q = \sum_{R \text{ routing from } S \text{ to } T} \operatorname{sign}(R)\operatorname{wt}(R).$$

In particular, if all edge weights are 1 *and if every routing takes s_i to t_i for all i, then*

$$\det Q = \text{ number of routings from } S \text{ to } T.$$

Proof. We have $\det A = \sum_P \operatorname{sign}(P)\operatorname{wt}(P)$ summing over **all** path systems $P = \{P_1, \dots, P_n\}$ from S to T; we need to cancel out the path systems that are not routings. For each such P, consider the lexicographically first pair of paths P_i and P_j that intersect, and let v be their first vertex of intersection. Now exchange the subpath of P_i from s_i to v and the subpath of P_j from s_j to v, to obtain new paths P_i' and P_j'. Replacing $\{P_i, P_j\}$ with $\{P_i', P_j'\}$, we obtain a new path system $\varphi(P)$ from S to T. Notice that $\varphi(\varphi(P)) = P$, and $\operatorname{sign}(\varphi(P))\operatorname{wt}(\varphi(P)) + \operatorname{sign}(P)\operatorname{wt}(P) = 0$; so for all non-routings P, the path systems P and $\varphi(P)$ cancel each other out. ■

This theorem was also anticipated by Karlin and McGregor [114] in the context of birth-and-death Markov processes.

Determinants via routings. The Lindström-Gessel-Viennot lemma is also a useful combinatorial tool for computing determinants of interest, usually by enumerating routings in a lattice. We illustrate this with several examples.

1. (Binomial determinants) Consider the **binomial determinant**

$$\binom{a_1, \dots, a_n}{b_1, \dots, b_n} = \det\left[\binom{a_i}{b_j}\right]_{1 \le i,j \le n}$$

where $0 \le a_1 < \cdots < a_n$ and $0 \le b_1 < \cdots < b_n$ are integers. These determinants arise as coefficients of the Chern class of the tensor product of two vector bundles. [129] This algebro-geometric interpretation implies these numbers are positive integers; as combinatorialists, we would like to know what they count.

A **SE path** is a lattice path in the square lattice $\mathbb{N}^2$ consisting of unit steps south and east. Consider the sets of points $A = \{A_1, \dots, A_n\}$ and $B = \{B_1, \dots, B_n\}$ where $A_i = (0, a_i)$ and $B_i = (b_i, b_i)$ for $1 \le i \le n$. Since there are $\binom{a_i}{b_j}$ SE paths from A_i to B_j, and since every SE routing from A to B takes A_i to B_i for all i, we have

$$\binom{a_1, \dots, a_n}{b_1, \dots, b_n} = \text{ number of SE routings from } A \text{ to } B.$$

This is the setting in which Gessel and Viennot discovered Theorem 1.4.14; they also evaluated these determinants in several special cases. [88] We now discuss one particularly interesting special case.

2. (Counting permutations by descent set) The **descent set** of a permutation π is the set of indices i such that $\pi_i > \pi_{i+1}$. We now prove that

$$\binom{c_1,\ldots,c_k,n}{0,c_1,\ldots,c_k} = \text{ number of permutations of } [n] \text{ with descent set } \{c_1,\ldots,c_k\},$$

for any $0 < c_1 < \cdots < c_k < n$. It is useful to define $c_0 = 0, c_{k+1} = n$.

Encode such a permutation π by a routing as follows. For each i let f_i be the number of indices $j \leq i$ such that $\pi_j \leq \pi_i$. Note that the descents $c_1,\ldots,c_k$ of π are the positions where f does not increase. Splitting f at these positions, we are left with $k+1$ increasing subwords $f^1,\ldots,f^{k+1}$. Now, for $1 \leq i \leq n+1$ let P_i be the NW path from (c_{i-1},c_{i-1}) to $(0,c_i)$ taking steps north precisely at the steps listed in f^i. These paths give one of the routings enumerated by the binomial determinant in question, and this is a bijection. See Figure 1.19 for an illustration. [88]

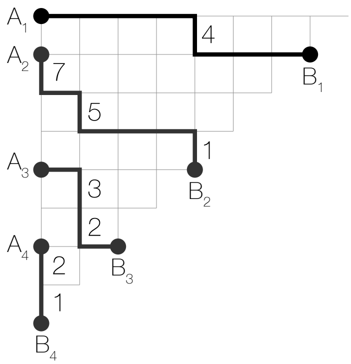


Figure 1.19
The routing corresponding to $\pi = 28351674$ and $f(\pi) = 12.23.157.4$.

3. (Rhombus tilings and plane partitions) Let R_n be the number of tilings of a regular hexagon of side length n using unit rhombi with angles 60° and 120°. Their enumeration is due to MacMahon [133]. There are several equivalent combinatorial models for this problem, illustrated in Figure 1.20, which we now discuss.

Firstly, it is almost inevitable to view these tilings as three-dimensional pictures. This shows that R_n is also the number of ways of stacking unit cubes into the corner of a cubical box of side length n. Incidentally, this three-dimensional view makes it apparent that there are exactly n^2 rhombi of each one of the three possible orientations.

Secondly, we may consider the triangular grid inside our hexagon, and place a dot on the center of each triangle. These dots form a hexagonal grid, where

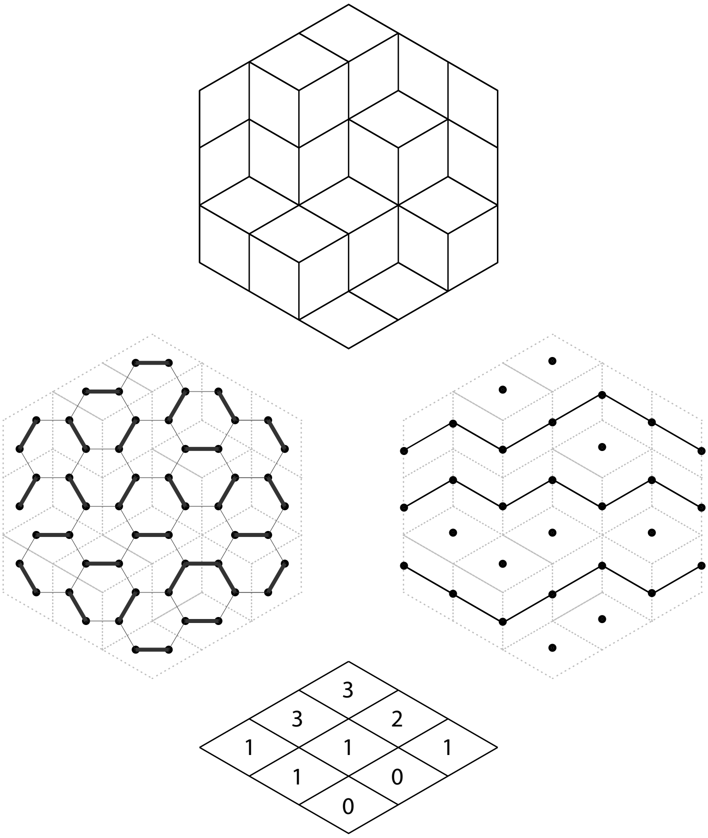


Figure 1.20
Four models for the rhombus tilings of a hexagon.

two dots are neighbors if they are at distance 1 from each other. Finally, join two neighboring dots when the corresponding triangles are covered by a tile. The result is a perfect matching of the hexagonal grid.

Next, on each one of the n^2 squares of the floor of the box, write down the number of cubes above it. The result is a **plane partition**: an array of non-negative integers (finitely many of which are non-zero) that is weakly decreasing in each row and column. We conclude that R_n is also the number of plane partitions whose non-zero entries are at most n, and fit inside an $n \times n$ square.

Finally, given such a rhombus tiling, construct n paths as follows. Each path starts at the center of one of the vertical edges on the western border of the hexagon, and successively crosses each tile splitting it into equal halves. It eventually comes out at the southeast side of the diamond, at the same height where it started (as is apparent from the 3-D picture). The final result is a routing from the n sources $S_1, \ldots, S_n$ on the left to the sinks $T_1, \ldots, T_n$ on the right in the "rhombus" graph shown below. It is clear how to recover the tiling from the routing. Since there are $\binom{2n}{n+i-j}$ paths from S_i to T_j, the Lindström–Gessel–Viennot lemma tells us that R_n is given by the determinant

$$R_n = \det\left[\binom{2n}{n+i-j}\right]_{1\le i,j\le n} = \prod_{i,j,k=1}^{n} \frac{i+j+k-1}{i+j+k-2}.$$

We will prove this product formula in Section 1.4.2.

