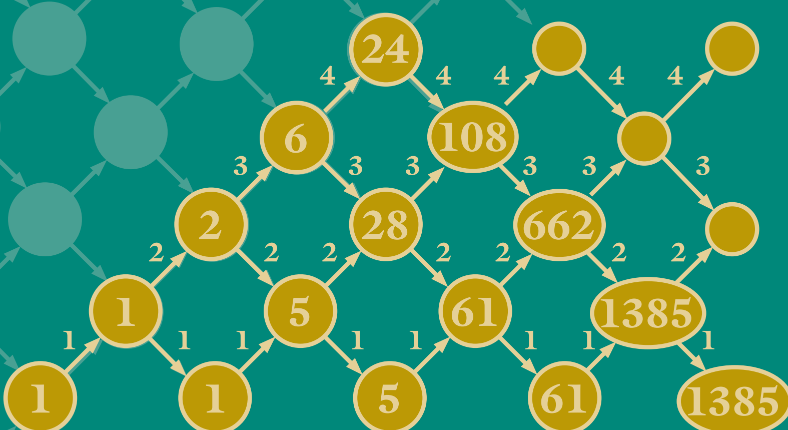
STUDENT MATHEMATICAL LIBRARY
Volume 23

Lectures on Generating Functions

S. K. Lando



$$\Longrightarrow \quad \cfrac{1}{1-\cfrac{1^2 s^2}{1-\cfrac{2^2 s^2}{1-\cfrac{3^2 s^2}{1-\dots}}}}$$

AMS
American Mathematical Society

Lectures on Generating Functions

STUDENT MATHEMATICAL LIBRARY
Volume 23

Lectures on Generating Functions

S. K. Lando

AMS
AMERICAN MATHEMATICAL SOCIETY

This work was originally published in Russian by МЦНМО under the title "Лекции о производящих функциях" ©2002.
The present translation was created under license for the American Mathematical Society and is published by permission.

Translated from the Russian by the author.

2000 *Mathematics Subject Classification.* Primary 05A15, 05C30.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-23

Library of Congress Cataloging-in-Publication Data

Lando, S. K. (Sergei K.), 1955–
[Lektsii o proizvodiaschikh funktsiiakh. English]
Lectures on generating functions / S. K. Lando.
p. cm. — Student mathematical library ; ISSN 1520-9121 ; v. 23
Includes bibliographical references and index
ISBN 0-8218-3481-9 (alk. paper)
1. Generating functions. I. Title. II. Series.

QA164.8.L3613 2003
511′.6–dc22 200355939

∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at http://www.ams.org/

10 9 8 7 6 5 4 3 2 1   08 07 06 05 04 03

To A. A. Kirillov,
from whom I have first heard the words
"generating function"

Contents

Preface to the English Edition

Modern combinatorics speaks the language of generating functions. The study of this language does not require a bulky knowledge of numerous parts of mathematics; although some preliminary acquaintance with calculus and algebra is more than welcome. On the other hand, generating functions may prove to be extremely useful in further mathematical education because of their deep involvement in various mathematical activities, including computer science. The goal of the present book is to serve as a basis for a one-semester undergraduate course in combinatorics, based on the notion of generating function. It contains many exercises both for class and home work. Of course, it is an introductory book not containing a complete theory. I hope, however, that some of its readers will find in it a good entrance point into the fascinating world of generating functions.

All of the main ideas in the book are introduced on the basis of examples. Sometimes the choice of examples is classical, and in other cases it is justified by my own research experience. This experience concerns first of all graph embeddings into two-surfaces and enumeration of the embeddings. This subject plays a central role in contemporary theoretical physics, and specialists know that it incorporates many advanced mathematical theories. A variety of generating functions appears naturally in these studies and some of them found their way into this book.

I would like to use this opportunity to express my gratitude to the American Mathematical Society for the suggestion to publish the English translation of the book. In the translation, some minor corrections and changes were made.

Sergei Lando, July 2003

Preface

After multiplying by $(2n-1)!$, the coefficient of x^{2n-1} in the power expansion of the function $\tan x$ becomes a positive integer. What is more surprising, this number appears to be equal to the number of up-down permutations of the set $\{1, \ldots, 2n-1\}$. This shows that $\tan x$ is the "exponential generating function" for the sequence of numbers of up-down permutations. This fact can be proved, but we cannot be sure that we understand the phenomenon completely. The function $\tan x$ is not unique in this sense: coefficients in the expansions of many classical functions have a combinatorial interpretation. Trigonometric, hypergeometric and elliptic functions, elliptic integrals and so on fall into this class. One can even affirm that the coefficients of every function which is interesting by itself and not only as an element of some functional class must have a combinatorial meaning.

Mathematicians of the 18th and 19th centuries knew functions "personally". I doubt whether there are more specialists nowadays possessing these skills than there were a hundred years ago, in spite of the fact that the roots, the asymptotics, the disk of convergence, the singularities, and the topology of the corresponding Riemann surface can say a lot about the nature of the objects under enumeration.

Generating functions admit a natural splitting into classes. The simplest is the class of rational functions. It is well studied and a huge bunch of problems leading to rational generating functions is known.

Algebraic generating functions also appear frequently. In the beginning of 1960s Schützenberger showed that their non-commutative analogues arise naturally as languages generated by unambiguous formal grammars. However, the class of algebraic functions (in contrast to that of rational ones) is not closed under the natural operation of the Hadamard product. Generally, the Hadamard product of two algebraic functions is an algebro-logarithmic function. And the class of algebro-logarithmic functions, which is closed under the Hadamard product, seems to be natural.

The relationship between algebraic functions and formal grammars indicates that the class of objects under enumeration is essentially one-dimensional: words in languages admit a linear recording. Modern quantum field theory models require enumeration of objects of essentially two-dimensional origin, and the nature of generating functions arising in these problems is far from being understood completely. The elegant method of matrix integration invented by physicists leads to explicit results only in a few cases.

I wanted to write a simple and accessible introduction to generating functions, paying attention first of all to striking examples, not to (often non-existing) general theories. As a result, many important applications of the generating functions method, including Polya's enumeration theory and q-analogues, Poincaré's generating polynomials and generating families, the theory of ramified coverings and many other important topics are not even mentioned in the book.

My interest in enumerative combinatorics was inspired by a series of problems posed by V. I. Arnold in connection with some problems of the singularity theory as well as his own activities in this field. I was influenced a lot by the combinatorial team of the University Bordeaux I (G. Viennot and others) and by P. Flajolet. The book is based on the series of optional courses I gave for many years to freshmen of the Higher College of Mathematics of the Independent University of Moscow in 1992–99. In giving these courses, I enjoyed substantial help from M. N. Vyalyi, who also helped greatly in preparing the book for publication. The main source of my knowledge in combinatorics is my

friend and long-time coauthor Alexander Zvonkin, whose mastery of creating texts is — alas — beyond my reach.

S. K. Lando

Chapter 1

Formal Power Series and Generating Functions. Operations with Formal Power Series. Elementary Generating Functions

1.1. The lucky tickets problem

In the early 70s A. A. Kirillov usually opened his seminar with the following problem. A bus passenger, then, had to buy a ticket from a cashier. The tickets had a 6-digit number.

A ticket is said to be *lucky* if the sum of the first three digits of its number coincides with the sum of the last three digits.

Thus, the ticket with the number 123060 is lucky, while the one with the number 123456 is not lucky.

Now, *how many lucky tickets are there*?

A person possessing elementary skills in computer programming will have no difficulty in writing a computer program counting the number of lucky tickets. The simplest such program just searches through all numbers from 000000 to 999999 selecting the lucky ones. However, let us look at the problem more attentively.

First, split all lucky tickets into classes formed by tickets with a given sum of the first three digits. This sum runs from 0 (for the triple 000) to 27 (for the triple 999). Therefore, the total number of classes is 28. Denote by a_n the number of triples of digits with the sum n. It is easy to compute the first values of a_n:

- $a_0 = 1$ (there is a single triple with the sum 0);
- $a_1 = 3$ (the three triples 001, 010, 001 have the sum 1);
- $a_2 = 6$ (the triples are 002, 020, 200, 011, 101, 110).

The number of lucky tickets with the sum of the first three digits equal to n is a_n^2. Indeed, we are able to put arbitrary triple of digits with the same sum n both at the beginning and at the end of the number of a lucky ticket. Hence, in order to compute the number of lucky tickets it suffices to compute the numbers a_n and then to find the sum of their squares.

Before computing a_n, let us first compute the number of one- and two-digit numbers with the sum n. For each $n = 0, 1, 2, \dots, 9$ there is a single one-digit number with the sum of digits n (when written down, this number just coincides with n written down). We will describe one-digit numbers by the polynomial

$$A_1(s) = 1 + s + s^2 + \dots + s^9.$$

The coefficients of this polynomial have the following meaning:

the coefficient of s^n in the polynomial A_1 coincides with the number of one-digit numbers having the sum of digits equal to n.

In other words, the coefficient of s^n in A_1 is 1 provided that $0 \le n \le 9$ and is 0 for $n > 9$.

Now let us write down the polynomial $A_2(s)$ which describes the two-digit numbers. The coefficient of s^n in $A_2(s)$ is the number of two-digit numbers having the sum n. (We take into account also two-digit numbers such that their first, or even both digits are 0.)

It is easy to see that the degree of A_2 is 18. Indeed, 18 is the largest possible sum of the digits of a two-digit number. The computation of the first few coefficients of this polynomial also meets no trouble:

$$A_2(s) = 1 + 2s + 3s^2 + 4s^3 + \dots$$

It turns out that the polynomial A_2 is closely related to the polynomial A_1.

Statement 1.1. $A_2(s) = (A_1(s))^2$.

Proof. The product of two monomials s^k and s^m contributes to the coefficient of the monomial s^n in the polynomial $(A_1(s))^2$ if and only if $n = k + m$. Therefore, the coefficient of s^n in $(A_1(s))^2$ is exactly the number of ways to represent n as a sum $n = k + m, \quad k, m = 0, 1, \dots, 9$. Hence, the polynomial on the right-hand side of the identity coincides with A_2.

Now we are able to write down the polynomial $A_3(s) = a_0 + a_1 s + \dots + a_{27} s^{27}$.

Statement 1.2. $A_3(s) = (A_1(s))^3$.

Proof. The proof repeats that of the previous statement almost word for word: the coefficient of s^n in the polynomial $(A_1(s))^3$ is equal to the number of representations of n as a sum of three digits, $n = m + k + l, \quad m, k, l = 0, 1, \dots, 9$.

Thus, the lucky tickets problem is almost solved; it remains only to compute the polynomial $(A_1(s))^3$ and then find the sum of squares of its coefficients. Note that the multiplication by the polynomial $A_1(s)$ is a rather simple operation. The calculation can be done by hand and it takes about 10 minutes. No computer program is required.

However, being not absolutely satisfied with the result, we can proceed further[1]. The approach to the lucky tickets problem which we explain below belongs to V. Drinfeld, then a high school student.

[1] The reader may skip the rest of the section since it requires some knowledge of calculus and will not be used in the future.

Together with the polynomial $A_3(s)$, consider the "Laurent polynomial" $A_3(1/s)$ in the variable s:

$$A_3\left(\frac{1}{s}\right) = a_0 + \frac{a_1}{s} + \frac{a_2}{s^2} + \cdots + \frac{a_{27}}{s^{27}}.$$

The product $A_3(s)A_3\left(\frac{1}{s}\right)$ also is a Laurent polynomial (i.e., it contains monomials of the form s^k both with positive and negative k's, but the values of k are bounded from below as well as from above). The free term in this product (the coefficient of s^0) has the form $a_0^2 + a_1^2 + \cdots + a_{27}^2$, and we conclude that

the number of lucky tickets coincides with the free term of the Laurent polynomial $A_3(s)A_3(1/s)$.

This free term can be computed using the basic fact of the theory of functions of one complex variable, the Cauchy theorem.

Theorem 1.3 (Cauchy)**.** *For any Laurent polynomial $p(s)$ its free term p_0 is*

$$p_0 = \frac{1}{2\pi i}\int \frac{p(s)ds}{s},$$

where the integral is taken over an arbitrary counterclockwise oriented circle in the complex plane containing the origin.

In other words, the integral of $s^k ds$ over such a circle is $2\pi i$ if $k = -1$, and it is 0 otherwise. This fact can be easily verified.

The most convenient circle for our purposes is the unit circle centered at the origin. Using the fact that

$$A_1(s) = 1 + s + \cdots + s^9 = \frac{1 - s^{10}}{1 - s}$$

one can represent the Laurent polynomial in question in the form

$$\begin{aligned} P(s) &= A_3(s)A_3\left(\frac{1}{s}\right) = A_1^3(s)A_1^3\left(\frac{1}{s}\right) \\ &= \left(\frac{1-s^{10}}{1-s}\right)^3\left(\frac{1-s^{-10}}{1-s^{-1}}\right)^3 = \left(\frac{2-s^{10}-s^{-10}}{2-s-s^{-1}}\right)^3. \end{aligned}$$

Introducing the standard parameter φ in the unit circle and restricting the Laurent polynomial $P(s)$ to this circle we obtain the

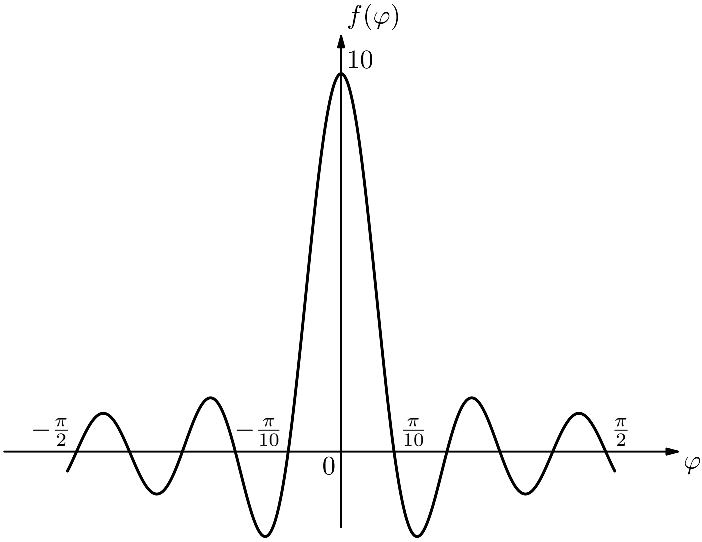


Figure 1. The shape of the graph $f(\varphi) = \dfrac{\sin(10\varphi)}{\sin\varphi}$

following expression for the free term of the polynomial:

$$p_0 = \frac{1}{2\pi}\int_0^{2\pi}\left(\frac{2-2\cos(10\varphi)}{2-2\cos\varphi}\right)^3 d\varphi = \frac{1}{2\pi}\int_0^{2\pi}\left(\frac{\sin^2(5\varphi)}{\sin^2\left(\frac{\varphi}{2}\right)}\right)^3 d\varphi$$

$$(1.1)\quad = \frac{1}{\pi}\int_0^{\pi}\left(\frac{\sin(10\varphi)}{\sin\varphi}\right)^6 d\varphi = \frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\left(\frac{\sin(10\varphi)}{\sin\varphi}\right)^6 d\varphi.$$

Let us try to estimate the value of the last integral. The graph of the function $f(\varphi) = \frac{\sin(10\varphi)}{\sin\varphi}$ on the segment $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ is shown in Fig. 1. The function has the maximum, equal to 10, at the origin. The value of f out of the segment $\left[-\frac{\pi}{10}, \frac{\pi}{10}\right]$ is less than $\frac{1}{\sin\frac{\pi}{10}} \approx 3$. Therefore, the contribution of the complement to the segment $\left[-\frac{\pi}{10}, \frac{\pi}{10}\right]$ to the integral (1.1) is less than $\pi \cdot 3^6 \approx 2100$ (actually, this contribution is significantly smaller).