4. (Catalan determinants, multitriangulations, and Pfaffian rings) The **Hankel matrices** of a sequence $A = (a_0, a_1, a_2, \ldots)$ are

$$H_n(A) = \begin{pmatrix} a_0 & a_1 & \cdots & a_n \\ a_1 & a_2 & \cdots & a_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_n & a_{n+1} & \cdots & a_{2n} \end{pmatrix},$$

and

$$H'_n(A) = \begin{pmatrix} a_1 & a_2 & \cdots & a_{n+1} \\ a_2 & a_3 & \cdots & a_{n+2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n+1} & a_{n+2} & \cdots & a_{2n+1} \end{pmatrix}.$$

Note that if we know the Hankel determinants $\det H_n(A)$ and $\det H'_n(A)$ and they are nonzero for all n, then we can use them as a recurrence relation to recover each a_k from $a_0, \ldots, a_{k-1}$.

There is a natural interpretation of the Hankel matrices of the Catalan sequence $C = (C_0, C_1, C_2, \ldots)$. Consider the "diagonal" grid on the upper half plane with steps $(1,1)$ and $(1,-1)$. Let $A_i = (-2i, 0)$ and $B_i = (2i, 0)$. Then there are C_{i+j} paths from A_i to B_j, and there is clearly a unique routing from $(A_0, \ldots, A_n)$ to $(B_0, \ldots, B_n)$. See Figure 1.21 for an illustration. This proves that $\det H_n(C) = 1$, and an analogous argument proves that $\det H'_n(C) = 1$. Therefore

$$\det H_n(A) = \det H'_n(A) = 1 \text{ for all } n \ge 0 \quad \Longleftrightarrow \quad A \text{ is the Catalan sequence.}$$

The Hankel determinants of the shifted Catalan sequences also arise naturally in several contexts; they are given by:

$$\det \begin{pmatrix} C_{n-2k} & C_{n-2k+1} & \cdots & C_{n-k-1} \\ C_{n-2k+1} & C_{n-2k+2} & \cdots & C_{n-k} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n-k-1} & C_{n-k} & \cdots & C_{n-2} \end{pmatrix} = \prod_{i+j\le n-2k-1} \frac{i+j+2(k-1)}{i+j}. \tag{1.9}$$

There are several ways of proving this equality; for instance, it is a consequence of [122, Theorem 26]. We describe three appearances of this determinant.

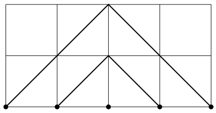


Figure 1.21
Routing interpretation of the Hankel determinant $H_n(C)$.

(a) A **k-fan of Dyck paths of length** $2n$ is a collection of k Dyck paths from $(-n,0)$ to $(n,0)$ that do not cross (although they necessarily share some edges). Shifting the $(i+1)$th path i units up and adding i upsteps at the beginning and i downsteps at the end, we obtain a routing of k Dyck paths starting at the points $A = \{-(n+k-1),\dots,-(n+1),-n\}$ and ending at the points $B = \{n, n+1,\dots,n+k-1\}$ on the x-axis. See Figure 1.22 for an illustration. It follows that the number of k-fans of Dyck paths of length $2(n-2k)$ is given by (1.9).

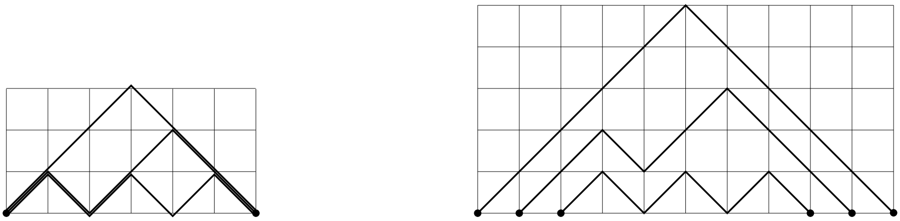


Figure 1.22
A k-fan of Dyck paths.

(b) There is also an extension of the classical one-to-one correspondence between Dyck paths and triangulations of a polygon. Define a **k-crossing** in an n-gon to be a set of k diagonals that cross pairwise. A **k-triangulation** is a maximal set of diagonals with no $(k+1)$-crossings. The main enumerative result, due to Jonsson [110], is that the number of k-triangulations of an n-gon is also given by (1.9). A subtle bijection with fans of Dyck paths is given in [177].

Several properties of triangulations extend non-trivially to this context. For example, every k-triangulation has exactly $k(2n-2k-1)$ diagonals [148, 69]. The k-triangulations are naturally the facets of a simplicial complex called the **multiassociahedron**, which is topologically a sphere [109]; it is not currently known whether it is polytopal. There is a further generalization in the context of Coxeter groups, with connections to cluster algebras [49].

(c) These determinants also arise naturally in the commutative algebraic properties of Pfaffians, defined earlier in this section. Let A be a skew-symmetric $n \times n$ matrix whose entries above the diagonal are indeterminates $\{a_{ij} : 1 \leq i < j \leq n\}$ over a field $\Bbbk$. Consider the **Pfaffian ideal** $I_k(A)$ generated by the $\binom{n}{2k}$ Pfaffian minors of A of size $2k \times 2k$, and the **Pfaffian ring** $R_k(A) = \Bbbk[a_{ij}]/I_k(A)$. Then the multiplicity of the Pfaffian ring $R_k(X)$ is also given by (1.9). [99, 89]

5. (Schröder determinants and Aztec diamonds) Recall from Section 1.3.2.2 that a **Schröder path** of length n is a path from $(0,0)$ to $(2n,0)$ using steps $NE = (1,1), SE = (1,-1)$, and $E = (2,0)$ that stays above the x-axis. The Hankel determinant $\det H_n(R)$ counts the routings of Schröder paths from the points $A = \{0,-2,\ldots,-(2n)\}$ to the points $B = \{0,2,\ldots,2n\}$ on the x-axis.

   These Hankel determinants have a natural interpretation in terms of tilings. Consider the **Aztec diamond** * AD_n consisting of $2n$ rows centered horizontally, consisting successively of $2,4,\ldots,2n,2n,\ldots,4,2$ squares. We are interested in counting the tilings of the Aztec diamond into dominoes.

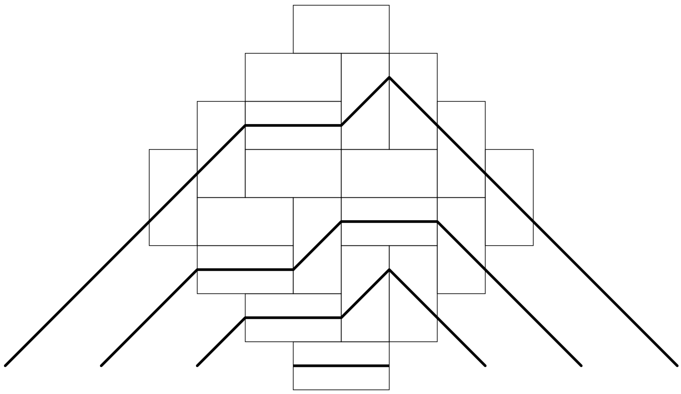


Figure 1.23
A tiling of the Aztec diamond and the corresponding routing.

   Given a domino tiling of AD_n, construct n paths as follows. Each path starts at the center of one of the vertical unit edges on the southwest border of the diamond, and successively crosses each tile that it encounters following a straight line through the center of the tile. It eventually comes out at the

*This shape is called the **Aztec diamond** because it is reminiscent of designs of several Native American groups. Perhaps the closest similarity is with Mayan pyramids, such as the Temple of Kukulcán in Chichén Itzá; the name **Mayan diamond** would have been more appropriate.

southeast side of the diamond, at the same height where it started. See Figure 1.23 for an illustration. If we add i initial NE steps and i final SE steps to the $(i+1)$th path for each i, the result will be a routing of Schröder paths from $A = \{-(2n), \dots, -2, 0\}$ to $B = \{0, 2, \dots, 2n\}$. In fact this correspondence is a bijection [76].

We will prove in Section 1.4.2 that

$$\begin{cases} \det H_n(A) = 2^{n(n-1)/2} \\ \det H'_n(A) = 2^{n(n+1)/2} \end{cases} \text{for all } n \geq 0 \quad \Longleftrightarrow \quad A \text{ is the Schröder sequence.}$$

It will follow that

$$\text{number of domino tilings of the Aztec diamond } AD_n = 2^{n(n+1)/2}.$$

This elegant result is originally due to Elkies, Kuperberg, Larsen, and Propp. For several other proofs, see [73, 74].

1.4.2 Computing determinants

In light of Section 1.4.1, it is no surprise that combinatorialists have become talented at computing determinants. Fortunately, this is a very classical topic with connections to many branches of mathematics and physics, and by now there are numerous general techniques and guiding examples available to us. Krattenthaler's surveys [122] and [123] are excellent references that have clearly influenced the exposition in this section. We now highlight some of the key tools and examples.

1.4.2.1 Is it known?

Of course, when we wish to evaluate a new determinant, one first step is to check whether it is a special case of some known determinantal evaluation. Starting with classical evaluations such as the Vandermonde determinant

$$\det(x_i^{j-1})_{1 \leq i,j \leq n} = \prod_{1 \leq i < j \leq n} (x_j - x_i), \tag{1.10}$$

there is now a wide collection of powerful results at our disposal. A particularly useful one [122, Lemma 3] states that for any $x_1, \dots, x_n, a_2, \dots, a_n, b_2, \dots, b_n$ we have:

$$\det\left[(x_i + b_2) \cdots (x_i + b_j)(x_i + a_{j+1}) \cdots (x_i + a_n)\right]_{1 \leq i,j \leq n} = \prod_{1 \leq i < j \leq n} (x_j - x_i) \prod_{1 \leq i < j \leq n} (b_i - a_j). \tag{1.11}$$

For instance, as pointed out in [89] and [122], the Catalan determinant (1.9) is a special case of this formula. Recognizing it as such is not immediate, but the product formula for Catalan numbers gives an indication of why this is feasible.

In fact, here is a counterintuitive principle: Often the easiest way to prove a determinantal identity is to generalize it. It is very useful to introduce as many parameters as possible into a determinant, while making sure that the more general determinant still evaluates nicely. We will see this principle in action several times in what follows.

1.4.2.2 Row and column operations

A second step is to check whether the standard methods of computing determinants are useful: Laplace expansion by minors, or performing row and column operations until we get a matrix whose determinant we can compute easily. For example, recall the determinant $L_0(K_n)$ of the $(n-1)\times(n-1)$ reduced Laplacian of the complete graph K_n, discussed in Section 1.4.1.3. We can compute it by first adding all rows to the first row, and then adding the first row to all rows:

$$\begin{aligned}\det L_0(K_n) &= \begin{vmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & n-1 \end{vmatrix} = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ -1 & n-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & n-1 \end{vmatrix} \\ &= \begin{vmatrix} 1 & 1 & \cdots & 1 \\ 0 & n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n \end{vmatrix} = n^{n-2},\end{aligned}$$

reproving Theorem 1.4.6.1.

1.4.2.3 Identifying linear factors

Many $n\times n$ determinants of interest have formulas of the form $\det M(\mathbf{x}) = cL_1(\mathbf{x})\cdots L_n(\mathbf{x})$ where c is a constant and the $L_i(\mathbf{x})$ are linear functions in the variables $\mathbf{x}=(x_1,\ldots,x_k)$. We may prove such a formula by first checking that each $L_i(\mathbf{x})$ is indeed a factor of M, and then computing the constant c.

The best known application of this technique is the proof of the formula (1.10) for Vandermonde's determinant $V(x_1,\ldots,x_n)$. If $x_i=x_j$ for $i\neq j$, then rows i and j are equal, and the determinant is 0. It follows that x_i-x_j must be a factor of the polynomial $\det V(x_1,\ldots,x_n)$. Since this polynomial is homogeneous of degree $\binom{n}{2}$, it must equal a constant times $\prod_{i<j}(x_i-x_j)$. Comparing the coefficients of $x_1^0x_2^1\cdots x_n^{n-1}$ we see that the constant equals 1.

A similar argument may be used to prove the more general formula (1.11).

To use this technique, it is sometimes necessary to introduce new variables into our determinant. For example, the formula $\det(i^{j-1})_{1\leq i,j\leq n} = 1^{n-1}2^{n-2}\cdots(n-1)^1$ cannot immediately be treated with this technique. However, the factorization of the answer suggests that this may be a special case of a more general result where this method does apply; in this case, Vandermonde's determinant.

1.4.2.4 Computing the eigenvalues

Sometimes we can compute explicitly the eigenvalues of our matrix, and multiply them to get the determinant. One common technique is to produce a complete set of eigenvectors.

1. (The Laplacian of the complete graph K_n) Revisiting the example above, the Laplacian of the complete graph is $L(K_n) = nI - J$ where I is the identity matrix and J is the matrix all of whose entries equal 1. We first find the eigenvalues of J: 0 is an eigenvalue of multiplicity $n-1$, as evidenced by the linearly independent eigenvectors $\mathbf{e}_1 - \mathbf{e}_2, \ldots, \mathbf{e}_{n-1} - \mathbf{e}_n$. Since the sum of the eigenvalues is $\mathrm{tr}(J) = n$, the last eigenvalue is n; an eigenvector is $\mathbf{e}_1 + \cdots + \mathbf{e}_{n-1}$. Now, if v is an eigenvector for J with eigenvalue λ, then it is an eigenvector for $nI - J$ with eigenvalue $n - \lambda$. Therefore the eigenvalues of $nI - J$ are $n, n, \ldots, n, 0$. Using Theorem 1.4.5, we have reproved yet again that $\det L_0(K_n) = \frac{1}{n}(n^{n-1}) = n^{n-2}$.

2. (The Laplacian of the n-cube C_n) A more interesting example is the reduced Laplacian $L_0(C_n)$ of the graph of the n-dimensional cube, from Theorem 1.4.6.4. By producing explicit eigenvectors, one may prove that if the Laplacians $L(G)$ and $L(H)$ have eigenvalues $\{\lambda_i : 1 \leq i \leq a\}$ and $\{\mu_j : i \leq j \leq b\}$ then the Laplacian of the product graph $L(G \times H)$ has eigenvalues $\{\lambda_i + \mu_j : 1 \leq i \leq a, 1 \leq j \leq b\}$. Since C_1 has eigenvalues 0 and 2, this implies that $C_n = C_1 \times \cdots \times C_1$ has eigenvalues $0, 2, 4, \ldots, 2n$ with multiplicities $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$, respectively. Therefore the number of spanning trees of the cube C_n is
$$\det L_0(C_n) = \frac{1}{2^n} 2^{\binom{n}{1}} 4^{\binom{n}{2}} \cdots (2n)^{\binom{n}{n}} = 2^{2^n - n - 1} 1^{\binom{n}{1}} 2^{\binom{n}{2}} \cdots n^{\binom{n}{n}}.$$

3. (The perfect matchings of a rectangle) An even more interesting example comes from the perfect matchings of the $a \times b$ rectangle, which we discussed in Section 1.4.1.5. Let V be the $4mn$-dimensional vector space of functions $f : [2m] \times [2n] \to \mathbb{C}$, and consider the linear transformation $L : V \to V$ given by
$$(Lf)(x,y) = f(x-1,y) + f(x+1,y) + if(x,y-1) + if(x,y+1),$$
where $f(x,y) = 0$ when $x \in \{0, a+1\}$ or $y \in \{0, b+1\}$. The matrix of this linear transformation is precisely the one we are interested in. A straightforward computation shows that the following are eigenfunctions and eigenvalues of L:
$$g_{k,l}(x,y) = \sin\frac{k\pi x}{a+1} \sin\frac{l\pi y}{b+1}, \qquad \lambda_{k,l} = 2\cos\frac{k\pi}{a+1} + 2i\cos\frac{l\pi}{b+1}$$
for $1 \leq k \leq a$ and $1 \leq l \leq b$. (Note that $g_{k,l}(x,y) = 0$ for $x \in \{0, a\}$ or $y \in \{0, b\}$.) This is then the complete list of eigenvalues for L, so
$$\det L = 2^{ab} \prod_{k=1}^{a} \prod_{l=1}^{b} \left(\cos\frac{k\pi}{a+1} + i\cos\frac{l\pi}{b+1} \right),$$
which is easily seen to equal the expression in Theorem 1.4.13.

1.4.2.5 LU factorizations

A classic result in linear algebra states that, under mild hypotheses, a square matrix M has a unique factorization

$$M = LU$$

where L is a lower triangular matrix and U is an upper triangular matrix with all diagonal entries equal to 1. Computer algebra systems can compute the LU-factorization of a matrix, and if we can guess and prove such a factorization it will follow immediately that $\det M$ equals the product of the diagonal entries of L.

An interesting application of this technique is the determinant

$$\det(\gcd(i,j))_{1\le i,j\le n} = \prod_{i=1}^{n} \varphi(i), \tag{1.12}$$

where $\varphi(k) = \{i \in \mathbb{N} : (\gcd(i,k) = 1 \text{ and } 1 \le i \le k\}$ is Euler's totient function. This is a special case of a more general formula for semilattices that is easier to prove. For this brief computation, we assume familiarity with the Möbius function μ and the zeta function ζ of a poset; these will be treated in detail in Section 1.5.5.3.

Let P be a finite meet semilattice and consider any function $F : P \times P \to \Bbbk$. We will prove the **Lindström–Wilf determinantal formula**:

$$\det F(p \vee q, p)_{p,q\in P} = \prod_{p\in P}\left(\sum_{r\ge p} \mu(p,r)F(r,p)\right). \tag{1.13}$$

Computing some examples will suggest that the LU factorization of F is $F = MZ$ where

$$M_{pq} = \begin{cases} \sum_{r\ge q} \mu(q,r)F(r,p) & \text{if } p \le q, \\ 0 & \text{otherwise,} \end{cases} \qquad Z_{pq} = \begin{cases} 1 & \text{if } p \ge q, \\ 0 & \text{otherwise.} \end{cases}$$

This guess is easy to prove, and it immediately implies (1.13). In turn, applying the Lindström–Wilf to the poset of integers $\{1,\dots,n\}$ ordered by reverse divisibility and the function $F(x,y) = x$, we obtain (1.12).