The main contribution to the integral (1.1) is produced by the segment $\left[-\frac{\pi}{10}, \frac{\pi}{10}\right]$. To estimate this contribution we make use of the *method of the stationary phase*. This method allows one to find the asymptotics of the integral

$$\int_{-\frac{\pi}{10}}^{\frac{\pi}{10}} f^t d\varphi = \int_{-\frac{\pi}{10}}^{\frac{\pi}{10}} e^{t\ln f} d\varphi$$

as $t \to \infty$. For t large enough the value of the integral is determined by the behavior of the function $\ln f$ (the "phase") in the neighborhood of its stationary point 0 (the point, where $(\ln f)' = 0$, or, what is the same, $f' = 0$). In a neighborhood of the stationary point we have $f(\varphi) \approx 10\left(1 - \frac{33}{2}\varphi^2\right)$, and $\ln f(\varphi) \approx \ln 10 - \frac{33}{2}\varphi^2$. For t large enough this gives

$$\int_{-\frac{\pi}{10}}^{\frac{\pi}{10}} e^{t\left(\ln 10 - \frac{33}{2}\varphi^2\right)} d\varphi = e^{t\ln 10}\int_{-\frac{\pi}{10}}^{\frac{\pi}{10}} e^{-\frac{33}{2}t\varphi^2} d\varphi \approx e^{t\ln 10}\frac{\sqrt{2\pi}}{\sqrt{33t}}.$$

Setting $t = 6$ and recalling Eq. (1.1) we obtain the value

$$p_0 \approx \frac{10^6}{3\sqrt{11\pi}} \approx 56700.$$

This result approximates the exact value reasonably well (the error is not greater than 3%).

Another approach to enumeration of lucky tickets is explained in Sec. 7.1.

1.2. First conclusions

The example considered in the previous section allows one to make some conclusions concerning the problems we are going to study and the methods we will use.

The main subject of our study will be problems of enumerative combinatorics. They concern enumeration of objects belonging to a family of finite sets. Each set has a number (in the lucky ticket problem the number is the sum n of the three leftmost digits).

As a rule, an enumerative problem is solvable "in principle": we are able to write down all elements belonging to each set of the family and find their number. The problem, however, is to find a "good" solution without exhausting all elements.

On the other hand, it is a complicated task to define what a good solution is. We are able often only to compare two solutions and say which one of them is better.

When solving enumerative problems, generating polynomials (or, more generally, generating series) are of great use. In our case, the generating polynomial A_3 happened to be extremely useful. Operations with combinatorial objects can be naturally translated into operations with generating functions. For instance, the passage from one-digit numbers to three-digit ones consists just in taking the cube of the the generating polynomial A_1.

We have also seen that methods of close mathematical areas (analysis, for example) can provide new points of view on combinatorial problems and allow one to find unexpected approaches to solving them.

1.3. Generating functions and operations with them

Now let us give formal definitions.

Definition 1.4. Let $a_0, a_1, a_2, \dots$ be an arbitrary (infinite) sequence of numbers. The *generating function* (*generating series*) for this sequence is the expression of the form

$$a_0 + a_1 s + a_2 s^2 + \dots,$$

or, briefly,

$$\sum_{n=0}^{\infty} a_n s^n.$$

If all elements in the sequence starting from some element are equal to zero, then the corresponding generating function is called a *generating polynomial.*

The elements of the sequence a_n may be of arbitrary nature. We will consider sequences of natural, integer, rational, real, and complex numbers. As is usual for ordinary functions, we will often denote a generating function by a single letter followed by the argument in brackets:

$$A(s) = a_0 + a_1 s + a_2 s^2 + \dots$$

Remark 1.5. When using the term "function" we do not mean that the expression we write down indeed is a function. For example, we are not able to say what is the "value $A(s_0)$ of a function A at a point s_0". The variable s is *formal* and the sum of the series $a_0+a_1s_0+a_2s_0^2+\dots$ makes no sense. However, the statement $A(0) = a_0$ is true, that is, the value of a generating function at 0 is well defined.

A generating function represents a number sequence as a series in powers of the formal variable. That is why we use the term "formal power series" along with the term "generating function".

Definition 1.6. The *sum* of two generating functions

$$A(s) = a_0 + a_1s + a_2s^2 + \dots$$

and

$$B(s) = b_0 + b_1s + b_2s^2 + \dots$$

is the generating function

$$A(s) + B(s) = (a_0 + b_0) + (a_1 + b_1)s + (a_2 + b_2)s^2 + \dots$$

The *product* of the generating functions A and B is the generating function

$$A(s)B(s) = a_0b_0 + (a_0b_1 + a_1b_0)s + (a_0b_2 + a_1b_1 + a_2b_0)s^2 + \dots$$

Obviously, both the addition and the multiplication are commutative and associative.

Definition 1.7. Let

$$A(s) = a_0 + a_1s + a_2s^2 + \dots$$

and

$$B(t) = b_0 + b_1t + b_2t^2 + b_3t^3 + \dots$$

be two generating functions and suppose $B(0) = b_0 = 0$. The *substitution* of B into A is the generating function

$$A(B(t)) = a_0+a_1b_1t+(a_1b_2+a_2b_1^2)t^2+(a_1b_3+2a_2b_1b_2+a_3b_1^3)t^3+\dots$$

If, for example, $B(t) = -t$, then

$$A(B(t)) = A(-t) = a_0 - a_1 t + a_2 t^2 - a_3 t^3 + \ldots$$

Note that the substitution of a function different from zero at 0 is not well defined. Its application would cause the necessity to sum up infinite number series.

Of course, if both functions A and B are polynomials, then their sum, their product, and the result of the substitution coincide with the usual sum, product and substitution for polynomials.

For a closer look on generating functions let us prove the following important theorem.

Theorem 1.8 (about the inverse function)**.** *Let a function*

$$B(t) = b_1 t + b_2 t^2 + b_3 t^3 + \ldots$$

be such that $B(0) = b_0 = 0$, *and* $b_1 \neq 0$. *Then there exist functions*

$$A(s) = a_1 s + a_2 s^2 + a_3 s^3 + \ldots, \quad A(0) = 0,$$

and

$$C(u) = c_1 u + c_2 u^2 + c_3 u^3 + \ldots, \quad C(0) = 0,$$

such that

$$A(B(t)) = t \quad \text{and} \quad B(C(u)) = u.$$

Each of the functions A and C is a unique function possessing this property.

The function A is said to be *left inverse* and the function C is said to be *right inverse* to the function B.

Proof. Let us prove the existence and uniqueness of the left inverse function. For the right inverse function the proof is similar. We compute the coefficients of the function A step by step. The coefficient a_1 is the solution of the equation $a_1 b_1 = 1$, whence

$$a_1 = \frac{1}{b_1}.$$

Now suppose the coefficients $a_1, a_2, \ldots, a_n$ are already known. Then the coefficient a_{n+1} is the solution of the equation

$$a_{n+1} b_1^{n+1} + \cdots = 0,$$

where dots denote some polynomial in $a_1, \dots, a_n$ and $b_1, \dots, b_n$. Hence, the equation is a linear equation with respect to a_{n+1} and the coefficient of a_{n+1} is b_1^{n+1}. This coefficient is non-zero, therefore, the equation has a unique solution and the proof of the theorem is completed.

Thus, we have learned how to add, multiply and substitute power series. We would also like to divide power series. The division is not always well defined. In this respect it reminds the division of integers: the ratio of two integers is not necessarily an integer. However, we are always capable to divide by a power series whose value at 0 is non-zero.

Statement 1.9. *Let*

$$A(s) = a_0 + a_1 s + a_2 s^2 + a_3 s^3 + \dots$$

be a formal power series such that $A(0) \neq 0$*. Then there exists a formal power series*

$$B(s) = b_0 + b_1 s + b_2 s^2 + b_3 s^3 + \dots$$

such that $A(s)B(s) = 1$*.*

Proof. Once again we proceed by induction. We have $b_0 = \frac{1}{a_0}$. Now suppose that all coefficients of the series B up to the degree $n-1$ are known. The coefficient of s^n is determined from the condition

$$a_0 b_n + a_1 b_{n-1} + \dots + a_n b_0 = 0.$$

This is a linear equation with respect to b_n, and the coefficient a_0 of b_n in this equation is non-zero. Therefore, the equation has a unique solution.

1.4. Elementary generating functions

Writing each generating function as a power series is not always convenient. Therefore, some frequently used generating functions have a special short notation.

Definition 1.10.

1) $(1+s)^\alpha = 1 + \frac{\alpha}{1!}s + \frac{\alpha(\alpha-1)}{2!}s^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!}s_3 + \dots,$
where $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n$ and α is an arbitrary complex number.

Coefficients in this generating function are called the *binomial coefficients*; the nth binomial coefficient is denoted by

$$\binom{\alpha}{n} = \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!};$$

2) $e^s = \exp s = 1 + \frac{1}{1!}s + \frac{1}{2!}s^2 + \frac{1}{3!}s^3 + \dots;$

3) $\ln\left(\frac{1}{1-s}\right) = s + \frac{1}{2}s^2 + \frac{1}{3}s^3 + \dots;$

4) $\sin s = s - \frac{1}{3!}s^3 + \frac{1}{5!}s^5 - \dots;$

5) $\cos s = 1 - \frac{1}{2!}s^2 + \frac{1}{4!}s^4 - \dots$

Expansion 1) in Definition 1.10 was introduced by Newton and is called the *Newton binomial*. For positive integer values of α it coincides with the usual definition of the power of the binomial. This fact allows one to deduce elementary combinatorial identities. For example, the substitutions $s=1$ and $s=-1$ yield respectively

$$(1.2)\quad \binom{\alpha}{0} + \binom{\alpha}{1} + \dots + \binom{\alpha}{\alpha} = 1 + \frac{\alpha}{1!} + \frac{\alpha(\alpha-1)}{2!} + \dots + \frac{\alpha!}{\alpha!} = 2^\alpha,$$

$$(1.3)\quad \binom{\alpha}{0} - \binom{\alpha}{1} + \dots + (-1)^\alpha\binom{\alpha}{\alpha} = 0$$

for arbitrary positive integer α.

Besides, the functions introduced above satisfy some natural relations which also have combinatorial meaning. Let us prove, for example, that

$$e^s e^{-s} = 1.$$

Indeed, the free term in the product on the left is 1, while for $n>0$ the coefficient of s^n is

$$\frac{1}{n!\,0!} - \frac{1}{(n-1)!\,1!} + \frac{1}{(n-2)!\,2!} - \dots + \frac{(-1)^n}{0!\,n!}.$$

Multiplying this expression by $n!$ we reduce it to the left-hand side of Eq. (1.3) for $\alpha = n$, which completes the proof of the statement.

1.5. Differentiating and integrating generating functions

Definition 1.11. Let $A(s) = a_0 + a_1 s + a_2 s^2 + \dots$ be a generating function. The *derivative* of this function is the function

$$A'(s) = a_1 + 2a_2 s + 3a_3 s^2 + \dots + n a_n s^{n-1} + \dots$$

The *integral* is the function

$$\int A(s) = a_0 s + a_1 \frac{s^2}{2} + a_2 \frac{s^3}{3} + \dots + a_n \frac{s^{n+1}}{(n+1)} + \dots$$

The differentiation is the inverse operation to the integration:

$$\left(\int A(s)\right)' = A(s).$$

On the contrary, the integration of a derivative leads to a function with the zero free term; therefore, generally speaking, the result differs from the original function by an additive constant.

Remark 1.12. It is easy to see that for functions admitting a power series expansion the formal derivative coincides with the ordinary one. The formula for the integral corresponds to the conventional integral with the variable right end of the interval,

$$\int A(s) = \int_0^s A(\xi) d\xi.$$

The last remark allows one to compute (that is, to express in terms of elementary functions) generating functions for various sequences. Let us compute, for example, the generating function

$$f(s) = \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} s + \frac{1}{3 \cdot 4} s^2 + \dots + \frac{1}{(n+1)(n+2)} s^n + \dots$$

Multiplying f by s^2 and taking the derivative we obtain

$$(s^2 f(s))' = s + \frac{1}{2} s^2 + \frac{1}{3} s^3 + \dots = \ln(1-s)^{-1},$$

whence

$$f(s) = s^{-2} \int \ln(1-s)^{-1}$$
$$= s^{-2} \left((s-1)\ln(1-s)^{-1} + s \right).$$

1.6. The algebra and the topology of formal power series

The reader will find below some information from the theory of formal power series. This information will not be used in the future, but it can help to place this theory among other mathematical theories.

From the algebraic point of view, the set of formal power series (with coefficients in the field of complex, real, or rational numbers) forms an (infinite dimensional) *vector space* over this field. The multiplication makes this vector space into an *algebra* denoted by $\mathbb{C}[[s]]$ (respectively $\mathbb{R}[[s]]$ and $\mathbb{Q}[[s]]$). An important role is played by *ideals* in this algebra, i.e., subsets $I \subset \mathbb{C}[[s]]$ such that $fI \subset I$ for any $f \in \mathbb{C}[[s]]$. All ideals in the algebra of formal power series are *principal*, i.e., each of them has the form $f\mathbb{C}[[s]]$ for some element $f \in \mathbb{C}[[s]]$. Moreover, it is easy to describe all these ideals: they are $I_k = s^k\mathbb{C}[[s]],\ k = 0, 1, 2, \ldots$ (this means that the ideal I_k consists of all power series divisible by s^k). One of the ideals I_k, namely I_1, is the maximal one: it is not contained in any other ideal different from the algebra itself. An algebra having a single maximal ideal is said to be *local*. The locality property shows that the algebra of formal power series is close to coordinate algebras in a neighborhood of the origin (the algebras of *germs* of smooth or analytic functions). The *quotient algebras* $\mathbb{C}[[s]]/I_k$ are called the *algebras of truncated polynomials* and are also very useful.

The algebra of formal power series is equipped with a *topology*. The ideals $I_k,\ k = 0, 1, 2, \ldots$, as well as the empty set are the *open sets* in this topology. This topology defines the notion of *convergence*: a sequence $F_1(s), F_2(s), \ldots$ *converges* to a formal power series $F(s)$ if for any n there is a number N such that the coefficients of $s^0, s^1, \ldots, s^n$ in all the series $F_k(s)$ for $k > N$ coincide with those in $F(s)$. In particular, the sequence of partial sums of a power series $F(s)$ converges to $F(s)$.

1.7. Problems

1.1. Compute the polynomials A_2 and A_3 explicitly and enumerate lucky tickets.

1.2. Find an expression for the number of $2r$-figure lucky tickets in the number system to the base q.

1.3. Prove the following identities:

a) $\sin^2 s + \cos^2 s = 1$;

b) $(1+s)^\alpha (1+s)^\beta = (1+s)^{\alpha+\beta}$;

c) $\exp(\ln((1-s)^{-1})) = (1-s)^{-1}$;

d) $\ln(1+s) = s - \frac{1}{2}s^2 + \frac{1}{3}s^3 - \dots + \frac{(-1)^{n+1}}{n}s^n + \dots$;

e) $\ln((1-s)^\alpha) = \alpha \ln(1-s)$.

1.4. Suppose a function $B = B(s) = b_1 s + b_2 s^2 + b_3 s^3 + \dots$ is such that $b_1 \neq 0$. Prove that the left inverse function $A(t)$ and the right inverse function $C(t)$ to it coincide. This common *inverse function* is denoted by $B^{-1}(t)$.

1.5. Prove that the power series of the form

$$a_1 s + a_2 s^2 + \dots, \qquad a_1 \neq 0,$$

form a group with respect to the composition.

1.6. Prove that there is no power series $A(s)$ satisfying the equation $sA(s) = 1$.

1.7. Prove that if each of the power series $A(s)$ and $B(s)$ is non-zero, then their product $A(s)B(s)$ also is non-zero.

1.8. Suppose the series $A(s) = a_0 + a_1 s + a_2 s^2 + \dots$, $a_0 \neq 0$, and $B(s) = b_1 s + b_2 s^2 + \dots$, $b_1 \neq 0$, have integer coefficients. Find simple assumptions on their coefficients which guarantee that the series $\frac{1}{A(s)}$, $B^{-1}(s)$ also have integer coefficients.

1.9. Find the generating function for the sequences

a) $1, 2, 3, 4, 5, 6, \dots$;

b) $1 \cdot 2, 2 \cdot 3, 3 \cdot 4, \dots$;

c) $1^2, 2^2, 3^2, 4^2, \dots$

1.10. Prove that for any series $B = B(t)$ with zero free term, $B(0) = 0$, and for arbitrary series $A = A(s)$ one has

$$\left(\int A\right)(B(t)) = \int (A(B(t))B'(t))$$

(the substitution of variable in the integral).