Another interesting special case is the determinant

$$\det(x^{\mathrm{rank}(p\vee q)})_{p,q\in P} = \prod_{p\in P}\left(x^{\mathrm{rank}(p)}\,\chi_{[p,\hat{1}]}(1/x)\right),$$

where $\chi_{[p,\hat{1}]}(x)$ is the characteristic polynomial of the interval $[x,\hat{1}]$. When P is the partition lattice Π_n, this determinant arises in Tutte's work on the Birkhoff-Lewis equations [203].

1.4.2.6 Hankel determinants and continued fractions

For Hankel determinants, the following connection with continued fractions [209] is extremely useful. If the expansion of the generating function for a sequence $f_0, f_1, \dots$

as a **J-fraction** is

$$\sum_{n=0}^{\infty} f_n x^n = \cfrac{f_0}{1 + a_0 x - \cfrac{b_1 x^2}{1+a_1 x - \cfrac{b_2 x^2}{1+a_2 x - \cdots}}},$$

then the Hankel determinants of $f_0, f_1, \ldots$ equal

$$\det H_n(A) = f_0^n b_1^{n-1} b_2^{n-2} \cdots b_{n-2}^2 b_{n-1}$$

For instance, using the generating function for the Schröder numbers r_n, it is easy to prove that

$$\sum_{n=0}^{\infty} r_n x^n = \cfrac{1}{1 - 2x - \cfrac{2x^2}{1-3x - \cfrac{2x^2}{1-3x-\cdots}}}, \qquad \sum_{n=0}^{\infty} r_{n+1} x^n = \cfrac{2}{1 - 3x - \cfrac{2x^2}{1-3x - \cfrac{2x^2}{1-3x-\cdots}}}.$$

Therefore

$$\det H_n(R) = 2^{n(n-1)/2}, \qquad \det H_n'(R) = 2^{n(n+1)/2},$$

as stated in Example 5 of Section 1.4.1.6.

By computer calculation, it is often easy to guess J-fractions experimentally. With a good guess in place, there is an established procedure for proving their correctness, rooted in the theory of orthogonal polynomials; see [122, Section 2.7].

Dodgson condensation. It is often repeated that Lewis Carroll, author of *Alice in Wonderland*, was also an Anglican deacon and a mathematician, publishing under his real name, Rev. Charles L. Dodgson. His contributions to mathematics are discussed less often, and one of them is an elegant method for computing determinants.

To compute an $n \times n$ determinant A, we create a square pyramid of numbers, consisting of $n+1$ levels of size $n+1, n \ldots, 1$, respectively. On the bottom level we place an $(n+1) \times (n+1)$ array of 1s, and on the next level we place the $n \times n$ matrix A. Each subsequent floor is obtained from the previous two by the following rule: Each new entry is given by $f = (ad - bc)/e$ where f is directly above the entries $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and two floors above the entry e. * The top entry of the pyramid is the determinant. For example, the computation

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \to \begin{pmatrix} 2 & 7 & 5 & 4 \\ 1 & 9 & 7 & 7 \\ 2 & 3 & 2 & 1 \\ 5 & 7 & 6 & 3 \end{pmatrix} \to \begin{pmatrix} 11 & 4 & 7 \\ -15 & -3 & -7 \\ -1 & 4 & 0 \end{pmatrix} \to \begin{pmatrix} 3 & -1 \\ -21 & 28 \end{pmatrix} \to (21)$$

shows that the determinant of the 4×4 determinant is 21.

Dodgson's condensation method relies on the following fact, due to Jacobi. If A is an $n \times n$ matrix and $A_{i_1,\ldots,i_k; j_1,\ldots,j_k}$ denotes the matrix A with rows $i_1, \ldots, i_k$ and columns $j_1, \ldots, j_k$ removed, then

$$\det A \cdot \det A_{1,n;1,n} = \det A_{1;1} \cdot \det A_{n;n} - \det A_{1;n} \cdot \det A_{n;1}. \tag{1.14}$$

*Special care is required when 0s appear in the interior of the pyramid.

This proves that the numbers appearing in the pyramid are precisely the determinants of the "contiguous" submatrices of A, consisting of consecutive rows and columns.

If we have a guess for the determinant of A, as well as the determinants of its contiguous submatrices, Dodgson condensation is an extremely efficient method to prove it. All we need to do is to verify that our guess satisfies (1.14).

To see how this works in an example, let us use Dodgson condensation to prove the formula in Section 1.4.1.6 for $R_n = \det \binom{2n}{n+i-j}_{1\le i,j\le n}$, the number of stacks of unit cubes in the corner of an $n\times n\times n$ box. The first step is to guess the determinant of the matrix in question, as well as all its contiguous submatrices; they are all of the form $R(a,b,c) = \det \binom{a+b}{a+i-j}_{1\le i,j\le c}$, where $a+b=2n$. This more general determinant is equally interesting combinatorially: it counts the stacks of unit cubes in the corner of an $a\times b\times c$ box. By computer experimentation, it is not too difficult to arrive at the following guess:

$$R(a,b,c) = \det\left[\binom{a+b}{a+i-j}\right]_{1\le i,j\le c} = \prod_{i=1}^{a}\prod_{j=1}^{b}\prod_{k=1}^{c}\frac{i+j+k-1}{i+j+k-2}.$$

Proving this formula by Dodgson condensation is then straightforward; we just need to check that our conjectural product formula holds for $c=0,1$ and that it satisfies (1.14); that is,

$$R(a,b,c+1)R(a,b,c-1) = R(a,b,c)^2 - R(a+1,b-1,c)R(a-1,b+1,c).$$

For more applications of Dodgson condensation, see for example [3].

There is a wonderful connection between Dodgson condensation, Aztec diamonds, and **alternating sign matrices**, which we now describe. Let us construct a square pyramid of numbers where levels $n+2$ and $n+1$ are given by two matrices $\mathbf{y} = (y_{ij})_{1\le i,j\le n+2}$ and $\mathbf{x} = (x_{ij})_{1\le i,j\le n+1}$, respectively, and levels $n-1,\ldots,2,1$ are computed in terms of the lower rows using Dodgson's recurrence $f = (ad-bc)/e$. Let $f_n(\mathbf{x},\mathbf{y})$ be the entry at the top of the pyramid.

Remarkably, all the entries of the resulting pyramid will be Laurent monomials in the x_{ij}s and y_{ij}s; that is, their denominators are always monomials. This is obvious for the first few levels, but it becomes more and more surprising as we divide by more and more intricate expressions.

The combinatorial explanation for this fact is that each entry in the $(n-k)$th level of the pyramid encodes the domino tilings of an Aztec diamond AD_k. For instance, if $n=3$, the entry at the top of the pyramid is

$$\begin{aligned} f_2(\mathbf{x},\mathbf{y}) = {} & \frac{x_{11}x_{22}x_{33}}{y_{22}y_{33}} - \frac{x_{11}x_{23}x_{32}}{y_{22}y_{33}} - \frac{x_{12}x_{21}x_{33}}{y_{22}y_{33}} + \frac{x_{12}x_{21}x_{23}x_{32}}{x_{22}y_{22}y_{33}} \\ & - \frac{x_{12}x_{21}x_{23}x_{32}}{x_{22}y_{23}y_{32}} + \frac{x_{12}x_{23}x_{31}}{y_{23}y_{32}} + \frac{x_{13}x_{21}x_{32}}{y_{23}y_{32}} - \frac{x_{13}x_{31}x_{32}}{y_{23}y_{32}} \end{aligned}$$

There is a simple bijection between the eight terms of f_3 and the eight domino tilings of AD_2. Regard a tiling of AD_2 as a graph with vertices on the underlying lattice, and add a vertical edge above and below the tiling, and a horizontal edge to the

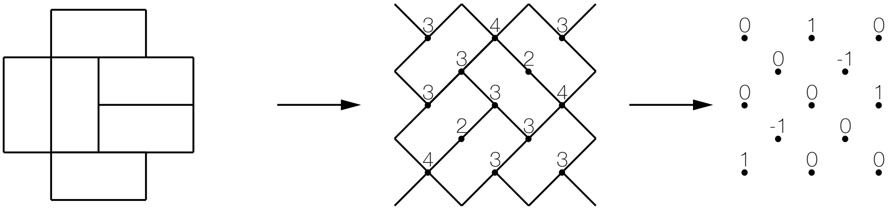


Figure 1.24
A domino tiling of AD_2 and the corresponding monomial in $f_2(\mathbf{x},\mathbf{y})$.

left and to the right of T. Now rotate the tiling 45°. Record the degree of each vertex, ignoring the outside corners on the boundary of the diamond, and subtract 3 from each vertex. This leaves us with an $n \times n$ grid of integers within an $(n+1) \times (n+1)$ grid of integers. Assign to it the monomial whose x exponents are given by the outer grid and whose y exponents are given by the inner grid. For example, the tiling in Figure 1.24 corresponds to the monomial $(x_{12}x_{23}x_{31})/(y_{23}y_{32})$.