1.11. Prove the Newton–Leibniz identity

$$(A(s)B(s))' = A'(s)B(s) + A(s)B'(s).$$

1.12. Prove the integration by parts formula

$$\int (A(s)B'(s) + A'(s)B(s)) = A(s)B(s) - A(0)B(0).$$

1.13. (Binomial number system) Prove that for a given positive integer k each positive integer n admits a unique representation in the form

$$n = \binom{b_1}{1} + \binom{b_2}{2} + \cdots + \binom{b_k}{k},$$

where $0 \le b_1 < b_2 < \cdots < b_k$. For example, for $k = 2$, we have the following representations:

$$\begin{aligned}
1 &= \binom{0}{1} + \binom{2}{2} \\
2 &= \binom{1}{1} + \binom{2}{2} \\
3 &= \binom{0}{1} + \binom{3}{2} \\
4 &= \binom{1}{1} + \binom{3}{2} \\
5 &= \binom{2}{1} + \binom{3}{2} \\
6 &= \binom{0}{1} + \binom{4}{2}
\end{aligned}$$

and so on. (Recall that, by definition, $\binom{p}{q} = 0$ for integers p, q such that $0 \le p < q$.)

Chapter 2

Generating Functions for Well-known Sequences

2.1. Geometric series

The simplest sequence is the constant sequence $1, 1, 1, \ldots$ The generating function for this sequence has the form

$$G(s) = 1 + s + s^2 + s^3 + \ldots, \tag{2.1}$$

and it is easy to express it in terms of elementary functions. Indeed, multiplying Eq. (2.1) by s we obtain

$$\begin{aligned} sG(s) &= s + s^2 + s^3 + s^4 + \ldots \\ &= G(s) - 1, \end{aligned}$$

whence

$$G(s) = \frac{1}{1-s}. \tag{2.2}$$

After an appropriate modification the same argument works for an arbitrary sequence of the form $a, ar, ar^2, ar^3, \ldots$:

$$\begin{aligned} G_{a,r}(s) &= a + ars + ar^2s^2 + ar^3s^3 + \ldots \\ &= a\left(1 + (rs) + (rs)^2 + (rs)^3 + \ldots\right), \end{aligned}$$

whence

$$rsG_{a,r}(s) = G_{a,r}(s) - a$$

and

$$G_{a,r}(s) = \frac{a}{1 - rs}. \tag{2.3}$$

The above computations are nothing but a well-known derivation of the formula for the sum of a geometric series. Obviously, their result is consistent with the definition of the generating function $(1 - s)^{-1}$.

2.2. The Fibonacci sequence

The famous *Fibonacci sequence* is defined by its first two terms $f_0 = f_1 = 1$ and the relation

$$f_{n+2} = f_{n+1} + f_n. \tag{2.4}$$

This relation allows one to easily produce the first few terms of the Fibonacci sequence,

$$1, 1, 2, 3, 5, 8, 13, 21, 34, \dots;$$

starting with f_2, each element of this sequence is the sum of the two preceding elements. To compute the generating function

$$\mathrm{Fib}(s) = 1 + s + 2s^2 + 3s^3 + 5s^4 + \dots, \tag{2.5}$$

let us multiply both parts of Eq. (2.5) by $s + s^2$. We obtain

$$\begin{aligned}(s + s^2)\,\mathrm{Fib}(s) &= s + s^2 + 2s^3 + 3s^4 + 5s^5 + \dots \\ &\quad + s^2 + s^3 + 2s^4 + 3s^5 + \dots \\ &= s + 2s^2 + 3s^3 + 5s^4 + 8s^5 + \dots,\end{aligned}$$

or

$$(s + s^2)\,\mathrm{Fib}(s) = \mathrm{Fib}(s) - 1,$$

whence

$$\mathrm{Fib}(s) = \frac{1}{1 - s - s^2}. \tag{2.6}$$

The resulting formula may be treated as the composition of two generating functions, namely, $(1 - s)^{-1}$ and $s + s^2$, i.e.,

$$\mathrm{Fib}(s) = 1 + (s + s^2) + (s + s^2)^2 + (s + s^2)^3 + \dots$$

This expansion, however, is inconvenient because the summands on the right contain different powers of s, which makes the explicit formulas for the coefficients complicated. More useful is a representation of the fraction (2.6) as a sum of two elementary fractions:

$$\begin{aligned}\frac{1}{1-s-s^2} &= \frac{1}{\sqrt{5}}\left(\frac{1}{s-s_2}-\frac{1}{s-s_1}\right)\\ &= \frac{1}{\sqrt{5}}\left(\frac{1}{s_1\left(1-\frac{s}{s_1}\right)}-\frac{1}{s_2\left(1-\frac{s}{s_2}\right)}\right),\end{aligned}$$

where $s_1 = (-1+\sqrt{5})/2$ and $s_2 = (-1-\sqrt{5})/2$ are the roots of the quadratic equation $1-s-s^2=0$. The latter decomposition immediately yields

$$\begin{aligned}\mathrm{Fib}(s) = &\frac{1}{\sqrt{5}s_1}\left(1+\frac{s}{s_1}+\frac{s^2}{s_1^2}+\dots\right)\\ &-\frac{1}{\sqrt{5}s_2}\left(1+\frac{s}{s_2}+\frac{s^2}{s_2^2}+\dots\right).\end{aligned}$$

Therefore,

$$\begin{aligned}f_n &= \frac{1}{\sqrt{5}}(s_1^{-1-n}-s_2^{-1-n})\\ &= \frac{(-1)^n}{\sqrt{5}}(s_1^{n+1}-s_2^{n+1})\\ &= \frac{(-1)^n}{\sqrt{5}}\left(\left(\frac{-1+\sqrt{5}}{2}\right)^{n+1}-\left(\frac{-1-\sqrt{5}}{2}\right)^{n+1}\right).\end{aligned} \tag{2.7}$$

Here we made use of the equation $s_1s_2=-1$.

Another way to deduce the generating function for the Fibonacci numbers uses elementary notions of linear algebra. Consider a pair of consecutive Fibonacci numbers f_n, f_{n+1} as the coordinates of a vector in the two-dimensional real vector space $\mathbb{R}^2$:

$$\begin{pmatrix} f_n \\ f_{n+1}\end{pmatrix} \in \mathbb{R}^2.$$

Then Eq. (2.4) may be interpreted as the rule of transformation of the vector $\binom{f_n}{f_{n+1}}$ to the vector $\binom{f_{n+1}}{f_{n+2}}$:

$$\Phi\colon \begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} \mapsto \begin{pmatrix} f_{n+1} \\ f_{n+2} \end{pmatrix} = \begin{pmatrix} f_{n+1} \\ f_n + f_{n+1} \end{pmatrix}.$$

This transformation is linear, and it can be written in the matrix form:

$$\Phi\colon \begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = \Phi \begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix}.$$

The transition from the vector $\binom{f_{n+1}}{f_{n+2}}$ to the vector $\binom{f_{n+2}}{f_{n+3}}$ is achieved by iterating Φ, and so on. Hence, the generating function for the vector Fibonacci sequence becomes

$$\begin{aligned}
\overline{F}(s) &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} s + \begin{pmatrix} 2 \\ 3 \end{pmatrix} s^2 + \begin{pmatrix} 3 \\ 5 \end{pmatrix} s^3 + \dots \\
&= \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} + \Phi \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} s + \Phi^2 \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} s^2 + \Phi^3 \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} s^3 + \dots \\
&= (I + \Phi s + \Phi^2 s^2 + \Phi^3 s^3 + \dots) \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} \\
&= (I - s\Phi)^{-1} \begin{pmatrix} f_0 \\ f_1 \end{pmatrix}.
\end{aligned}$$

Here I denotes the identity matrix, $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and we have applied the derivation of the formula for the geometric series to the vector generating function. The only difference is in the result: the expression $(I - s\Phi)^{-1}$ is treated as the *inverse matrix* to the matrix $I - s\Phi$.

An explicit expression for the Fibonacci numbers can be obtained if we manage to find an explicit form of the matrix Φ^n for arbitrary n. In order to do this, diagonalize the matrix Φ, that is, represent it in the form

$$\Phi = L^{-1} \widetilde{\Phi} L,$$

where $\widetilde{\Phi}$ is a diagonal matrix, and L is a non-degenerate one. Then

$$\Phi = \frac{1}{s_2^{-1} - s_1^{-1}} \begin{pmatrix} 1 & 1 \\ s_1^{-1} & s_2^{-1} \end{pmatrix} \begin{pmatrix} s_1^{-1} & 0 \\ 0 & s_2^{-1} \end{pmatrix} \begin{pmatrix} s_2^{-1} & -1 \\ -s_1^{-1} & 1 \end{pmatrix}.$$

Now, using the relation

$$\Phi^n = L^{-1}\widetilde{\Phi}^n L,$$

and the values s_1, s_2 we obtain Eq. (2.7).

2.3. Recurrence relations and rational generating functions

The Fibonacci sequence is defined by the linear recurrence relation $f_{n+2} = f_{n+1} + f_n$. Using this relation and the two first terms of the sequence we managed to find the generating function explicitly. It happens to be a rational function, that is, a ratio of two polynomials. In fact, our derivation did not use the special form of the recurrence relation. Proceeding in the same way we are able to prove a similar theorem for the generating function of an arbitrary sequence produced by a linear recurrence relation.

Theorem 2.1. *Suppose a sequence a_n is given by a linear recurrence relation of order k with constant coefficients $c_1, \dots, c_k$,*

$$a_{n+k} = c_1 a_{n+k-1} + c_2 a_{n+k-2} + \dots + c_k a_n, \tag{2.8}$$

and let the elements $a_0, a_1, \dots, a_{k-1}$ be given. Then the generating function $A(s) = a_0 + a_1 s + a_2 s^2 + \dots$ is rational, $A(s) = \frac{P(s)}{Q(s)}$, where Q is a polynomial of degree k, and P is a polynomial of degree at most $k-1$.

Proof. The proof of the theorem repeats the argument for the Fibonacci sequence almost word for word. Multiplying the generating function $A(s)$ by $c_1 s + c_2 s^2 + \dots + c_k s^k$ we obtain

$$\begin{aligned}
&(c_1 s + \dots + c_k s^k)A(s) \\
&\quad = c_1 a_0 s + c_1 a_1 s^2 + c_1 a_2 s^3 + \dots + c_1 a_{k-1} s^k + \dots \\
&\qquad\qquad + c_2 a_0 s^2 + c_2 a_1 s^3 + \dots + c_2 a_{k-2} s^k + \dots \\
&\qquad\qquad\qquad + c_3 a_0 s^3 + \dots + c_3 a_{k-3} s^k + \dots \\
&\qquad\qquad\qquad\qquad \dots \\
&\qquad\qquad\qquad\qquad + c_k a_0 s^k + \dots \\
&\quad = P(s) + A(s),
\end{aligned}$$

where the degree of the polynomial P is at most $k-1$. Indeed, the coefficient of s^{n+k} on the right-hand side of the first identity coincides with the right-hand side of Eq. (2.8). Now the theorem is straightforward.

Note that in the course of the proof of Theorem 2.1 we obtained a sharper statement: we have proved that the polynomial Q is

$$Q(s) = 1 - c_1 s - c_2 s^2 - \cdots - c_k s^k.$$

The derivation of the vector generating function for the Fibonacci sequence also admits an immediate generalization to the case of an arbitrary recursive sequence. In the general case, the two-dimensional vector must be replaced with the k-dimensional vector

$$\begin{pmatrix} a_n \\ a_{n+1} \\ \vdots \\ a_{n+k-1} \end{pmatrix} \in \mathbb{R}^k,$$

and the transformation matrix $\mathcal{A}$ corresponding to the recurrence relation acquires the form

$$(2.9) \qquad \begin{pmatrix} a_{n+1} \\ a_{n+2} \\ \vdots \\ a_{n+k} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \multicolumn{6}{c}{\dotfill} \\ 0 & 0 & 0 & \dots & 0 & 1 \\ c_k & c_{k-1} & c_{k-2} & \dots & c_2 & c_1 \end{pmatrix} \begin{pmatrix} a_n \\ a_{n+1} \\ \vdots \\ a_{n+k-1} \end{pmatrix}.$$

As the result, we obtain the vector generating function

$$\overline{A}(s) = (I - s\mathcal{A})^{-1} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{pmatrix}.$$

Generally speaking, the matrix $\mathcal{A}$ is non-diagonalizable. It is diagonalizable if and only if its eigenvalues (that is, the roots of the polynomial Q) are pairwise distinct. However, in the general case it admits a Jordan normal form, whose powers also are easily computable.

It happens that rational generating functions are exactly the generating functions for recursive sequences. To be more precise, the following statement takes place.

Theorem 2.2. *If the generating function* $A(s) = a_0 + a_1 s + a_2 s^2 + \dots$ *is rational,* $A(s) = \frac{P(s)}{Q(s)}$*, where the polynomials* P *and* Q *are coprime, then, starting from some number* n*, the sequence* $a_0, a_1, a_2, \dots$ *is given by a linear recurrence relation*

$$a_{n+k} = c_1 a_{n+k-1} + c_2 a_{n+k-2} + \dots + c_k a_n,$$

where k *is the degree of* Q*, and* $c_1, c_2, \dots, c_k$ *are some constants.*

We leave the proof as an exercise to the reader.

2.4. The Hadamard product of generating functions

One of the most attractive features of rational generating functions is their closedness with respect to the Hadamard product.

Definition 2.3. The *Hadamard product* of two generating functions

$$A(s) = a_0 + a_1 s + a_2 s^2 + \dots$$

and

$$B(s) = b_0 + b_1 s + b_2 s^2 + \dots$$

is the generating function

$$A \circ B(s) = a_0 b_0 + a_1 b_1 s + a_2 b_2 s^2 + \dots$$

Hence, the Hadamard product of two sequences is the sequence whose elements are the products of corresponding elements of the sequences. We have already met a situation, where the generating function for the Hadamard product was necessary: in the lucky tickets problem we have to compute the sum of squares of the coefficients of the generating polynomial A_3. This situation is reproduced each time when we need to enumerate pairs of objects of the same order n: if the number of objects of the first kind is a_n, while that of the second kind is b_n, then the number of pairs of objects, one of the first and one of the second kind, is $a_n b_n$.

Theorem 2.4. *The Hadamard product of two rational generating functions is rational.*

To prove the theorem we will need a new characterization of the rational generating functions.

Lemma 2.5. *The generating function for a sequence* $a_0, a_1, a_2, \dots$ *is rational if and only if there are numbers* $q_1, \dots, q_l$ *and polynomials* $p_1(n), \dots, p_l(n)$ *such that starting from some number* n *we have*

$$a_n = p_1(n)q_1^n + \dots + p_l(n)q_l^n. \tag{2.10}$$

The expression on the right-hand side of Eq. (2.10) is called a *quasipolynomial* in the variable n.

Proof. Note first of all that the generating function $(1-qs)^{-k}$ has the form

$$\begin{aligned}(1-qs)^{-k} &= 1 - \binom{-k}{1}qs + \binom{-k}{2}q^2s^2 - \binom{-k}{3}q^3s^3 + \dots \\ &= 1 + \binom{k}{1}qs + \binom{k+1}{2}q^2s^2 + \binom{k+2}{3}q^3s^3 + \dots \\ &= 1 + \binom{k}{k-1}qs + \binom{k+1}{k-1}q^2s^2 + \binom{k+2}{k-1}q^3s^3 + \dots\end{aligned}$$

The coefficient of s^n in this generating function has the form

$$\binom{k+n-1}{k-1}q^n = \frac{(n+1)(n+2)\dots(n+k-1)}{(k-1)!}q^n = P_{k-1}(n)q^n, \tag{2.11}$$

where $P_{k-1}(n)$ is a polynomial in n of degree $k-1$. Any rational function of s admits a representation as a linear combination of a polynomial and elementary fractions of the form $(1-q_is)^{-k_i}$; therefore, the coefficients of a rational function are quasipolynomials.

Conversely, suppose that the coefficients of a generating function are quasipolynomials starting from some number. Let us show that the generating function with quasipolynomial coefficients $a_n = p(n)q^n$ is rational. Suppose the degree of the polynomial p is $k-1$. The polynomials $P_0, P_1, \dots, P_{k-1}$ defined by Eq. (2.11) form a basis in the vector space of polynomials of degree at most $k-1$. Indeed, any sequence of polynomials of degrees $0, 1, \dots, k-1$ forms a basis in this vector space. Therefore, p is a linear combination of the polynomials P_i and the corresponding generating function is simply a linear combination of the rational functions $(1-qs)^{-j}$, $j=0,1,\ldots,k-1$. For an arbitrary quasipolynomial we arrive at a linear combination of elementary fractions with different q_i. This completes the proof of the lemma.