In general, this gives a bijection between the terms of $f_n(\mathbf{x},\mathbf{y})$ and the domino tilings of the Aztec diamond AD_n. One may also check that there are no cancellations, so Dodgson condensation tells us that the number m_n of terms in f_n satisfies $m_{n-1}m_{n+1} = 2m_n^2$. This gives an alternative proof that the Aztec diamond AD_n has $2^{n(n+1)/2}$ domino tilings.

We may also consider the patterns formed by the x_{ij}s by themselves (or of the y_{ij}s by themselves). In each individual monomial of $f_n(\mathbf{x},\mathbf{y})$, the exponents of the x_{ij}s form an $n \times n$ **alternating sign matrix** (ASM): a matrix of 1s, 0s, and -1s such that the nonzero entries in any row or column alternate $1, -1, \ldots, -1, 1$. Similarly, the negatives of the exponents of the y_{ij}s form an ASM of size $n-1$.

Alternating sign matrices are fascinating objects in their own right, with connections to representation theory, statistical mechanics, and other fields. The number of alternating sign matrices of size n is

$$\frac{1!\,4!\,7!\cdots(3n-2)!}{n!\,(n+1)!\,(n+2)!\cdots(2n-1)!}.$$

For details on the history and solution of this difficult enumeration problem see [44, 167, 222].

1.5 Posets

This section is devoted to the enumerative aspects of the theory of partially ordered sets (posets). Section 1.5.1 introduces key definitions and examples. Section 1.5.2 discusses some families of lattices that are of special importance. In Section 1.5.3 we count chains and linear extensions of posets.

The remaining sections are centered around the *Möbius Inversion Formula*, which is perhaps the most useful enumerative tool in the theory of posets. This formula helps us count sets that have an underlying poset structure; it applies to many combinatorial settings of interest.

In Section 1.5.4 we discuss the Inclusion-Exclusion Formula, a special case of great importance. In Section 1.5.5 we introduce Möbius functions and the Möbius Inversion Formula. In particular, we catalog the Möbius functions of many important posets. The incidence algebra, a nice algebraic framework for understanding and working with the Möbius function, is discussed in Section 1.5.5.3. In Section 1.5.5.4 we discuss methods for computing Möbius functions of posets, and sketch proofs for the posets of Section 1.5.5. Finally, in Section 1.5.6, we discuss Eulerian posets and the enumeration of their flags, which gives rise to the **ab**-index and **cd**-index.

1.5.1 Basic definitions and examples

A **partially ordered set** or **poset** $(P, \leq)$ is a set P together with a binary relation $\leq$, called a **partial order**, such that

- For all $p \in P$, we have $p \leq p$.
- For all $p, q \in P$, if $p \leq q$ and $q \leq p$ then $p = q$.
- For all $p, q, r \in P$, if $p \leq q$ and $q \leq r$ then $p \leq r$.

We say that $p < q$ if $p \leq q$ and $p \neq q$. We say that p and q are **comparable** if $p < q$ or $p > q$, and they are **incomparable** otherwise. We say that q **covers** p if $q > p$ and there is no $r \in P$ such that $q > r > p$. When q covers p we write $q \gtrdot p$.

Example 1.5.1 *Many sets in combinatorics come with a natural partial order, and often the resulting poset structure is very useful for enumerative purposes. Some of the most important examples are the following:*

1. (Chain) *The poset* $\mathbf{n} = \{1, 2, \ldots, n\}$ *with the usual total order.* $(n \geq 1)$
2. (Boolean lattice) *The poset* 2^A *of subsets of a set* A*, where* $S \leq T$ *if* $S \subseteq T$.
3. (Divisor lattice) *The poset* D_n *of divisors of* n*, where* $c \leq d$ *if* c *divides* d. $(n \geq 1)$
4. (Young's lattice) *The poset* Y *of integer partitions, where* $\lambda \leq \mu$ *if* $\lambda_i \leq \mu_i$ *for all* i.
5. (Partition lattice) *The poset* Π_n *of set partitions of* $[n]$*, where* $\pi \leq \rho$ *if* π refines ρ*; that is, if every block of* ρ *is a union of blocks of* π. $(n \geq 1)$
6. (Non-crossing partition lattice) *The subposet* NC_n *of* Π_n *consisting of the non-crossing set partitions of* $[n]$*, where there are no elements* $a < b < c < d$ *such that* a, c *are together in one block and* b, d *are together in a different block.* $(n \geq 1)$

7. (Bruhat order on permutations) *The poset S_n of permutations of $[n]$, where π covers ρ if π is obtained from ρ by choosing two adjacent numbers $\rho_i = a < b = \rho_{i+1}$ in ρ and exchanging their positions. ($n \geq 1$)*

8. (Subspace lattice) *The poset $L(\mathbb{F}_q^n)$ of subspaces of a finite dimensional vector space $\mathbb{F}_q^n$, where $U \leq V$ if U is a subspace of V. ($n \geq 1$, q a prime power)*

9. (Distributive lattice) *The poset $J(P)$ of order ideals of a poset P (subsets $I \subseteq P$ such that $j \in P$ and $i < j$ imply $i \in P$) ordered by containment.*

10. (Face poset of a polytope) *The poset $F(P)$ of faces of a polytope P, ordered by inclusion.*

11. (Face poset of a subdivision of a polytope) *The poset $\widehat{\mathscr{T}}$ of faces of a subdivision $\mathscr{T}$ of a polytope P ordered by inclusion, with an additional maximum element.*

12. (Subgroup lattice of a group) *The poset $L(G)$ of subgroups of a group G, ordered by containment.*

The **Hasse diagram** of a finite poset P is obtained by drawing a dot for each element of P and an edge going down from p to q if p covers q. Figure 1.25 shows the Hasse diagrams of some of the posets above. In particular, the Hasse diagram of $2^{[n]}$ is the 1-skeleton of the n-dimensional cube.

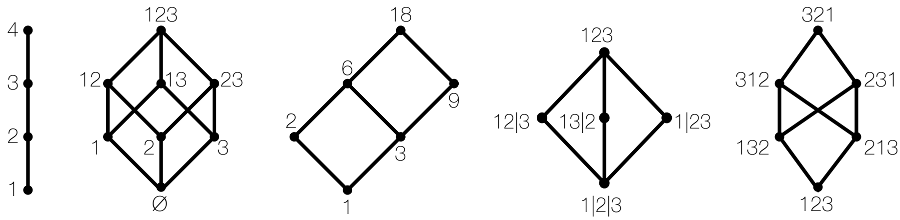


Figure 1.25
The Hasse diagrams of the chain **4**, Boolean lattice $2^{[3]}$, divisor lattice D_{18}, partition lattice Π_3, and Bruhat order S_3.

A subset Q of P is a **chain** if every pair of elements is comparable, and it is an **antichain** if every pair of elements is incomparable. The length of a chain C is $|C| - 1$. If there is a **rank function** $r : P \to \mathbb{N}$ such that $r(x) = 0$ for any minimal element x and $r(y) = r(x) + 1$ whenever $y \gtrdot x$, then P is called **graded** or **ranked**. The largest rank is called the **rank** or **height** of P. The **rank-generating function** of a finite graded poset is

$$R(P;x) = \sum_{p \in P} x^{r(p)}.$$

All the posets of Example 1.5.1 are graded except for subgroup lattices.

A poset P induces a poset structure on any subset $Q \subseteq P$; a special case of interest is the **interval** $[p,q] = \{r \in P : p \leq r \leq q\}$. We call a poset **locally finite** if all its intervals are finite. Given posets P and Q on disjoint sets, the **direct sum** $P+Q$ is the poset on $P \cup Q$ inheriting the order relations from P and Q, and containing no additional order relations between elements of P and Q. The **direct product** $P \times Q$ is the poset on $P \times Q$ where $(p,q) \leq (p',q')$ if $p \leq p'$ and $q \leq q'$.

We have already seen examples of product posets. The Boolean lattice is $2^A \cong \mathbf{2} \times \cdots \times \mathbf{2}$. Also, if $n = p_1^{t_1} \cdots p_k^{t_k}$ is the prime factorization of n, then $D_n \cong (\mathbf{t_1}+\mathbf{1}) \times \cdots \times (\mathbf{t_k}+\mathbf{1})$.

1.5.2 Lattices

A poset is a **lattice** if every two elements p and q have a least upper bound $p \vee q$ and a greatest lower bound $p \wedge q$, called their **meet** and **join**, respectively. We will see this additional algebraic structure can be quite beneficial for enumerative purposes.