Proof of Theorem 2.4. The theorem follows from the fact that the product of two quasipolynomials is also a quasipolynomial. This is an immediate corollary of Eq. (2.10).

2.5. Catalan numbers

In arithmetic expressions, the order of computations is determined by brackets, e.g.,

$$(3-1)\cdot(4+(15-9)\cdot(2+6)).$$

After erasing all elements of an arithmetic expression except for the brackets we obtain what is called a *regular bracket structure*:

$$()(()()).$$

Here are all regular bracket structures with one, two, and three pairs of brackets:

$$()$$

$$()() \qquad (())$$

$$()()() \qquad ()(()) \qquad (())() \qquad (()()) \qquad ((()))$$

Definition 2.6. The *Catalan number* c_n is the number of regular bracket structures with n pairs of brackets.

It is convenient to set $c_0 = 1$. Then the sequence of Catalan numbers starts as follows:

$$1, 1, 2, 5, 14, 42, 132, \ldots$$

In order to deduce the generating function for the Catalan numbers let us first deduce a recurrence relation for these numbers.

Each regular bracket structure satisfies the following two conditions:

1) the number of left and right brackets in a regular bracket structure is the same;

2) the number of left brackets in any starting segment of a regular bracket structure is not less than the number of right brackets in the same segment.

Conversely, each (finite) sequence of left and right brackets possessing properties 1) and 2) is a regular bracket structure.

All brackets in a regular bracket structure are split into pairs: a right bracket is associated to each left bracket. The right bracket associated to a given left one is defined by the following rule: this is the first right bracket to the right of the given left bracket such that the sequence of brackets between the two is a regular bracket structure.

Consider a regular bracket structure consisting of $n+1$ pairs of brackets and the pair of brackets in it that contains the leftmost left bracket. Then the sequence of brackets inside the chosen pair forms a regular bracket structure, and the sequence of pairs outside this pair also forms a regular bracket structure: $(\dots)\dots$, where dots denote regular bracket structures. If the inside regular bracket structure consists of k pairs of brackets, then there are $n-k$ pairs of brackets in the outside bracket structure. Conversely, for each pair of regular bracket structures containing respectively k and $n-k$ pairs of brackets we can construct a new regular bracket structure, consisting of $n+1$ pairs of brackets, by bracketing the first structure and concatenating the result with the second one.

This procedure produces a recurrence relation for the Catalan numbers. This time the relation is non-linear:

$$c_{n+1} = c_0c_n + c_1c_{n-1} + \cdots + c_nc_0. \tag{2.12}$$

For example, for $n = 4$ we have

$$\begin{aligned} c_5 &= c_0c_4 + c_1c_3 + c_2c_2 + c_3c_1 + c_4c_0 \\ &= 1 \cdot 14 + 2 \cdot 5 + 2 \cdot 2 + 5 \cdot 1 + 14 \cdot 1 \\ &= 42. \end{aligned}$$

Consider the generating function for the Catalan numbers:

$$\begin{aligned}\mathrm{Cat}(s) &= c_0 + c_1 s + c_2 s^2 + \dots \\ &= 1 + s + 2s^2 + 5s^3 + \dots\end{aligned}$$

Taking its square and multiplying the result by s we obtain

$$\begin{aligned}s\,\mathrm{Cat}^2(s) &= c_0^2 s + (c_0c_1 + c_1c_0)s^2 + (c_0c_2 + c_1c_1 + c_2c_0)s^2 + \dots \\ &= s + 2s^2 + 5s^3 + 14s^4 + \dots \\ &= \mathrm{Cat}(s) - 1,\end{aligned}$$

which leads to the following quadratic equation:

$$s\,\mathrm{Cat}^2(s) - \mathrm{Cat}(s) + 1 = 0, \tag{2.13}$$

whence

$$\mathrm{Cat}(s) = \frac{1 - \sqrt{1-4s}}{2s}. \tag{2.14}$$

(We choose the minus sign before the root because the expression with the plus sign, $(1 + \sqrt{1-4s})/2s = 1/s + \dots$, contains a negative power of s and hence cannot coincide with $\mathrm{Cat}(s)$.)

The generating function (2.14) allows one to find an explicit formula for the Catalan numbers. By the Newton binomial formula,

$$c_n = \frac{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{3}{2} \cdot \dots \cdot \frac{2n-1}{2} \cdot 4^{n+1}}{2(n+1)!},$$

and multiplying both the numerator and the denominator by $n!$ and dividing by 2^{n+1} we obtain

$$c_n = \frac{(2n)!}{n!(n+1)!} = \frac{1}{n+1}\binom{2n}{n}. \tag{2.15}$$

The last formula also yields a simpler (but containing variable coefficients) recurrence relation for the Catalan numbers:

$$c_{n+1} = \frac{4n+2}{n+2} c_n. \tag{2.16}$$

The Catalan numbers enumerate objects of various nature. They are discussed in numerous papers and books. Dozens of their definitions are known. We describe here only two more of their interpretations.

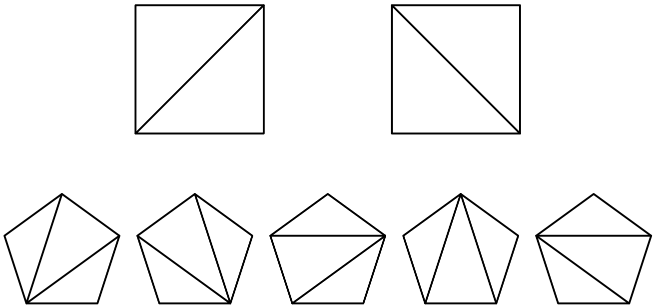


Figure 1. Diagonal triangulations of the square and the pentagon

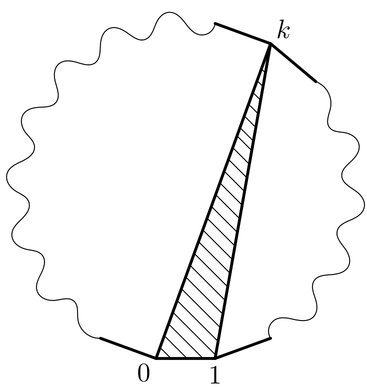


Figure 2. The triangle adjacent to the side 01

Consider a convex $(n+2)$-gon whose vertices are numbered counterclockwise from 0 to $n+1$. A *diagonal triangulation* is a partition of the polygon into triangles by non-intersecting diagonals. Each diagonal triangulation contains $n-1$ diagonals. All diagonal triangulations of the square and the pentagon are shown in Fig. 1.

Let t_n be the number of triangulations of the $(n+2)$-gon, $n \geq 1$; we set $t_0 = 1$. Consider an arbitrary triangulation and the triangle in it adjacent to the side 01 (see Fig. 2). Let k be the number of the third vertex of this triangle. The chosen triangle splits the $(n+2)$-gon into a k-gon and a $(n-k+3)$-gon, each triangulated by diagonals. Let us number the vertices of both polygons counterclockwise starting with 0 (see Fig. 3). As a result, we obtain a pair of diagonal triangulations of a k-gon and a $(n-k+3)$-gon.

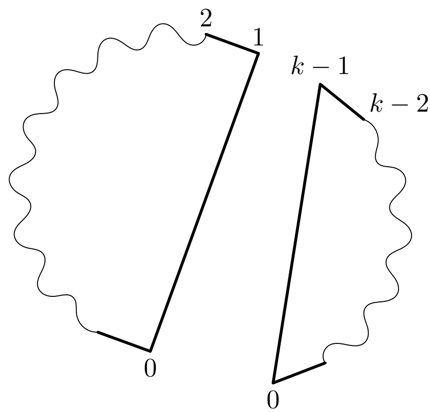


Figure 3. Renumbering the vertices of the parts

Conversely, each pair of triangulations of a k-gon and a $(n-k+3)$-gon uniquely determines a triangulation of the initial polygon. Therefore,

$$t_{n+1} = t_0 t_n + t_1 t_{n-1} + \cdots + t_n t_0,$$

and since $t_0 = 1$, the sequence t_n coincides with the Catalan sequence.

The procedure associating to a triangulation of a $(n+2)$-gon a pair of triangulations of a k-gon and a $(n-k+3)$-gon described above allows one to establish a one-to-one correspondence between triangulations of a $(n+2)$-gon and regular bracket structures of n pairs of brackets. Indeed, suppose that such correspondence is already established for all smaller values of n. We have assigned to each triangulation of a $(n+2)$-gon a pair of triangulations of polygons with fewer vertices. By the induction hypothesis, a pair of regular bracket structures is associated to this pair of triangulations. Bracketing the first bracket structure and concatenating the result with the second one we obtain a new regular bracket structure which we assign to the initial triangulation of the entire $(n+2)$-gon.

Another important realization of the Catalan numbers is related to the Dyck paths in the plane. Consider the integer square lattice in the positive quadrant in the plane. A *Dyck path* is a continuous broken line consisting of vectors $(1,1)$ and $(1,-1)$, starting at the origin and ending at the x-axis (see Fig. 4).

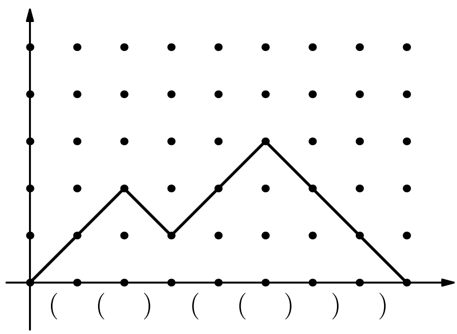


Figure 4. A Dyck path and the corresponding regular bracket structure

It is clear how to establish a correspondence between the Dyck paths and the regular bracket structures: to the vector $(1,1)$ we associate the left bracket, and to the vector $(1,-1)$ the right one (Fig. 4). Then the assumption that the path belongs to the upper half-plane and ends at the x-axis just coincides with the regularity requirement for the bracket structure. Therefore:

The number of Dyck paths consisting of $2n$ *steps coincides with the* n*th Catalan number* c_n.

2.6. Problems

2.1. Prove that if a sequence a_n admits a representation of the form (2.10) and all q_i are distinct, then such a representation is unique up to a permutation of the summands.

2.2. Using the previous problem, show that the generating function $\ln((1-s)^{-1}) = s + s^2/2 + s^3/3 + \dots$ is not rational.

2.3. Are the generating functions for the following sequences rational?

a) $1, -2, 3, -4, 5, \dots;$

b) $2, -6, 12, \dots, (-1)^k(k+1)(k+2), \dots;$

c) $1, -4, 9, -16, \dots, (-1)^k(k+1)^2, \dots$;

d) $1, \frac{1}{4}, \frac{1}{9}, \dots, \frac{1}{k^2}, \dots$;

e) f_n^2, where f_n are the Fibonacci numbers?

Find the generating functions in those cases, where they are rational.

2.4. Let $A(s) = a_0 + a_1 s + a_2 s^2 + \dots$ be the generating function for a sequence $a_0, a_1, a_2, \dots$. Express in terms of A the generating functions for the following sequences:

a) $a_0 + a_1, a_1 + a_2, a_2 + a_3, \dots$;

b) $a_0, a_0 + a_1, a_0 + a_1 + a_2, \dots$;

c) $a_0, a_1 b, a_2 b^2, a_3 b^3, \dots,\quad b$ is a constant;

d) $a_0, 0, a_2, 0, a_4, 0, a_6, 0, a_8, 0, \dots$

2.5. Using the generating function for the Fibonacci numbers, prove the following identities:

a) $f_0 + f_1 + \dots + f_n = f_{n+2} - 1$;

b) $f_0 + f_2 + \dots + f_{2n} = f_{2n+1}$;

c) $f_1 + f_3 + \dots + f_{2n-1} = f_{2n} - 1$;

d) $f_0^2 + f_1^2 + \dots + f_n^2 = f_n f_{n+1}$.

2.6. Prove that in the Jordan normal form the matrix of Eq. (2.9) has exactly one Jordan block, of dimension equal to the multiplicity of the eigenvalue, for each eigenvalue.

2.7. Find the generating functions and explicit formulas for the sequences given by the following recurrence relations:

a) $a_{n+2} = 4a_{n+1} - 4a_n,\ a_0 = a_1 = 1$;

b) $a_{n+3} = -3a_{n+2} - 3a_{n+1} - a_n,\ a_0 = 1, a_1 = a_2 = 0$;

c) $a_{n+3} = \frac{3}{2}a_{n+2} - \frac{1}{2}a_n,\ a_0 = 0, a_1 = 1, a_2 = 2$.

2.8. Find the Hadamard products for the following functions of s:

$$(1-qs)^{-1} \circ (1-rs)^{-1},\ (1-qs)^{-1} \circ (1-qs)^{-1},\ (1-qs)^{-2} \circ (1-rs)^{-1},$$

$$(1-qs)^{-2} \circ (1-rs)^{-2},\ (1-qs)^{-3} \circ (1-qs)^{-1}.$$

2.9. The *Motzkin paths* are defined in the same way as Dyck paths, but they can also include horizontal vectors $(1, 0)$ (see Fig. 5). The

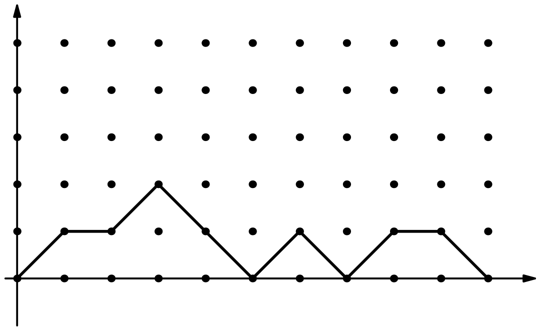


Figure 5. A Motzkin path

number of Motzkin paths consisting of n vectors is called the nth *Motzkin number* and is denoted by m_n. Here is the beginning of the Motzkin sequence: $m_0 = 1, m_1 = 1, m_2 = 2, m_3 = 3$. Compute some further terms of this sequence. Find a recurrence relation and the generating function for this sequence.

2.10. Invent algorithms producing successively a) regular bracket structures; b) Motzkin paths. Try to make the algorithms as efficient as possible.

2.11. Consider the set of paths on the line consisting of steps of length 1 to the left or to the right. Find the generating functions for the numbers of such paths issuing from 0, consisting of n steps and a) returning back to 0; b) returning to 0 and not entering the negative half-line; c) ending at some given point $N > 0$; d) ending at some given point $N > 0$ and not entering the negative half-line.

2.12. Consider the set of paths on the plane consisting of vectors $(1, 0)$, $(-1, 0)$, $(0, 1)$. Find the generating function for the numbers of such paths of length n issuing from the origin and non-selfintersecting (this means that the vectors $(1, 0)$ and $(-1, 0)$ cannot follow each other immediately in a path).

2.13. Two players play the following game. The first picks an integer number between 1 and 144 (inclusive), while the second tries to guess the number by asking questions to which the first player answers (honestly) “yes” or “no”. Receiving the answer “yes” the second player pays 1 rouble, while the price of the answer “no” is 2 roubles. What is the strategy for the second player minimizing the loss in the worst situation possible? And what if 144 is replaced by another number?

2.14. (The Hipparchus Problem) The following citation is taken from Plutarch, *Moralia*, vol. 9, Cambridge MA, Harward University Press, 1961, §VIII.9, p. 732:

“Chrysippus says that the number of compound propositions that can be made from only ten simple propositions exceeds a million. (Hipparchus, to be sure, refuted this by showing that on the affirmative side there are 103,049 compound statements, and on the negative side 310,952.)”

Verify the Hipparchus statement admitting that

- a “compound proposition on the affirmative side” is constructed from n simple propositions by bracketing them in all possible ways (with at most $n-2$ pairs of brackets). For example, there are three compound propositions for $n = 3$:

$$a, b, c\,; \quad (a, b), c\,; \quad a, (b, c),$$

  and 11 compound propositions for $n = 4$:

$$a, b, c, d\,; \quad a, (b, c, d)\,; \quad (a, b, c), d\,; \quad a, b, (c, d);$$
$$a, (b, c), d\,; \quad (a, b), c, d\,; \quad a, (b, (c, d))\,; \quad a, ((b, c), d)\,;$$
$$(a, (b, c)), d\,; \quad ((a, b), c), d\,; \quad (a, b), (c, d)\,;$$

- a “compound proposition on the negative side” is constructed from n simple propositions by bracketing them, preceded by the negation, in all possible ways (with at most $n-1$ pairs of brackets). For example, there are 7 compound propositions on the negative side constructed from three simple propositions:

$$\text{no}\, a, b, c\,; \quad (\text{no}\, a, b), c\,; \quad \text{no}\, a, (b, c)\,; \quad \text{no}(a, b), c\,;$$
$$\text{no}(a, b, c)\,; \quad \text{no}((a, b), c)\,; \quad \text{no}(a, (b, c))\,.$$

(In the second case, the answer differs slightly from the one given by Hypparchus.)

Deduce the generating functions for the corresponding sequences.

2.15. (Fibonacci computational system) Prove that each positive integer admits a unique representation in the form $a_1 f_1 + a_2 f_2 + \dots$, where f_n are the Fibonacci numbers, each of the numbers a_i is either 0 or 1, the number of ones in the representation is finite, and no two subsequent elements of the sequence a_i are equal to 1 simultaneously. For example, the first few representations are $1 = f_1$, $2 = f_2$, $3 = f_3$, $4 = f_3 + f_1$, $5 = f_4$, $6 = f_4 + f_1$, $7 = f_4 + f_2$. (Pay attention to the fact that the number $f_0 = 1$ is not used in this computational system, so that the Fibonacci sequence starts with $1, 2, 3, 5, 8, \dots$.) Invent algorithms for converting numbers from the Fibonacci system to the decimal positional number system and back, and algorithms for adding and multiplying numbers written in the Fibonacci sequence.

2.16. Let

$$A_1(s) = \frac{P_1(s)}{Q_1(s)}, \quad A_2(s) = \frac{P_2(s)}{Q_2(s)}$$

be two rational generating functions given by irreducible fractions, and let

$$A_1 \circ A_2(s) = \frac{P(s)}{Q(s)}$$

be their Hadamard product represented as an irreducible fraction. What can be said about the polynomial Q in the denominator, provided the polynomials Q_1 and Q_2 are known?

Chapter 3

Unambiguous Formal Grammars. The Lagrange Theorem

3.1. The Dyck Language

We know already that regular bracket structures are enumerated by the Catalan numbers. Let us write out all regular bracket structures with up to 4 pairs of brackets:

( )	(( ))	((( )))	(((( ))))	((( ))) ( )	( ) (( ) ( ))
$a\,b$	$a\,a\,b\,b$	$a\,a\,a\,b\,b\,b$	$a\,a\,a\,a\,b\,b\,b\,b$	$a\,a\,a\,b\,b\,b\,a\,b$	$a\,b\,a\,a\,b\,a\,b\,b$
	( ) ( )	(( ) ( ))	((( ) ( )))	(( ) ( )) ( )	( ) (( )) ( )
	$a\,b\,a\,b$	$a\,a\,b\,a\,b\,b$	$a\,a\,a\,b\,a\,b\,b\,b$	$a\,a\,b\,a\,b\,b\,a\,b$	$a\,b\,a\,a\,b\,b\,a\,b$
		(( )) ( )	((( )) ( ))	(( )) (( ))	( ) ( ) (( ))
		$a\,a\,b\,b\,a\,b$	$a\,a\,a\,b\,b\,a\,b\,b$	$a\,a\,b\,b\,a\,a\,b\,b$	$a\,b\,a\,b\,a\,a\,b\,b$
		( ) (( ))	(( ) (( )))	(( )) ( ) ( )	( ) ( ) ( ) ( )
		$a\,b\,a\,a\,b\,b$	$a\,a\,b\,a\,a\,b\,b\,b$	$a\,a\,b\,b\,a\,b\,a\,b$	$a\,b\,a\,b\,a\,b\,a\,b$
		( ) ( ) ( )	(( ) ( ) ( ))	( ) ((( )))	
		$a\,b\,a\,b\,a\,b$	$a\,a\,b\,a\,b\,a\,b\,b$	$a\,b\,a\,a\,a\,b\,b\,b$	

Denoting the left bracket by a and the right bracket by b we can rewrite regular bracket structures as "words" over the alphabet $\{a, b\}$. In the above table, the corresponding word is written under each bracket structure.

Not all words over the alphabet $\{a, b\}$ can be obtained in this way. For example, no regular bracket structure corresponds to any of the words a, b, aa, ba.

Definition 3.1. Let $A = \{a_1, a_2, \dots, a_k\}$ be an arbitrary finite set of distinct letters. A *word* over the alphabet A is an arbitrary finite sequence $\alpha_1\alpha_2\dots\alpha_m$, where $\alpha_i \in A,\ i = 1, \dots, m$. The number m is called the *length* of the word. A *language* over the alphabet A is an arbitrary (either finite or infinite) set of words over A.

The length of the *empty word* λ is 0; the empty word may either belong or not belong to a given language.

Example 3.2. Let $\mathcal{F}$ denote the language over the alphabet $\{a, b\}$ consisting of words not containing two consecutive occurrences of the letter b: $\lambda, a, b, ab, ba, aaa, aab, aba, baa, bab, aaaa, \dots$

The set of regular bracket structures, together with the empty word, also form a language over the alphabet $\{a, b\}$. This language is called the *Dyck language*. Of course, we are able to speak about the same language over the alphabet $\{(,)\}$; however, for us the symbols a, b resemble letters more than the brackets.

Definition 3.3. The *generating function* of a language L is the generating function

$$L(s) = l_0 + l_1 s + l_2 s^2 + \dots,$$

where l_k is the number of words of length k in L.

3.2. Productions in the Dyck language

Writing out all regular bracket structures is a hard job. The process requires some order which allows one to include all regular bracket structures, and not to mention any of them twice. A useful tool to achieve this goal is the following set of *production rules in the Dyck language*:

$$\begin{array}{lccl} 1) & r & \longrightarrow & \lambda; \\ 2) & r & \longrightarrow & arbr. \end{array} \tag{3.1}$$

Here r is just an auxiliary letter. Instead of it, any letter not belonging to the alphabet $\{a, b\}$ would do.

The arrow in each of the production rules (3.1) is a substitute for the following phrase:

an occurrence of the letter r in a word can be replaced with the word on the right of the arrow.

To show how the production rules work, let us derive a given regular bracket structure using them.

Suppose we want to derive the word *aabaabbb*. Here is the derivation:

$$\underline{r} \xrightarrow{2)} arb\underline{r} \xrightarrow{1)} a\underline{r}b \xrightarrow{2)} aa\underline{r}brb \xrightarrow{1)} aab\underline{r}b \xrightarrow{2)} aaba\underline{r}brb \xrightarrow{1)}$$
$$\xrightarrow{1)} aaba\underline{r}bb \xrightarrow{2)} aabaa\underline{r}b\underline{r}bb \xrightarrow{1),1)} aabaabbb.$$

Above each arrow, the number of the rule applied is displayed. The occurrence of the letter r, to which the rule has been applied, is underlined.

The production rules in the Dyck language can be interpreted as follows:

Each word in the Dyck language is

1) *either the empty word,*

2) *or the concatenation of a bracketed word of the Dyck language and a word of the Dyck language.*

Clearly, each word of the Dyck language admits a unique representation of this form.

Let us compute, by means of the production rules, the generating function for the Dyck language. In order to do this, consider the "non-commutative generating series", enumerating all words in the language. This series is nothing but the formal sum of all words of the language, arranged in increasing order of their lengths:

$$\mathcal{D}(a, b) = \lambda + ab + aabb + abab + aaabbb + aababb + \dots \tag{3.2}$$

Theorem 3.4. *The series* (3.2) *satisfies the equation*

$$\mathcal{D}(a, b) = \lambda + a\mathcal{D}(a, b)b\mathcal{D}(a, b). \tag{3.3}$$

Proof. Indeed, the left-hand side of Eq. (3.3) is simply the sum of all words in the Dyck language. The identity means that the following statement is true:

Each word in the Dyck language is

1) *either the empty word,*

2) *or the concatenation of a bracketed word of the Dyck language and a word of the Dyck language.*

Such a representation is unique, which completes the proof of the theorem.

To obtain an ordinary generating series instead of the non-commutative one, let us make the substitution $a = s,\ b = s,\ \lambda = s^0 = 1$. Then Eq. (3.3) becomes

$$\mathcal{D}(s,s) = 1 + s^2\mathcal{D}(s,s).$$

Denoting $\mathcal{D}(s,s)$ by $d(s)$ we arrive at the equation

$$d(s) = 1 + s^2 d^2(s). \tag{3.4}$$

Of course, the solution

$$d(s) = \frac{1 - \sqrt{1-4s^2}}{2s^2}$$

of this equation coincides (up to squaring of the formal variable) with the generating function for the Catalan numbers (2.14). The appearance of s^2 instead of s is due to the fact that the length of a word in the Dyck language consisting of n pairs of brackets is $2n$, while previously we enumerated regular bracket structures with respect to the number of *pairs* of brackets.

3.3. Unambiguous formal grammars

Let us formalize and generalize the argument of the previous section.

Definition 3.5. A word $w = \beta_1 \dots \beta_m$ in a language L is said to be *indecomposable* in this language if none of its non-empty subwords $\beta_i\beta_{i+1}\dots\beta_{i+l}$, $1 \le i, i+l \le m,\ l \ge 0$, different from w belong to L.

In particular, the empty word is indecomposable in any language. Suppose a language L possesses the following properties:

(3.5)
1) the empty word belongs to L;
2) no beginning of an indecomposable word coincides with an end of the same or other indecomposable word;
3) if we insert a word belonging to L between any two neighboring letters of a word in L, then we obtain a word in L;
4) if we erase a subword belonging to L from a word in L, then we obtain a word in L.

Denote by $n(t) = n_0 + n_1 t + n_2 t^2 + \dots$ the generating function for the number of indecomposable words in L and by $l(s) = l_0 + l_1 s + l_2 s^2 + \dots$ the generating function for L.

Theorem 3.6. *The generating function for a language L possessing properties* (3.5) *and the generating function for the sublanguage of indecomposable words in L are related by the Lagrange equation*

$$l(s) = n(sl(s)). \tag{3.6}$$

Proof. Associate to each indecomposable word $\alpha_{i_1} \dots \alpha_{i_m}$ in L the production rule

$$r \longrightarrow \alpha_{i_1} r \alpha_{i_2} r \dots \alpha_{i_m} r.$$

It is clear that each word in L can be derived using these rules. Indeed, let $\beta_1 \dots \beta_k$ be an arbitrary word in L. If it is indecomposable, then it is represented by the right-hand side of the production rule

$$r \longrightarrow \beta_1 r \beta_2 r \dots \beta_k r,$$

where each occurrence of r must be replaced with the empty word. By the definition of an indecomposable word, such representation is unique. Otherwise, we may delete an indecomposable subword in $\beta_1 \dots \beta_k$, not containing the leftmost letter β_1, and proceed by induction over the length of the word.

Now suppose there are decomposable words admitting several representations. Let w be the shortest such word. It contains an indecomposable subword. Take the rightmost indecomposable subword in w. This is possible since indecomposable subwords in a word cannot overlap. After deleting this indecomposable subword we obtain a new word w', which has the same representations as w as the right-hand sides of the production rules. Hence, w' is a shorter word also admitting several representations. The contradiction thus obtained proves the uniqueness of representation.

Hence, the non-commutative generating function for the language L satisfies the equation

$$\mathcal{L}(a_1,\dots,a_m)=\lambda+\alpha_{11}\mathcal{L}(a_1,\dots,a_m)\alpha_{12}\mathcal{L}(a_1,\dots,a_m)\cdots+\dots$$

Making the substitution $\lambda=1,\ a_i=t$ and taking into account that $l(t)=\mathcal{L}(t,t,\dots,t)$ we arrive at the conclusion of the theorem.

Example 3.7. For the Dyck language, $n(t)=1+t^2$. The only indecomposable words are λ and ab. This gives us immediately Eq. (3.4) for the generating function for the Dyck language.

One additional symbol is not always enough to construct a grammar. Let us give a formal definition of a grammar.

Definition 3.8. Let $R=\{r_1,\dots,r_l\}$ be a finite set of symbols having empty intersection with A. A *production rule* is a string of the form

$$r_i \longrightarrow w,$$

where $r_i\in R$ and w is a word over the alphabet $A\sqcup R$. A (finite or infinite) set Γ of production rules

$$\begin{array}{ccc} r_1 & \longrightarrow & w_{11}, \\ r_1 & \longrightarrow & w_{12}, \\ & \dots & \\ r_l & \longrightarrow & w_{l1}, \\ r_l & \longrightarrow & w_{l2}, \\ & \dots & \end{array}$$

is called a (*context free*) *grammar* over the alphabet A. A word over A is *generated* by the symbol r_i if it can be obtained from r_i by a sequence of productions in Γ. A language L_i is *generated* by the symbol

r_i if all the words in L_i and only these words are generated by r_i. A grammar Γ is *unambiguous* if each word generated by r_i has a unique representation as the right-hand side of a production rule $r_i \longrightarrow w_{ik}$.

Enumeration problems are related first of all with unambiguous formal grammars. We have constructed already such a grammar for the Dyck language. Let us give one more example. An example of an ambiguous formal grammar is given in Problem 3.3 below.

Example 3.9. Consider the language $\mathcal{F}$ from Example 3.2. Here is a possible grammar for this language:

$$\begin{array}{rcl} r_1 & \longrightarrow & \lambda, \\ r_1 & \longrightarrow & b, \\ r_1 & \longrightarrow & r_2 b, \\ r_1 & \longrightarrow & r_2, \\ r_2 & \longrightarrow & r_1 a. \end{array}$$

The language $\mathcal{F}$ is generated by the symbol r_1. The symbol r_2 generates the sublanguage of $\mathcal{F}$ consisting of words ending with the letter a.

This grammar can be pronounced as follows:

1) each word of $\mathcal{F}$ is either the empty word, or the word b, or a word of $\mathcal{F}$ ending with a concatenated with b, or a word of $\mathcal{F}$ ending with a;

2) each word of $\mathcal{F}$ ending with a is a word of $\mathcal{F}$ concatenated with a.

Theorem 3.10. *Let Γ be an unambiguous grammar. Denote by $r_i(s)$ the generating functions for the languages L_i generated from the symbols r_i. Then these functions satisfy a system of equations*

$$r_i(s) = \sum_j s^{\nu_{ij}} \prod_k r_k^{\eta_{kj}}(s).$$

In particular, if there are finitely many production rules, then the generating functions r_i satisfy a system of polynomial equations and hence are *algebraic functions*.

Proof. We proceed in the same way as in the situation with one generating symbol. Namely, introduce the non-commutative generating series for each of the languages L_i. Since the grammar is unambiguous, we obtain a system of equations for unknown non-commutative series. The substitution $\lambda = s^0 = 1,\ a_i = s$ for $i = 1, \ldots, m$, reduces the system to one for non-commutative series. This completes the proof of the theorem.

3.4. The Lagrange equation and the Lagrange theorem

Let us look more attentively at Eq. (3.6). This is a functional equation relating the generating function for a language and the sublanguage of indecomposable words in it. We would like to know how to solve this equation provided that one of the functions is known. It happens that this is always possible.

First of all we rewrite this equation in the classical form by multiplying both parts by s introducing the notation $sl(s) = \tilde{l}(s)$. Then Eq. (3.6) becomes

$$\tilde{l}(s) = sn(\tilde{l}(s)). \tag{3.7}$$

The last equation is called the *Lagrange equation* and the following theorem holds for it.

Theorem 3.11 (Lagrange)**.** *Suppose one of the generating functions $\tilde{l}(s)$ $(\tilde{l}_0 = 0,\ \tilde{l}_1 \neq 0)$ or $n(t)$ $(n_0 \neq 0)$ in Eq.* (3.7) *is given. Then the second generating function can be uniquely reconstructed from it.*

Proof. One can rewrite Eq. (3.7) in the following form:

$$\begin{aligned} \tilde{l}_1 s + \tilde{l}_2 s^2 + \cdots = n_0 s &+ n_1 s\Big(\tilde{l}_1 s + \tilde{l}_2 s^2 + \tilde{l}_3 s^3 + + \ldots\Big) \\ &+ n_2 s\Big(\qquad \tilde{l}_1^2 s^2 + 2\tilde{l}_1\tilde{l}_2 s^3 + \ldots\Big) \\ &+ n_3 s\Big(\qquad\qquad \tilde{l}_1^3 s^3 \quad + \ldots\Big) \\ &+ \qquad\qquad\qquad \ldots \end{aligned}$$

Let us prove first that if the function $\tilde{l}(s)$ is known, then one can reconstruct the function $n(t)$. We will proceed by induction equating successively the coefficients of the same powers of s on both sides.