Example 1.5.2 *All the posets in Example 1.5.1 are lattices, except for the Bruhat order. In most cases, the meet and join have easy descriptions. In* $\mathbf{n}$*, the meet and join are the minimum and maximum, respectively. In* 2^A *they are the intersection and union. In* D_n *they are the greatest common divisor and least common multiple. In* Y *they are the componentwise minimum and maximum. In* Π_n *and in* NC_n *the meet of two partitions* π *and* ρ *is the collection of intersections of a block of* π *and a block of* ρ*. In* $L(\mathbb{F}_q^n)$ *the meet and join are the intersection and the span. In* $J(P)$ *they are the intersection and the union. In* $F(P)$ *the meet is the intersection. In* $L(G)$ *the meet is the intersection.*

Any lattice must have a unique minimum element $\widehat{0}$ and maximum element $\widehat{1}$. An element covering $\widehat{0}$ is called an **atom**; an element covered by $\widehat{1}$ is called a **coatom**. To prove that a finite poset P is a lattice, it is sufficient to check that it has a $\widehat{1}$ and that any $x,y \in P$ have a meet; then the join of x and y will be the (necessarily non-empty) meet of their common upper bounds. Similarly, it suffices to check that P has a $\widehat{0}$ and that any $x,y \in P$ have a join.

Distributive lattices. A lattice L is **distributive** if the join and meet operations satisfy the distributive properties:

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z), \qquad x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z) \tag{1.15}$$

for all $x,y,z \in L$. To prove that L is distributive, it is sufficient to verify that one of the equations in (1.15) holds for all $x,y,z \in L$.

Example 1.5.3 *There are several distributive lattices in Example 1.5.1: the chains* $\mathbf{n}$*, the Boolean lattices* 2^A*, the divisor lattices* D_n*, and Young's lattice* Y*. This follows from the fact that the pairs of operations* $(\min,\max)$*,* $(\gcd,\text{lcm})$ *and* $(\cap,\cup)$ *satisfy the distributive laws. The others are not necessarily distributive; for example,* Π_3 *and* S_3*.*

The most important, and in fact, the only, source of finite distributive lattices is the construction of Example 1.5.1.9: Given a poset P, a **downset** or **order ideal** I is a subset of P such that if $i \in I$ and $j < i$ then $j \in I$. A **principal** order ideal is one of the form $P_{\leq p} = \{q \in P : q \leq p\}$. Let $J(P)$ be the **poset of order ideals** of P, ordered by inclusion.

Theorem 1.5.4 (Fundamental Theorem for Finite Distributive Lattices) *A poset L is a distributive lattice if and only if there exists a poset P such that $L \cong J(P)$.*

Sketch of Proof. Since the collection of order ideals of a poset P is closed under union and intersection, $J(P)$ is a sublattice of 2^P. The distributivity of 2^P then implies that $J(P)$ is a distributive lattice.

For the converse, let L be a distributive lattice, and let P be the set of join-irreducible elements of L; that is, the elements $p > \widehat{0}$ that cannot be written as $p = q \vee r$ for $q, r < p$. These are precisely the elements of L that cover exactly one element. The set P inherits a partial order from L, and this is the poset such that $L \cong J(P)$. The isomorphism is given by

$$\begin{aligned} \phi : J(P) &\longrightarrow L \\ I &\longmapsto \bigvee_{p \in I} p \end{aligned}$$

and the inverse map is given by $\phi^{-1}(l) = \{p \in P : p \leq l\}$. ■

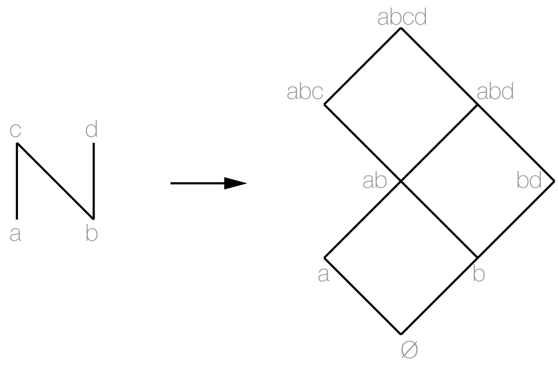


Figure 1.26
A poset and the corresponding distributive lattice.

Theorem 1.5.4 extends to some infinite posets with minor modifications. Let $J_f(P)$ be the set of finite order ideals of a poset P. Then the map $P \mapsto J_f(P)$ is a bijection between the posets whose principal order ideals are finite and the locally finite distributive lattices with $\widehat{0}$.

Example 1.5.5 *The posets P of join-irreducibles of the distributive lattices $L \cong J(P)$ of Example 1.5.3 are as follows. For $L = \mathbf{n}$, $P = \mathbf{n} - \mathbf{1}$ is a chain. For $L = 2^A$, $P = \mathbf{1} + \cdots + \mathbf{1}$ is an antichain. For $L = D_n$, where $n = p_1^{t_1} \cdots p_k^{t_k}$, $P = \mathbf{t_1} + \cdots + \mathbf{t_k}$ is the disjoint sum of k chains. For $L = Y$, $P = \mathbb{N} \times \mathbb{N}$ is a "quadrant."*

Theorem 1.5.4 explains the abundance of cubes in the Hasse diagram of a distributive lattice L. For any element $l \in L$ covered by n elements $l_1, \ldots, l_n$ of L, the joins of the 2^n subsets of $\{l_1, \ldots, l_n\}$ are distinct, and form a copy of the Boolean lattice $2^{[n]}$ inside L. The dual result holds as well.

The **width** of a poset P is the size of the largest antichain of P. **Dilworth's theorem** [68] states that this is the smallest integer w such that P can be written as the disjoint union of w chains.

Theorem 1.5.6 *The distributive lattice $J(P)$ can be embedded as an induced subposet of the poset $\mathbb{N}^w$, where w is the width of P.*

Proof. Decompose P as the disjoint union of w chains $C_1, \ldots, C_c$. The map

$$\begin{array}{rcl} \phi : J(P) & \longrightarrow & \mathbb{N}^w \\ I & \longmapsto & (|I \cap C_1|, \ldots, |I \cap C_w|) \end{array}$$

gives the desired inclusion. ∎

Geometric lattices. Now we introduce another family of lattices of great importance in combinatorics. We say that a lattice L is:

- **semimodular** if the following two equivalent conditions hold:
  - L is graded and $r(p) + r(q) \geq r(p \wedge q) + r(p \vee q)$ for all $p, q \in L$.
  - If p and q both cover $p \wedge q$, then $p \vee q$ covers both p and q.
- **atomic** if every element is a join of atoms.
- **geometric** if it is semimodular and atomic.

Example 1.5.7 *In Figure 1.25, the posets $2^{[3]}$ and Π_3 are geometric, while the posets $\mathbf{4}$, D_{18}, and S_3 are not.*

Not surprisingly, the prototypical example of a geometric lattice comes from a natural geometric construction, illustrated in Figure 1.27. Let $A = \{v_1, \ldots, v_n\}$ be a set of vectors in a vector space V. A **flat** is a subspace of V generated by a subset of A. We identify a flat with the set of v_is that it contains. Let L_A be the set of flats of A, ordered by inclusion. Then L_A is a geometric lattice.

The theory of geometric lattices is equivalent to the rich theory of *matroids*, which is the subject of Section 1.8.

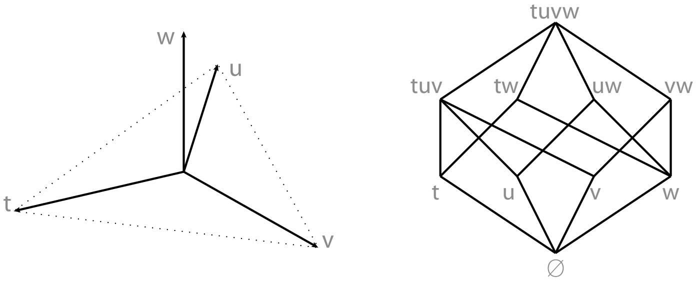


Figure 1.27
A vector configuration $\mathbf{t}, \mathbf{u}, \mathbf{v}, \mathbf{w}$ (where $\mathbf{t}, \mathbf{u}, \mathbf{v}$ are coplanar in $\mathbb{R}^3$) and the corresponding geometric lattice.

Supersolvable lattices. A lattice L is **supersolvable** if there exists a maximal chain, called an M-chain, such that the sublattice generated by C and any other chain of L is distributive. [182]

Again unsurprisingly, an important example comes from supersolvable groups, but there are several other interesting examples. Here is a list of supersolvable lattices, and an M-chain in each case.