The coefficient n_0 is determined from the equation

$$n_0 = \tilde{l}_1.$$

Now suppose the coefficients $n_0, n_1, \dots, n_{k-1}$ are already known. Then the coefficient n_k is determined from the equation obtained by equating the coefficients of s^{k+1}:

$$n_k \tilde{l}_1^k + n_{k-1}\lambda_{k-1} + \dots + n_1\lambda_1 = \tilde{l}_k. \tag{3.8}$$

Here λ_i, $i = 2, \dots, k-1$, denote the coefficients of s^k in the generating functions $\tilde{l}^i(s)$. Equation (3.8) is a linear equation with respect to n_k. The coefficient of n_k in it is $\tilde{l}_1^k$, which is non-zero by the assumptions of the theorem. Therefore, Eq. (3.8) determines n_k uniquely.

On the other hand, if the function $n(t)$ is given, then we must set $\tilde{l}_1 = n_0$. Then the coefficients $\tilde{l}_k$ are uniquely determined by (3.8), since each of the coefficients λ_i is a polynomial in $\tilde{l}_1, \dots, \tilde{l}_{k-1}$. The proof of the theorem is completed.

Remark 3.12. If the coefficients of the function n are non-negative integers, then the same is true for the coefficients of the function $\tilde{l}$. If the coefficients of the function $\tilde{l}$ are non-negative integers and $\tilde{l}_1 = 1$, then the coefficients of the function n also are integers, this time not necessarily non-negative.

3.5. Problems

3.1. Prove that the grammar of Example 3.9 indeed describes the language $\mathcal{F}$ of Example 3.2 and is unambiguous. Find the generating function for the language $\mathcal{F}$ using this grammar.

3.2. Invent an unambiguous grammar generating the language $\mathcal{F}$ and having one generating symbol.

3.3. Show that the grammar

$$\begin{aligned} r &\longrightarrow \lambda; \\ r &\longrightarrow ra; \\ r &\longrightarrow br; \\ r &\longrightarrow arb. \end{aligned}$$

is ambiguous. (It suffices to find a word which has two distinct representations as the right-hand side of these production rules.)

3.4. Write out production rules for the language of regular bracket structures over the alphabet of brackets of two kinds (round and square) and deduce the generating function for this language. For example, [([])][] is a regular bracket structure, while [(]) is not. This language is called the *Dyck language of the second order*. Generalize this result to Dyck languages of arbitrary order.

3.5. Invent formal grammars for the languages of systems of paths from Problems 2.11, 2.12; deduce from the constructed grammars the corresponding generating functions.

3.6. The *Motzkin language* is the language consisting of words over the alphabet $\{a, b, c\}$ such that erasing all occurrences of the letter c from it one obtains a word of the Dyck language. The words in the Motzkin language are in one-to-one correspondence with the Motzkin paths introduced in Problem 2.9. Construct an unambiguous grammar for the Motzkin language and find the generating function for the Motzkin language using this grammar.

3.7. Construct a grammar for the language of non-negative binary integers. (Take into account that the only binary integer starting with 0 is 0.)

3.8. Construct a grammar for the language of regular binary arithmetic expressions over the alphabet $\{(,), +, 1, 0\}$.

3.9. Construct grammars for the following languages:

a) $L_1 = \{a^{3i}b^i \,|\, i \geq 0\}$;

b) $L_2 = \{a^i b^j \,|\, i \geq j \geq 0\}$;

c) $L_3 = \{w \,|\,$ the number of occurrences of the letter a in w coincides with that of b $\}$;

d) $L_4 = \{w \,|\,$ the number of occurrences of the letter a in w is twice as much as that of $b\}$;

e) $L_5 = \{$words in a one-letter alphabet, whose length is divisible either by 2 or by 3$\}$;

f) $L_6 =$ the set of palindromes in the three-letter alphabet. (A *palindrome* is a word having the same shape when read from left to the right and from right to the left.)

g) $L_7 = \{$words in the alphabet $\{a, b\}$ containing an even number of a's$\}$.

Find the generating functions for these languages.

3.10. Prove Remark 3.12.

3.11. Find the generating functions for the languages over the two-letter alphabet consisting of words not containing

a) the subword *ba*; b) the subword *aabb*; c) the subword *aba*.

3.12. Prove that the generating function for the language over $\{a, b\}$ consisting of all words such that the number of occurrences of a in each subword differs from that of b at most by 2 is rational. Construct an unambiguous formal grammar for this language and find this generating function.

Chapter 4

Analytic Properties of Functions Represented as Power Series and the Asymptotics of their Coefficients

4.1. Exponential estimates for asymptotics

When solving enumerative problems, one is often interested in the behavior of the number of elements in the set with the growth of the enumeration parameter. This behavior is important, for example, if we plan to enumerate the objects by means of a computer program and try to predict the expected working time of the program.

Definition 4.1. Two functions $f\colon \mathbb{N} \to \mathbb{R}$ and $g\colon \mathbb{N} \to \mathbb{R}$ have *the same asymptotics*, or *the same growth rate*, as $n \to \infty$, if there is a limit $\lim\limits_{n\to\infty} \frac{f(n)}{g(n)}$ and this limit equals 1. A function f *grows slower* than g if the limit $\lim\limits_{n\to\infty} \frac{f(n)}{g(n)}$ exists and is 0. In the latter case one also says that g *grows faster* than f.

When computing the asymptotics, we usually choose some functions for "patterns" and "reduce" other functions to these patterns.

For the patterns, usually the most ordinary monotonous functions with simple well-known behavior are chosen. The usual patterns are

- the exponential function a^n for different values of the base a;
- the power function n^α for different values of the power α;
- the factorial $n!$;
- the logarithm $\ln n$;

and also various products and compositions of these functions.

It is easy to order the pattern functions in the decreasing rate of their growth:

$$n!;\quad a^n, a>1;\quad n^\alpha, \alpha>0;\quad \ln n;\quad n^\alpha, \alpha<0;\quad a^n, 0<a<1.$$

Example 4.2. The coefficients of the generating function $\ln(1-s)^{-1} = s + \frac{1}{2}s^2 + \frac{1}{3}s^3 + \dots$ grow as n^{-1} (although in this case it would be more natural to say that they "decrease as n^{-1}").

In the discussion below we treat the formal variable s as the *complex* variable $s \in \mathbb{C}$.

The simplest and most often used way to estimate the growth rate of coefficients of a generating function is provided by the following theorem based on the Cauchy criterion for the convergence of a number series.

Theorem 4.3. *Suppose the series $F(s)$ converges at some point s_0, $|s_0| = r$. Then the sequence of coefficients of the series grows slower than $\left(\frac{1}{r}+\varepsilon\right)^n$ for arbitrary positive real number ε.*

Corollary 4.4. *If a power series converges on the entire plane, then the sequence of its coefficients grows slower than ε^n for arbitrary positive real number ε.*

To each power series (with number coefficients that may be integer, real, or complex) its disc of convergence is associated. The *disc of convergence* of a power series is the largest open (that is, not containing the boundary) disc centered at the origin of the complex plane such that the series converges at each of its points. The disc of convergence can either be empty, or coincide with the entire plane, or have a finite non-zero radius R. In the last case R is called the *radius of convergence* of the power series and the following useful statement is true.

Statement 4.5. *The radius of convergence of a power series $F(s)$ coincides with the module of the singular point of the function F closest to the origin.*

For the proof, we refer the reader to standard calculus courses.

Instead of giving a formal definition of a *singular point* of a function, let us give several illustrations to this notion.

Example 4.6. Consider the generating function for the Fibonacci sequence $\mathrm{Fib}(s) = \frac{1}{1-s-s^2}$ (see Example (2.6)). The singularities of this function are those values of s, where the denominator is 0, that is, the roots $s_1 = (-1+\sqrt{5})/2$, $s_2 = (-1-\sqrt{5})/2$ of the quadratic polynomial in the denominator. The point s_1 is closer to the origin than the point s_2. Therefore, the radius of convergence of the Fibonacci series is

$$R_{\mathrm{Fib}} = \frac{\sqrt{5}-1}{2}.$$

Now Theorem 4.3 immediately implies that the Fibonacci numbers grow slower than $\left(\frac{\sqrt{5}+1}{2}+\varepsilon\right)^n$ for arbitrary $\varepsilon > 0$. Equation (2.7) allows one to give a more precise statement. In fact, the Fibonacci numbers grow as $\frac{1}{\sqrt{5}}\left(\frac{\sqrt{5}+1}{2}\right)^{n+1}$. Indeed, Eq. (2.7) yields

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

where $|c| = \left|\frac{\sqrt{5}-1}{\sqrt{5}+1}\right| < 1$. Clearly, c^n tends to 0 as $n \to \infty$.

Example 4.7. The generating function for the Catalan numbers (see Eq. (2.14)) is

$$\mathrm{Cat}(s) = \frac{1-\sqrt{1-4s}}{2s}.$$

The root $s = 0$ of the denominator is not a singularity of the function since at $s = 0$ the numerator also vanishes. The only singularity of $\mathrm{Cat}(s)$ is the point, where the square root becomes zero, i.e., the point $s = \frac{1}{4}$. Hence the Catalan numbers grow slower than $(4+\varepsilon)^n$ for arbitrary $\varepsilon > 0$. For more precise asymptotics, see below.

4.2. Asymptotics of hypergeometric sequences

In enumerative problems, one often meets sequences such that the ratio of two successive elements is expressed as the ratio of two polynomials of the same degree. For a geometric series, for example, this ratio is simply a constant. If the degrees of the polynomials are greater than zero, then the corresponding sequence is termed *hypergeometric*. The following statement gives a very good description of the asymptotics of hypergeometric sequences.

Lemma 4.8. *Suppose $a_0, a_1, \dots$ is a sequence of positive real numbers such that*

$$\frac{a_{n+1}}{a_n} = A\frac{n^k + \alpha_1 n^{k-1} + \dots + \alpha_k}{n^k + \beta_1 n^{k-1} + \dots + \beta_k} \tag{4.1}$$

for all n large enough, and suppose $\alpha_1 \neq \beta_1$. Then a_n grows as

$$a_n \sim cA^n n^{\alpha_1 - \beta_1} \tag{4.2}$$

for some constant $c > 0$.

Remark 4.9. The assumptions of the lemma do not allow one to determine the value of the constant c. Indeed, multiplying the sequence a_n by a constant $d > 0$ we obtain a new hypergeometric sequence with the same ratio of successive terms, but with c replaced by dc.

Example 4.10. For the Catalan numbers (cf. Eq. (2.16)) we have

$$\frac{c_{n+1}}{c_n} = \frac{4n+2}{n+2} = 4\frac{n+\frac{1}{2}}{n+2}.$$

Therefore, $c_n \sim c \cdot 4^n \cdot n^{-3/2}$ for some constant c.

Example 4.11. Let us find the asymptotics of the sequence of coefficients of the function $(a-s)^\alpha$, where α is real. We already know the asymptotics in several cases, for example, for $\alpha = -1$. By the definition of the function $(1-s)^\alpha$, one has

$$(4.3)\qquad \begin{aligned}(a-s)^\alpha &= a^\alpha\left(1-\frac{s}{a}\right)^\alpha\\ &= a^\alpha\Big(1-\frac{\alpha}{1!}\frac{s}{a}+\frac{\alpha(\alpha-1)}{2!}\left(\frac{s}{a}\right)^2\\ &\qquad -\frac{\alpha(\alpha-1)(\alpha-2)}{3!}\left(\frac{s}{a}\right)^3+\dots\Big).\end{aligned}$$

If α is a non-negative integer, then the series is finite, and no asymptotics is required. Otherwise, all coefficients in (4.3) have the same sign starting with some number. The asymptotics can be found by applying Lemma 4.8 to the sequence $a_n = (-1)^n\frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!a^n}$:

$$(4.4)\qquad \frac{a_{n+1}}{a_n} = \frac{1}{a}\frac{n-\alpha}{n+1}.$$

Therefore, $a_n \sim c\cdot a^{-n}\cdot n^{-\alpha-1}$. For example, the coefficients of the function $-(1-4s)^{1/2}$ have the asymptotics $c\cdot 4^n\cdot n^{-3/2}$, and we obtain the asymptotics for the Catalan numbers once again.

Proof of the lemma. The statement of the lemma is equivalent to the fact that there is a limit

$$\lim_{n\to\infty}\frac{a_n}{A^n n^{\alpha_1-\beta_1}},$$

and this limit is non-zero. Taking the logarithm of both parts, we arrive at the necessity to prove the existence of the limit

$$(4.5)\qquad \lim_{n\to\infty}(\ln a_n - n\ln A - (\alpha_1-\beta_1)\ln n).$$

We will prove the existence of the limit (4.5) by using the Cauchy criterion, i.e., we are going to prove that the sequence under consideration is fundamental. This means that for any $\varepsilon > 0$ there is a number N such that for any $n > N$ and each positive integer m one has

$$\begin{aligned}|\ln a_{n+m} &- \ln a_n - (n+m)\ln A + n\ln A\\ &- (\alpha_1-\beta_1)\ln(n+m) + (\alpha_1-\beta_1)\ln n| < \varepsilon,\end{aligned}$$

or

$$(4.6)\qquad |\ln a_{n+m} - \ln a_n - m\ln A - (\alpha_1-\beta_1)\ln(n+m) + (\alpha_1-\beta_1)\ln n| < \varepsilon.$$

Let us rewrite the ratio $\frac{a_{n+1}}{a_n}$ in the form

$$\begin{aligned} \frac{a_{n+1}}{a_n} &= A\frac{1+\alpha_1 n^{-1}+\dots+\alpha_k n^{-k}}{1+\beta_1 n^{-1}+\dots+\beta_k n^{-k}} \\ &= Af\left(\frac{1}{n}\right), \end{aligned} \tag{4.7}$$

where

$$f(x) = \frac{1+\alpha_1 x+\dots+\alpha_k x^k}{1+\beta_1 x+\dots+\beta_k x^k}. \tag{4.8}$$

Taking the logarithm of Eq. (4.7) we obtain

$$\ln a_{n+1} - \ln a_n = \ln A + \ln f\left(\frac{1}{n}\right). \tag{4.9}$$

Let us look at the function $\ln f(x)$. The first terms of the power series expansion at 0 of the function f defined by Eq. (4.8) are

$$f(x) = 1 + (\alpha_1 - \beta_1)x + \gamma x^2 + \dots$$

for some constant γ. This expansion is the central point of the proof. This is the coefficient $\alpha_1 - \beta_1$ (which is non-zero by the assumptions of the theorem) of the linear term that guarantees the appearance of the factor $n^{\alpha_1-\beta_1}$ in the asymptotics. For the logarithm of f one has

$$\ln f(x) = (\alpha_1 - \beta_1)x + \tilde{\gamma}x^2 + \dots$$

Therefore, for sufficiently small x one has $|\ln f(x) - (\alpha_1 - \beta_1)x| < Cx^2$ for some constant C. In particular, if N is sufficiently large, then $\forall n > N$,

$$\begin{aligned} &\left|\ln a_{n+1} - \ln a_n - \ln A - (\alpha_1-\beta_1)\frac{1}{n}\right| < C\frac{1}{n^2}, \\ &\left|\ln a_{n+2} - \ln a_{n+1} - \ln A - (\alpha_1-\beta_1)\frac{1}{n+1}\right| < C\frac{1}{(n+1)^2}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\left|\ln a_{n+m} - \ln a_{n+m-1} - \ln A - (\alpha_1-\beta_1)\frac{1}{n+m}\right| < C\frac{1}{(n+m)^2}. \end{aligned} \tag{4.10}$$

Now the expression on the left-hand side of inequality (4.6) we are interested in can be estimated by means of Eq. (4.10) and the triangle inequality:

$$
\begin{aligned}
(4.11)\quad & |\ln a_{n+m} - \ln a_n - m\ln A - (\alpha_1-\beta_1)(\ln(n+m)-\ln n)| \\
= \ & |\ln a_{n+m} - \ln a_{n+m-1} + \ln a_{n+m-1}\cdots + \ln a_{n+1} - \ln a_n \\
& -m\ln A - (\alpha_1-\beta_1)\sum_{k=0}^{m-1}\frac{1}{n+k} + (\alpha_1-\beta_1)\sum_{k=0}^{m-1}\frac{1}{n+k} \\
& -(\alpha_1-\beta_1)(\ln(n+m)-\ln n)| \\
\le \ & \left|\ln a_{n+1} - \ln a_n - \ln A - (\alpha_1-\beta_1)\frac{1}{n}\right| \\
& + \left|\ln a_{n+2} - \ln a_{n+1} - \ln A - (\alpha_1-\beta_1)\frac{1}{n+1}\right| \\
& \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
& + \left|\ln a_{n+m} - \ln a_{n+m-1} - \ln A - (\alpha_1-\beta_1)\frac{1}{n+m}\right| \\
& + |\alpha_1-\beta_1|\left|\sum_{k=0}^{m-1}\frac{1}{n+k} - \ln(n+m) + \ln n\right| \\
\le \ & C\left(\frac{1}{n^2} + \frac{1}{(n+1)^2} + \cdots + \frac{1}{(n+m-1)^2}\right) \\
& + |\alpha_1-\beta_1|\left|\sum_{k=0}^{m-1}\frac{1}{n+k} - \ln(n+m) + \ln n\right|.
\end{aligned}
$$

Since the series $\sum_{k=1}^{\infty} 1/k^2$ converges, the first summand on the right-hand side of the last inequality can be made arbitrarily small for n large enough. In order to estimate the second summand, remark that the sum in it is nothing but the area bounded by the graph of the stepwise function $\frac{1}{[x]}$ on the segment $[n, n+m]$; see Fig. 1. (Here $[x]$ denotes the integer part of x, that is, the maximal integer number not exceeding x.) This area lies between that under the graph of the function $y = \frac{1}{x}$, and the graph of the function $y = \frac{1}{x-1}$ on the same segment. The area bounded by the graph of the function $y = \frac{1}{x}$ is $\ln(n+m) - \ln n$, while that bounded by the graph of the function $\frac{1}{x-1}$ is $\ln(n+m-1) - \ln(n-1)$. Hence, the difference we are interested

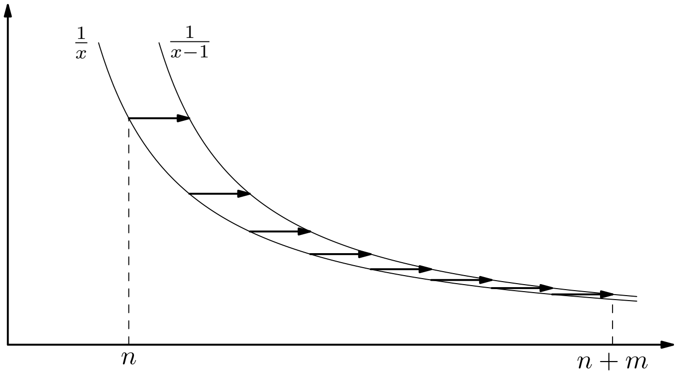


Figure 1. The graph of the function $y = \dfrac{1}{[x]}$ on the segment $[n, n+m]$

in is at most

$$\begin{aligned}&|(\ln(n+m-1)-\ln(n-1))-(-\ln(n+m)+\ln n)|\\&\qquad=\left|\ln\left(1-\frac{1}{n+m}\right)-\ln\left(1-\frac{1}{n}\right)\right|<\left|\ln\left(1-\frac{1}{n}\right)\right|<C\frac{1}{n}.\end{aligned}$$

This completes the proof of the lemma.

4.3. Asymptotics of coefficients of functions related by the Lagrange equation

Suppose two generating functions $\varphi = \varphi(s)$ and $\psi = \psi(t)$ are related by the Lagrange equation (see Eq. (3.7))

$$\varphi(s) = s\psi(\varphi(s)). \tag{4.12}$$

We would like to know whether their radii of convergence are related to each other. From the first glance, there is no connection at all. Indeed, in Example 3.7 we have seen that if $\psi(t) = 1 + t^2$ is the generating function for the sublanguage of indecomposable words in the Dyck language, then $\varphi(s)$ is s times the generating function for the Dyck language. While the first function is a polynomial and hence converges at the entire plane, the radius of convergence of the second one is $\frac{1}{16}$. The situation is similar for all languages generated by unambiguous grammars with a finite set of production rules. It changes dramatically, however, if there are infinitely many production rules (indecomposable words).

Theorem 4.12. *Suppose two generating functions* $\varphi = \varphi(s)$ *and* $\psi = \psi(t)$, $\psi(0) = 1$, *with non-negative coefficients are related by the Lagrange equation* (4.12). *Let* $r > 0$ *be the radius of convergence of the series* φ *and suppose the number series* $\varphi(r)$ *converges. Then the radius of convergence of the series* ψ *is at least* $\rho = \varphi(r)$. *If the number series* $\varphi'(r)$ *also converges, then the radius of convergence of the function* ψ *is exactly* $\rho = \varphi(r)$.

Remark 4.13. The non-negativity assumption for the coefficients of the series is natural if we consider generating functions for languages. In this case it is also natural to expect that the radius of convergence of the generating series for the sublanguage of indecomposable words is greater than that for the whole language (since the number of all words of given length is greater than the number of indecomposable words of the same length).

Proof. Let us prove that the series $\psi(s)$ converges absolutely at each point s, $|s| = q < \rho$. Since the function φ is monotonous and continuous on the segment $[0, r]$, there is a point $p \in [0, r]$ such that $\varphi(p) = q$. Therefore, for each truncated series $\psi^{[n]}(s) = \psi_0 + \psi_1 s + \dots + \psi_n s^n$ we have

$$|\psi^{[n]}(s)| \leq \psi^{[n]}(q) = \psi^{[n]}(\varphi(p)) \leq \varphi(p),$$

where the last inequality follows from the above remark.

The first statement of the theorem is proved.

Now let us rewrite the Lagrange equation (4.12) in the form

$$\psi(\lambda) = \frac{\lambda}{\varphi^{-1}(\lambda)}.$$

The functions $\psi(\lambda)$ and $\varphi^{-1}(\lambda)$ are defined and holomorphic in the disc of radius ρ. The theorem will be proved if we show that the function $\varphi^{-1}(\lambda)$ admits no holomorphic extension to a neighborhood of the point ρ. Suppose that such an extension exists. Then

$$\left(\varphi^{-1}\right)'(\rho) = \lim_{\lambda\to\rho-0}\left(\varphi^{-1}\right)'(\lambda) = \frac{1}{\lim\limits_{t\to r-0}\varphi'(t)}.$$

The last limit exists, and by the assumptions of the theorem is positive. Hence the function φ^{-1} is invertible in a neighborhood of the point ρ, which contradicts the assumptions of the theorem.

Note that the generating series for the Catalan numbers $\mathrm{Cat}(s)$ converges at $s = r = \frac{1}{4}$, since the Catalan numbers have the asymptotics $4^n \cdot n^{-3/2}$ and the series $\sum n^{-3/2}$ converges. On the other hand, the coefficients of the derivative $\mathrm{Cat}'(s)$ have the asymptotics $4^n \cdot n^{-1/2}$, whence the series $\mathrm{Cat}'(\frac{1}{4})$ diverges. Therefore, Theorem 4.12 cannot be applied to the Catalan series *in corpore*, and the second statement of the theorem is not true for this series.

4.4. Asymptotics of coefficients of generating series and singularities on the boundary of the disc of convergence

We have seen already that the radius of convergence of a generating series is determined by the closest to the origin singular point of the series. If the radius of convergence is finite (that is, it is neither zero nor infinity), then the asymptotics of the coefficients is closely related to the nature of the singularities on the boundary of the disc of convergence.

The simplest kind of singularities is a *pole*, the singularity of the form $(1 - s/a)^{-k}$ for a positive integer k. Rational generating functions possess only these singularities. The coefficients of a generating function with such singularity have asymptotics $n^{k-1}a^n$.

Algebraic and algebro-logarithmic singularities are more complicated.

Definition 4.14. A singular point $s = A$ is called an *algebro-logarithmic singular point* of a function $f(s)$ if in some neighborhood of A the function f admits a representation as a finite sum of functions of the form

$$(s - A)^{-\alpha} \ln^k (s - A)\varphi(s), \tag{4.13}$$

where α is a complex number, k is a non-negative integer, φ does not have a singularity at A, and $\varphi(A) \neq 0$.

The coefficients of the series (4.13) have the following asymptotics:

$$
\begin{array}{ll}
\text{const} \cdot A^n n^{\operatorname{Re}\alpha - 1} \ln^k n & \text{for } \alpha \neq 0, -1, -2, \dots \\
\text{const} \cdot A^n n^{\alpha - 1} \ln^{k-1} n & \text{for } \alpha = 0, -1, -2, \dots
\end{array}
\tag{4.14}
$$

We have shown in Sec. 3.4 that unambiguous formal grammars naturally lead to algebraic generating functions. The Hadamard product of rational functions is rational (see Theorem 2.4). A similar statement is true for the product of an algebraic and a rational generating function:

Theorem 4.15. *If $f(s)$ is a rational and $g(s)$ an algebraic generating function, then their Hadamard product is an algebraic function.*

However, in contrast to rational functions, the algebraic generating functions themselves are not closed with respect to the Hadamard product. For example, the Hadamard square of $(1-s)^{-1/2}$ (and, more generally, Hadamard products of functions $(1-s)^{-\alpha}$) is non-algebraic. A natural set of functions closed with respect to the Hadamard product is formed by functions with algebro-geometric singularities. More precisely, the following theorem is true.

Theorem 4.16. *If two functions $f(s)$ and $g(s)$ have only algebro-logarithmic singularities on the boundary of their disc of convergence, then the same is true for their Hadamard product. Moreover, the radius of convergence of the Hadamard product coincides with the product of the radii of convergence of the two factors.*

Here is another important result concerning the Hadamard product, due to Hurwitz.

Theorem 4.17 (Hurwitz)**.** *If each of the functions f and g is a solution of a homogeneous ordinary differential equation with polynomial coefficients, then the same is true for their Hadamard product.*

4.5. Problems

4.1. Find the asymptotics of the Motzkin numbers (see Problem 2.9).

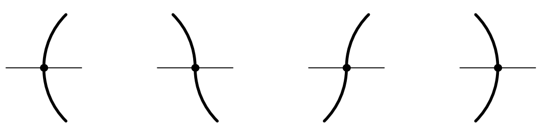
4.2. Find the asymptotics for the number of words in the Dyck language: a) of the second; b) of arbitrary order (see Problem 3.4).

4.3. Find the asymptotics for the number of paths of length k from Problems 2.11, 2.12.

4.4. Find the asymptotics for the number of words for the languages from Problems 3.9, 3.11.

4.5. Consider points $1, 2, \ldots, 2n$ on the horizontal line and join them in pairs by a set of n non-intersecting semicircles in the upper half-plane and n non-intersecting semicircles in the lower half-plane. Such a picture is called a *system of meanders* of order n. Find the asymptotics for the number of systems of meanders.

A system of meanders can be encoded by a word of length $2n$ over the four-letter alphabet $\{a, b, c, d\}$ by assigning to each of the points $1, \ldots, 2n$ one of the letters according to the following rule:



a b c d

(The picture shows the local behavior of the upper and of the lower semicircle passing through the chosen point.) Find the asymptotics for the number of indecomposable systems of meanders (i.e., for the sublanguage of indecomposable words in the language of systems of meanders). For example, the systems of meanders of order 2 have the encoding *aadd*, *adad*, *abcd*, *acbd*; the last two of them are indecomposable.

4.6. Prove that the Hadamard square of the generating function for the Catalan numbers is not an algebraic function.

4.7. Denote by a_k the number of ways to pack the quadrangle $3 \times 2k$ by non-overlapping 1×2-tiles. For example, $a_1 = 3$, $a_2 = 11$. Find the generating function for the numbers a_k and their asymptotics.

Chapter 5

Generating Functions of Several Variables

5.1. The Pascal triangle

Generating functions of two variables correspond to two-index sequences. It is convenient to write down such sequences in the form of a triangle (corresponding to the positive quadrant of the integer lattice).

The *Pascal triangle* is shown in Fig. 1. The entries of this triangle enumerate paths from the vertex of the triangle to the corresponding entry. Each path is a broken line consisting of unit vectors of either of the two kinds: going to the right down or to the left down.

The numbers in the Pascal triangle are the already well-known to us binomial coefficients,

$$c_{n,k} = \binom{n}{k}.$$

This can be easily shown by induction over n. Suppose that the numbers in the nth row of the Pascal triangle coincide with the coefficients in the expansion of $(1+s)^n$. The number of different paths going to the point $(n+1,k)$ coincides with the sum of the number of paths ending at the point $(n,k-1)$ and the number of paths ending at the point (n,k), $c_{n+1,k} = c_{n,k-1}+c_{n,k}$. Therefore, the number $c_{n+1,k}$ coincides with the coefficient of s^k in the polynomial $(1+s)\cdot(1+s)^n = (1+s)^{n+1}$.

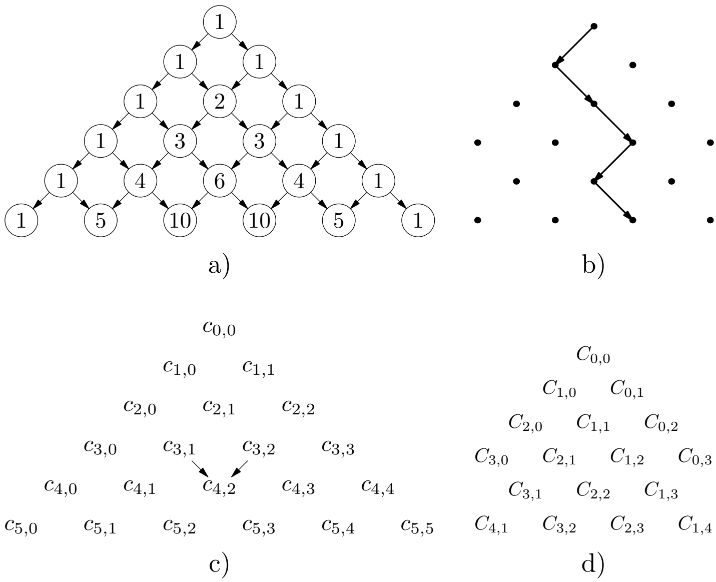


Figure 1. The Pascal triangle and the paths it enumerates

A generating function can be associated to the Pascal triangle in several different ways. For example, one can consider the generating function

$$\sum_{n,k=0}^{\infty} c_{n,k}x^k y^n = \sum_{n,k=0}^{\infty} \binom{n}{k} x^k y^n = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} \binom{n}{k} x^k \right) y^n$$
$$= \sum_{n=0}^{\infty} (1+x)^n y^n = \frac{1}{1-y-xy}.$$

Another possibility consists of numbering the entries of the triangle by the numbers of segments of each of the two types in any path leading to the entry (see Fig. 1 d)). For this numbering,

$$C_{n,m} = c_{n+m,m} = \binom{n+m}{m}$$

and the generating function is

$$\begin{aligned}\sum_{n,m=0}^{\infty} C_{n,m}x^n y^m &= \sum_{n,m=0}^{\infty} \binom{n+m}{m} x^n y^m \\ &= \sum_{k=0}^{\infty} \sum_{n+m=k} \binom{n+m}{n} x^n y^m \\ &= \sum_{k=0}^{\infty} (x+y)^k = \frac{1}{1-x-y}.\end{aligned}$$

This time the generating function is symmetric in x and y.

Finally, one can associate to the Pascal triangle the exponential generating function. In contrast to the ordinary generating function, the coefficients of the exponential generating function are not just the elements a_n, but the numbers $a_n/n!$.