1. Distributive lattices: every maximal chain is an M-chain.
2. Partition lattice Π_n: $1|2|\cdots|n < 12|3|\cdots|n < 123|4|\cdots|n < \cdots < 123\cdots n$.
3. Noncrossing partition lattice NC_n: the same chain as above.
4. Lattice of subspaces $L(\mathbb{F}_q^n)$ of the vector space $\mathbb{F}_q^n$ over a finite field $\mathbb{F}_q$: every maximal chain is an M-chain.
5. Subgroup lattices of finite supersolvable groups G: an M-chain is given by any normal series $1 = H_0 \triangleleft H_1 \triangleleft \cdots \triangleleft H_k = G$ where each H_i is normal and H_i/H_{i-1} is cyclic of prime order.

Fortunately, there is a simple criterion to verify semimodularity. An ***R*-labeling** of a poset P is a labeling of the edges of the Hasse diagram of P with integers such that for any $s \leq t$ there exists a unique maximal chain C from s to t.

Theorem 1.5.8 [139] *A finite graded lattice of rank n is supersolvable if and only if it has an R-labeling for which the labels on every maximal chain are a permutation of* $\{1, \ldots, n\}$.

1.5.3 Zeta polynomials and order polynomials

The **zeta polynomial** of a finite poset P counts the multichains of various lengths in P. A **multichain** of length k in P is a sequence of possibly repeated elements

$t_0, t_1, \ldots, t_k \in P$ such that $t_0 \leq t_1 \leq \cdots \leq t_k$. Let

$$Z_P(k) = \text{number of multichains of length } k-2 \text{ in } P \qquad (k \geq 2). \tag{1.16}$$

There is a unique polynomial $Z_P(k)$ satisfying (1.16) for all integers $k \geq 2$; it is given by

$$Z_P(k) = \sum_{i \geq 2} b_i \binom{k-2}{i-2}, \tag{1.17}$$

where b_i is the number of chains of length $i-2$ in P. This polynomial is called the **zeta polynomial** of P.

Example 1.5.9 *The following posets have particularly nice zeta polynomials:*

1. $P = \mathbf{n}$*:*
$$Z(k) = \binom{n+k-2}{n-1}$$

2. $P = B_n$*:*
$$Z(k) = k^n$$

3. $P = NC_n$*: (Kreweras, [124])*
$$Z(k) = \frac{1}{n}\binom{kn}{n-1}$$

The **order polynomial** of P counts the order-preserving labelings of P; it is defined by

$$\Omega_P(k) = \text{number of maps } f : P \to [k] \text{ such that } p < q \text{ implies } f(p) \leq f(q)$$

for $k \in \mathbb{N}$. The next proposition shows that, once again, there is a unique polynomial taking these values at the natural numbers.

Proposition 1.5.10 *For any poset P,* $\Omega_P(k) = Z_{J(P)}(k)$*.*

Proof. An order-preserving map $f : P \to [k]$ gives rise to a sequence of order ideals $f^{-1}(\{1\}) \subseteq f^{-1}(\{1,2\}) \subseteq \cdots f^{-1}(\{1,\ldots,k\})$, which is a multichain in $J(P)$. Conversely, every sequence arises uniquely in this way. ■

A **linear extension** of P is an order-preserving labeling of the elements of P with the labels $1, \ldots, n = |P|$, which extends the order of P; that is, a bijection $f : P \to [n]$ such that $p < q$ implies $f(p) < f(q)$. Let

$$e(P) = \text{ number of linear extensions of } P.$$

It follows from Proposition 1.5.10 and (1.17) that the order polynomial Ω_P has degree $|P|$, and leading coefficient $e(P)/|P|!$.

The following is a method for computing $e(P)$ recursively.

Proposition 1.5.11 *Define $e : J(P) \to \mathbb{N}$ recursively by*

$$e(I) = \begin{cases} 1 & \text{if } I = \widehat{0}, \\ \sum_{J \lessdot I} e(J) & \text{otherwise.} \end{cases}$$

Then $e(\widehat{1})$ is the number $e(P)$ of linear extensions of P.

Proof. Let $p_1, \ldots, p_k$ be the maximal elements of P. In a linear extension of P, one of $p_1, \ldots, p_k$ has to be labeled n, and therefore

$$e(P) = e(P \backslash \{p_1\}) + \cdots + e(P \backslash \{p_k\}).$$

This is equivalent to the desired recurrence. ■

It is useful to keep in mind that $J(P)$ is a subposet of $\mathbb{N}^w$ for $w = w(P)$. The recurrence of Proposition 1.5.11 generalizes Pascal's triangle, which corresponds to the case $P = \mathbb{N} + \mathbb{N}$. When we apply it to $P = \mathbb{N} + \cdots + \mathbb{N}$, we get the recursive formula for multinomial coefficients.

Example 1.5.12 *In some special cases, the problem of enumerating linear extensions is of fundamental importance.*

- $e(\mathbf{n}_1 + \cdots + \mathbf{n}_k) = \binom{n_1 + \cdots + n_k}{n_1, \ldots, n_k}$
- $e(\mathbf{2} \times \mathbf{n}) = C_n = \frac{1}{n+1}\binom{2n}{n}$.
- *Let T be a* tree poset *of n elements, such that the Hasse diagram is a tree rooted at $\widehat{0}$. For each vertex v let $t_v = |T_{\geq v}| = |\{w \in T \,:\, w \geq v\}|$. Then*

  $$e(T) = \frac{n!}{\prod_{v \in T} t_v}.$$

- *Let λ be a Ferrers diagram of n cells, partially ordered by decreeing that each cell is covered by the cell directly below and the cell directly to the right, if they are in λ. The* ***hook*** *H_c of a cell c consists of cells on the same row and to the right of c, those on the same column and below c, and c itself. Let $h_c = |H_c|$. Then*

  $$e(\lambda) = \frac{n!}{\prod_{c \in D} h_c}.$$

  This is the dimension of the irreducible representation of the symmetric group S_n corresponding to λ. [173]

1.5.4 The inclusion-exclusion formula

Our next goal is to discuss one of the most useful enumerative tools for posets: Möbius functions and the Möbius inversion theorem. Before we do that, we devote this section to a special case that preceded and motivated them: the inclusion-exclusion formula.

Theorem 1.5.13 (Inclusion-Exclusion Formula) *For any finite sets* $A_1,\dots,A_n \subseteq X$, *we have*

1. $|A_1\cup\cdots\cup A_n| = \sum_i |A_i| - \sum_{i<j}|A_i\cap A_j| + \sum_{i<j<k}|A_i\cap A_j\cap A_k| - \cdots \pm |A_1\cap\cdots\cap A_n|.$

2. $|\overline{A_1}\cap\cdots\cap\overline{A_n}| = |X| - \sum_i |A_i| + \sum_{i<j}|A_i\cap A_j| - \cdots \pm |A_1\cap\cdots\cap A_n|.$

Proof. It suffices to prove one of these two equivalent equations. To prove the first one, consider an element x appearing in $k\geq 1$ of the given sets. The number of times that x is counted in the right-hand side is $k-\binom{k}{2}+\cdots\pm\binom{k}{k}=1$. ■

We now present a slightly more general formulation.

Theorem 1.5.14 (Inclusion-Exclusion Formula) *Let A be a set and consider two functions* $f_=, f_\geq : 2^A \longrightarrow \Bbbk$ *from* 2^A *to a field* $\Bbbk$. *Then*

1. $f_\geq(S) = \sum_{T\supseteq S} f_=(T)$ *for* $S\subseteq A \iff f_=(S) = \sum_{T\supseteq S}(-1)^{|T-S|} f_\geq(T)$ *for* $S\subseteq A$.

2. $f_\leq(S) = \sum_{T\subseteq S} f_=(T)$ *for* $S\subseteq A \iff f_=(S) = \sum_{T\subseteq S}(-1)^{|S-T|} f_\leq(T)$ *for* $S\subseteq A$.

The most common interpretation is the following. Suppose we have a set U of objects and a set A of properties that each object in U may or may not satisfy. If we know, for each $S\subseteq A$, the number $f_\geq(S)$ of elements having *at least* the properties in S (or the number $f_\leq(S)$ of elements having *at most* the properties in S), then we obtain, for each $S\subseteq A$, the number $f_=(S)$ of elements having *exactly* the properties in S. We are often interested in the number $f_=(\emptyset)$ or $f_=(A)$ of elements satisfying none or all of the given properties.

Theorem 1.5.14 has a simple linear algebraic interpretation. Consider the two $2^A\times 2^A$ matrices C,D whose non-zero entries are $C_{S,T}=1$ for $S\subseteq T$, and $D_{S,T}=(-1)^{|T-S|}$ for $S\subseteq T$. Then the inclusion-exclusion formula is equivalent to the assertion that C and D are inverse matrices. This can be proved directly, but we prefer to deduce it as a special case of the Möbius inversion formula (Theorem 1.5.16). We now present two applications.

Derangements. One of the classic applications of the inclusion-exclusion formula is the enumeration of the **derangements** of $[n]$, which are the permutations $\pi\in S_n$ such that $\pi(i)\neq i$ for all i. Let $A=\{A_1,\dots,A_n\}$ where A_i is the property that $\pi(i)=i$. Then $f_\geq(T)=(n-|T|)!$, so the number D_n of derangements of $[n]$ is

$$\begin{aligned} D_n &= f_=(\emptyset) = \sum_T (-1)^{|T|} f_\geq(T) = \sum_{k=0}^{n}\binom{n}{k}(-1)^k(n-k)! \\ &= n!\left(\frac{1}{0!}-\frac{1}{1!}+\frac{1}{2!}-\cdots\pm\frac{1}{n!}\right). \end{aligned}$$

It follows that D_n is the integer closest to $n!/e$.

Discrete derivatives. Consider the $\Bbbk$-vector space Γ of functions $f : \mathbb{Z} \to \Bbbk$. The **discrete derivative** of f is the function Δf given by $\Delta f(n) = f(n+1) - f(n)$. We now wish to show that, just as with ordinary derivatives,

$$\Delta^{d+1} f = 0 \text{ if and only if } f \text{ is a polynomial of degree at most } d.$$

This was part of Theorem 1.3.6. Regarding Δ as a linear operator on Γ, we have $\Delta = E - 1$ where $Ef(n) = f(n+1)$ and 1 is the identity. Then $\Delta^k = (E-1)^k = \sum_{i=0}^{k} \binom{k}{i} E^i (-1)^{k-i}$, so the kth discrete derivative is $\Delta^k f(n) = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} f(n+i)$.

The functions $f_{\leq}(S) = f(n+|S|)$ and $f_{=}(S) = \Delta^k f(|S|)$ satisfy Theorem 1.5.14.2, so we have $f(n+k) = \sum_{i=0}^{k} \binom{k}{i} \Delta^i f(n)$. (This is equivalent to $E^k = (\Delta+1)^k$.) If $\Delta^{d+1} f = 0$, this gives $f(k) = \sum_{i=0}^{d} \binom{k}{i} \Delta^i f(0)$, which is a polynomial in k of degree at most d. The converse follows from the observation that Δ lowers the degree of a polynomial by 1.

1.5.5 Möbius functions and Möbius inversion

1.5.5.1 The Möbius function

Given a locally finite poset P, let $\mathrm{Int}(P) = \{[x,y] : x, y \in P, x \leq y\}$ be the set of intervals of P. The (two-variable) Möbius function of a poset P is the function $\mu : \mathrm{Int}(P) \to \mathbb{Z}$ defined by

$$\sum_{p \leq r \leq q} \mu(p,r) = \begin{cases} 1 & \text{if } p = q, \\ 0 & \text{otherwise.} \end{cases} \tag{1.18}$$

Here we are denoting $\mu(p,q) = \mu([p,q])$. We will later see that the Möbius function can be defined equivalently by the equations:

$$\sum_{p \leq r \leq q} \mu(r,q) = \begin{cases} 1 & \text{if } p = q, \\ 0 & \text{otherwise.} \end{cases} \tag{1.19}$$

When P has a minimum element $\widehat{0}$, the (one-variable) **Möbius function** $\mu : P \to \mathbb{Z}$ is $\mu(x) = \mu(\widehat{0}, x)$. If P also has a $\widehat{1}$, the **Möbius number** of P is $\mu(P) = \mu(\widehat{0}, \widehat{1})$.

Computing the Möbius function is a very important problem, because the Möbius function is the poset analog of a derivative; and as such, it is a fundamental invariant of a poset. This problem often leads to very interesting enumerative combinatorics, as can be gleaned from the following gallery of Möbius functions.

Theorem 1.5.15 *The Möbius functions of some key posets are as follows.*

1. *(Chain)* $P = \mathbf{n}$*:*

$$\mu_{\mathbf{n}}(i,j) = \begin{cases} 1 & \textit{if } j = i \\ -1 & \textit{if } j = i+1 \\ 0 & \textit{otherwise.} \end{cases}$$

2. *(Boolean lattice)* $P = 2^A$*:*

$$\mu_{2^A}(S,T) = (-1)^{T-S}.$$

3. *(Divisor lattice)* $P = D_n$*: We have* $\mu_{D_n}(k,l) = \mu(l/k)$ *where*

$$\mu(m) = \begin{cases} (-1)^t & \text{if } m \text{ is a product of } t \text{ distinct primes, and} \\ 0 & \text{otherwise.} \end{cases}$$

is the classical ***Möbius function*** *from number theory.*

4. *(Young's lattice)* $P = Y$*:*

$$\mu(\lambda,\mu) = \begin{cases} (-1)^{|\mu-\lambda|} & \text{if } \mu-\lambda \text{ has no two adjacent squares, and} \\ 0 & \text{otherwise.} \end{cases}$$

5. *(Partition lattice)* $P = \Pi_n$*: The Möbius number of* Π_n *is*

$$\mu(\Pi_n) = (-1)^{n-1}(n-1)!,$$

from which a (less elegant) formula for the complete Möbius function can be derived.

6. *(Non-crossing partition lattice)* $P = NC_n$*: The Möbius number of* NC_n *is*

$$\mu(NC_n) = (-1)^{n-1}C_{n-1},$$

where C_{n-1} *is the* $(n-1)$*th Catalan number. This gives a (less elegant) formula for the complete Möbius function.*

7. *(Bruhat order)* $P = S_n$*:*

$$\mu(u,v) = (-1)^{\ell(v)-\ell(u)},$$

where the ***length*** $\ell(w)$ *of a permutation* $w \in S_n$ *is the number of* ***inversions*** (i,j) *where* $1 \le i < j \le n$ *and* $w_i > w_j$*. (There is a generalization of this result to the Bruhat order on any Coxeter group* W*, or even on a parabolic subgroup* W^J*; see Section 1.5.5.4.)*

8. *(Subspace lattice)* $P = L(\mathbb{F}_q^n)$*:*

$$\mu(U,V) = (-1)^d q^{\binom{d}{2}},$$

where $d = \dim V - \dim U$*.*

9. *(Distributive lattice)* $L = J(P)$*:*

$$\mu(I,J) = \begin{cases} (-1)^{|J-I|} & \text{if } J-I \text{ is an antichain in } P\text{, and} \\ 0 & \text{otherwise.} \end{cases}$$

10. *(Face poset of a polytope)* $L = F(P)$*:*

$$\mu(F,G) = (-1)^{\dim G - \dim F}.$$

11. *(Face poset of a subdivision* $\mathscr{T}$ *of a polytope* P*)* $L = \widehat{T}$*:*

$$\mu(F,G) = \begin{cases} (-1)^{\dim G - \dim F} & \text{if } G < \widehat{1} \\ (-1)^{\dim P - \dim F + 1} & \text{if } G = \widehat{1} \text{ and } F \text{ is not on the boundary of } P \\ 0 & \text{if } G = \widehat{1} \text{ and } F \text{ is on the boundary of } P. \end{cases}$$

12. *(Subgroup lattice of a finite* p*-group) If* $|G| = p^n$ *for* p *prime,* $n \in \mathbb{N}$*, then in* $L = L(G)$*:*

$$\mu(A,B) = \begin{cases} (-1)^k p^{\binom{k}{2}} & \text{if } A \text{ is a normal subgroup of } B \text{ and } B/A \cong \mathbb{Z}_p^k, \text{ and} \\ 0 & \text{otherwise}. \end{cases}$$

Some of the formulas above follow easily from the definitions, while others require more sophisticated methods. In the following sections we will develop some of the basic theory of Möbius functions and discuss the most common methods for computing them. Along the way, we will sketch proofs of all the formulas above.

It is worth remarking that a version of item 11 of Theorem 1.5.15 holds more generally for the face poset of any finite regular cell complexes Γ such that the underlying space $|\Gamma|$ is a manifold with or without boundary; see [194, Prop. 3.8.9] for details.

1.5.5.2 Möbius inversion

In enumerative combinatorics, there are many situations where we have a set U of objects, and a natural way of assigning to each object u of U an element $f(u)$ of a poset P. We are interested in counting the objects in U that map to a particular element $p \in P$. Often we find that it is much easier to count the objects in s that map to an element **less than or equal to** * p in P. The following theorem tells us that this easier enumeration is sufficient for our purposes, as long as we can compute the Möbius function of P.

Theorem 1.5.16 (Möbius Inversion formula) *Let* P *be a poset and let* $f,g : P \to \Bbbk$ *be functions from* P *to a field* $\Bbbk$*. Then*

1. $\forall p \in P \;\; g(p) = \sum_{q \geq p} f(q) \quad \Longleftrightarrow \quad \forall p \in P \;\; f(p) = \sum_{q \geq p} \mu(p,q) g(q)$ *and*
2. $\forall p \in P \;\; g(p) = \sum_{q \leq p} f(q) \quad \Longleftrightarrow \quad \forall p \in P \;\; f(p) = \sum_{q \leq p} \mu(q,p) g(q).$

In his paper [169], which pioneered the use of the Möbius inversion formula as a tool for counting in combinatorics, Rota described this enumerative philosophy as follows:

*or greater than or equal to