5.2. Exponential generating functions

Fix an arbitrary sequence $\{\alpha_n\}$. One can associate to each sequence $\{a_n\}$ the generating function

$$\{a_n\} \mapsto \sum_{n=0}^{\infty} a_n \alpha_n s^n,$$

defined by the sequence $\{\alpha_n\}$. If there are no zeroes in the sequence $\{\alpha_n\}$, then this correspondence is one-to-one. Up to now we made use only of ordinary generating functions, i.e., those corresponding to the sequence $\alpha_n \equiv 1$. Other sequences also may prove to be useful. The choice of the sequence is determined by the problems we are trying to solve. The sequence $\alpha_n = \frac{1}{n!}$ is one of the most frequently used. The corresponding generating functions are said to be *exponential*. Exponential generating functions for integer sequences are called *Hurwitz functions*.

What is the difference between exponential generating functions and ordinary ones? Let us look at the behavior of the exponential generating functions under the usual operations. Their behavior under the summation is similar to that of ordinary ones:

$$\sum_{n=0}^{\infty} \frac{a_n}{n!} s^n + \sum_{n=0}^{\infty} \frac{b_n}{n!} s^n = \sum_{n=0}^{\infty} \frac{(a_n + b_n)}{n!} s^n,$$

while under the multiplication they behave differently:

$$\left(\frac{a_0}{0!} + \frac{a_1}{1!} s + \frac{a_2}{2!} s^2 + \dots\right)\left(\frac{b_0}{0!} + \frac{b_1}{1!} s + \frac{b_2}{2!} s^2 + \dots\right)$$
$$= \frac{a_0}{0!}\frac{b_0}{0!} + \left(\frac{a_0}{0!}\frac{b_1}{1!} + \frac{a_1}{1!}\frac{b_0}{0!}\right) s + \left(\frac{a_0}{0!}\frac{b_2}{2!} + \frac{a_1}{1!}\frac{b_1}{1!} + \frac{a_2}{2!}\frac{b_0}{0!}\right) s^2 + \dots$$

The coefficients $\frac{c_n}{n!}$ of the product are given by the formula

$$c_n = \binom{n}{0} a_0 b_n + \binom{n}{1} a_1 b_{n-1} + \dots + \binom{n}{n} a_n b_0.$$

Another essential difference between exponential and ordinary generating functions is their behavior under the differentiation (and integration). Both the differentiation and integration of an exponential generating function lead to the shift of the sequence of coefficients, without changing the coefficients:

$$\left(\frac{a_0}{0!} + \frac{a_1}{1!} s + \frac{a_2}{2!} s^2 + \dots\right)' = \frac{a_1}{0!} + \frac{a_2}{1!} s + \frac{a_3}{2!} s^2 + \dots ;$$
$$\int \left(\frac{a_0}{0!} + \frac{a_1}{1!} s + \frac{a_2}{2!} s^2 + \dots\right) = \frac{a_0}{1!} s + \frac{a_1}{2!} s^2 + \frac{a_2}{3!} s^3 + \frac{a_3}{4!} s^4 + \dots$$

The ordinary generating function $A(s) = a_0 + a_1 s + a_2 s^2 + \dots$ can be expressed in terms of the exponential one $\mathcal{A}(t) = \frac{a_0}{0!} + \frac{a_1}{1!} t + \frac{a_2}{2!} t^2 + \dots$ according to the formula

$$A(s) = \int_0^{\infty} e^{-t} \mathcal{A}(st) dt.$$

Indeed, it is easy to see that

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

Now we are in position to deduce the exponential generating function for the Pascal triangle:

$$\sum_{n,m=0}^{\infty} \frac{1}{(n+m)!} \binom{n+m}{m} x^n y^m = \sum_{n=0}^{\infty} \frac{(x+y)^n}{n!} = e^{x+y}.$$

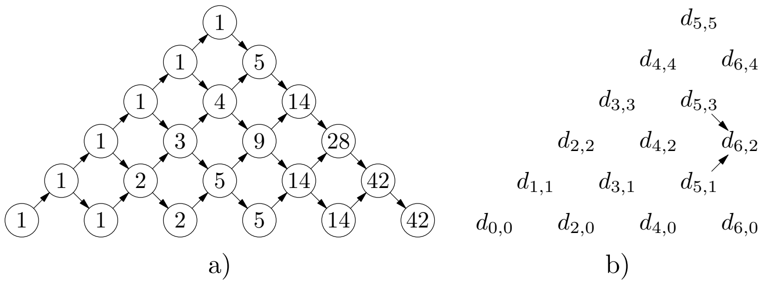


Figure 2. The Dyck triangle

Several more sophisticated examples of exponential generating functions will be discussed later.

5.3. The Dyck triangle

The *Dyck triangle* enumerates paths in the positive quadrant issuing from the origin and consisting of vectors $(1,1)$ and $(1,-1)$ (see Fig. 2). Those paths that end on the y-axis are the Dyck paths from Sec. 2.5.

Clearly, the elements d_{ij} of the Dyck triangle are non-zero if and only if $i \geq j$ and $i+j$ is even. Denote by $D(x,y)$ the generating function of two variables

$$D(x,y) = \sum_{i,j=0}^{\infty} d_{ij} x^i y^j.$$

The construction rule for the Dyck triangle hints the following equation for this generating function:

$$xyD(x,y) + (D(x,y) - D(x,0))\frac{x}{y} = D(x,y) - 1.$$

Indeed, the coefficient of any non-unit monomial $x^i y^j$ is the sum of the coefficients of the monomials $x^{i-1}y^{i-1}$ and $x^{i-1}y^{j+1}$. We already know the function

$$D(x,0) = \frac{1-\sqrt{1-4x^2}}{2x^2},$$

and the series $D(x,y)$ is found by solving the linear equation,

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

5.4. The Bernoulli–Euler triangle and enumeration of snakes

Similarly to the Pascal triangle, the *Bernoulli–Euler triangle* (see Fig. 3) possesses many remarkable properties. The left-hand side of this triangle is called the *Bernoulli side*, and the right-hand side is the *Euler side*[1].

An entry of the Bernoulli–Euler triangle also enumerates paths from the vertex of the triangle to the given entry. However, here we consider only alternating paths: odd steps go to the left, while even steps go to the right (not necessarily to the neighboring entry). That is why each element in the Bernoulli–Euler triangle is equal to the sum of all elements of the previous row situated either to the right, or to the left of the given entry, depending on the parity of the row's number.

One can also define the Bernoulli–Euler triangle by a simpler recurrence rule, after switching the sign in pairs of successive rows (see Fig. 4). In this alternated triangle each entry is the sum of two neighboring entries, to the right and to the up-right of the given one. To make this definition of the triangle unambiguous, one must define the Euler side of the triangle. In order to do this, we shall make use of two other interpretations of the Bernoulli–Euler triangle, those in terms of Morse polynomials and up-down permutations.

Definition 5.1. A point x_0 is a *critical point* of a polynomial $p = p(x)$ if it is a root of the derivative, $p'(x_0) = 0$. The tangent line to the graph of the polynomial at a critical point is horizontal. The value $p(x_0)$ of a polynomial p at a critical point is called a *critical value* of the polynomial. A polynomial p is *Morse* if

a) all its critical points are real and distinct;

b) all its critical values are distinct.

A Morse polynomial of degree $n + 1$ has n critical points and n critical values. We will consider polynomials of the form

$$p(x) = x^{n+1} + a_1 x^n + \cdots + a_{n+1}, \tag{5.1}$$

[1]The reader should not mix these sequences up with other two number sequences, also carrying the names of Bernoulli and Euler.

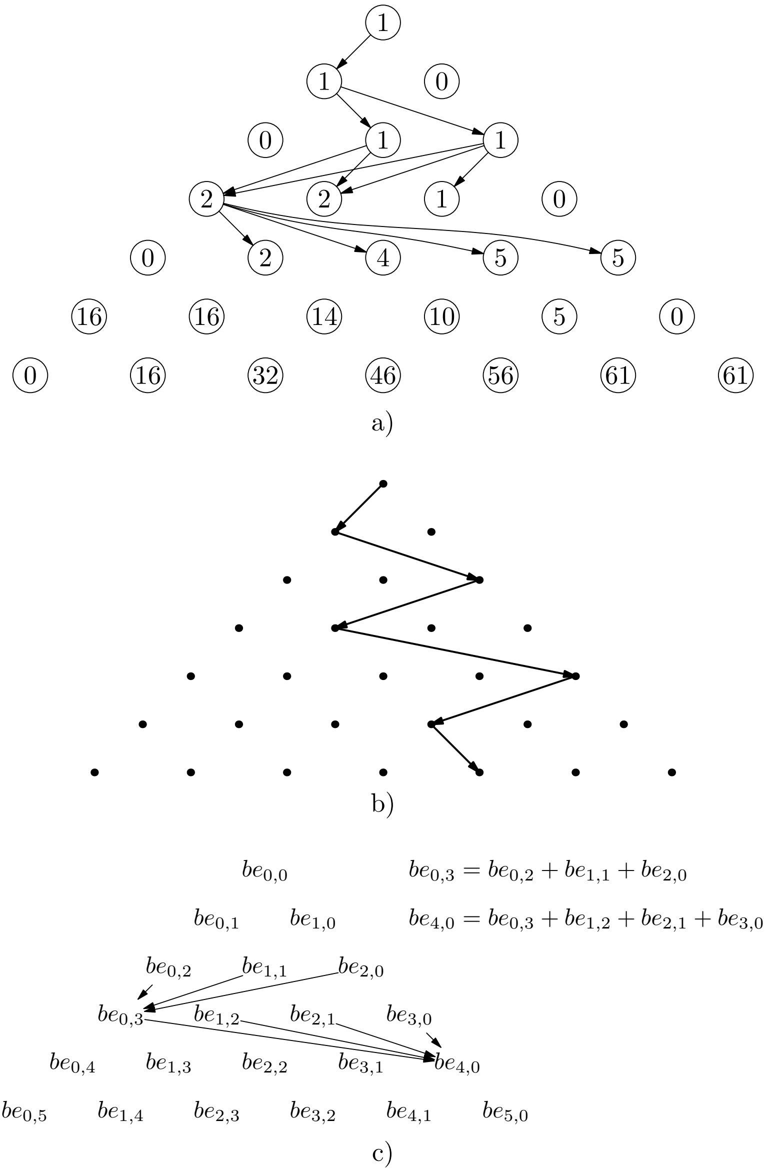


Figure 3. The Bernoulli–Euler triangle and the paths it enumerates

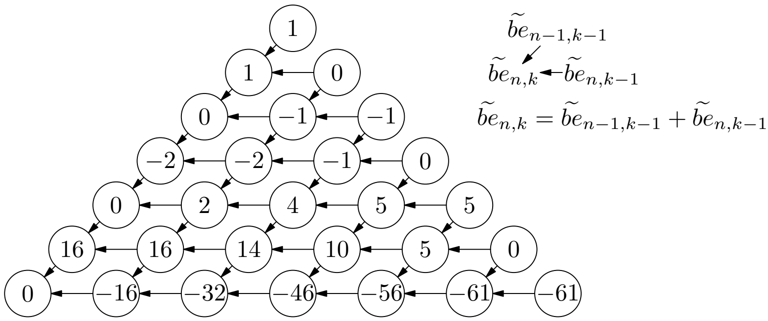


Figure 4. The alternated Bernoulli–Euler triangle

with the leading coefficient equal to 1.

One can associate to a Morse polynomial a permutation on the set of n elements. This permutation shows the order of the critical values of the polynomial. To construct the permutation, let us number the critical points and the critical values in the increasing order. The ith element of the desired permutation is the number of the critical value at the ith critical point (see Fig. 5). Clearly, each element of such a permutation is either bigger than both its neighbors (the corresponding critical value is a local maximum), or smaller than both of them (the corresponding critical value is a local minimum). Such permutations are called *up-down permutations*. The up-down permutation corresponding to a polynomial is called the *type* of this polynomial.

Note that not each up-down permutation can be the type of a polynomial of the form (5.1): the last element of such a permutation must be smaller than its left neighbor. As a result, the first element of such a permutation must be smaller than its right neighbor if n is odd, and must be bigger otherwise.

The types of Morse polynomials for small values of n are shown in Fig. 6. For $n = 1$ and $n = 2$ there is only one possibility. For $n = 3$ the number of cases is 2, and for $n = 4$ the number of cases is 5. Continuing this enumeration we obtain the sequence

$$1, 1, 2, 5, 16, 61, 272, \ldots$$

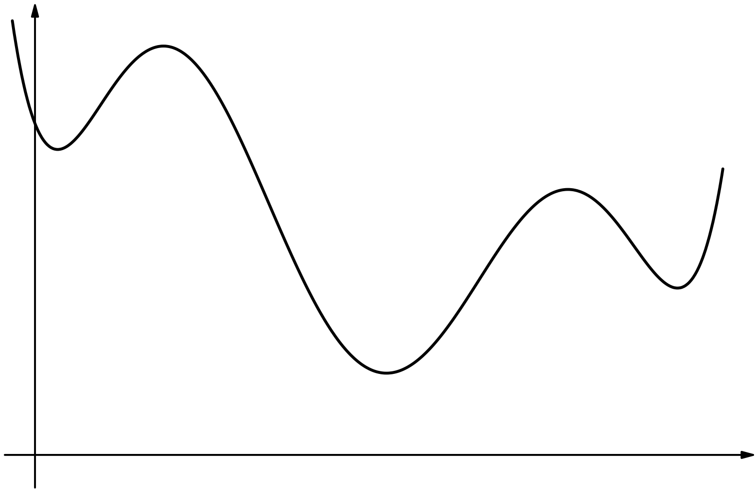


Figure 5. The permutation corresponding to a Morse polynomial

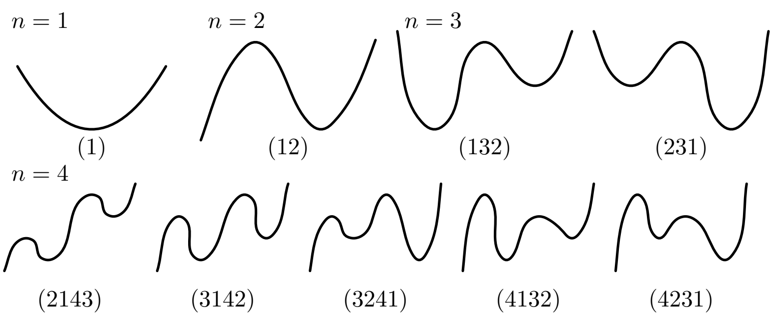


Figure 6. The types of Morse polynomials with $n = 1, 2, 3, 4$

The elements of this sequence corresponding to the odd values of n coincide with those on the Bernoulli side of the Bernoulli–Euler triangle, while the elements corresponding to the even values of n are on the Euler side.

In order to understand where the relationship with the Bernoulli–Euler triangle comes from, let us look at the types of Morse polynomials with the first critical value having a given number k.

Lemma 5.2. *Let $c_{n,k}$ be the number of types of Morse polynomials of degree $n+1$ with the first critical value having number k. Then $c_{n,k}$ is the kth entry in the nth row of the Bernoulli–Euler triangle.*

Proof. For the first two rows of the triangle the verification is absolutely straightforward. Now let us prove that if the statement is true for the nth row, then it remains true for the $n+1$st row as well. Suppose for definiteness that $n+1$ is even. Then both n and $n+2$ are odd; we study the types of polynomials of degree $n+2$. The first critical value of such a polynomial is a local maximum, whence the second one is a local minimum and hence it is smaller than the first one.

Throwing away the first critical value we obtain a uniquely defined type of Morse polynomial of degree $n+1$. The number of the first critical value of this polynomial can be k, $k+1$, $\dots$, n. Conversely, to each type of a polynomial of degree $n+1$ whose first critical value has number l ($l \geq k$) one can associate, in a unique way, a type of a polynomial of degree $n+2$ with the first critical value having number k. Thus,

$$c_{n+1,k} = c_{n,k} + c_{n,k+1} + \cdots + c_{n,n}.$$

For odd rows a similar argument also works. Therefore, the numbers $c_{n,k}$ satisfy the same recurrence relations as the entries of the Bernoulli–Euler triangle, and hence exactly these numbers constitute the triangle.

Consider the following two cases separately:

- n is odd; denote the corresponding number of the up-down permutations by b_n and introduce the exponential generating function

$$\mathcal{B}(x) = \frac{b_1}{1!}x + \frac{b_2}{2!}x^2 + \cdots = \frac{1}{1!}x + \frac{2}{3!}x^3 + \frac{16}{5!}x^5 + \dots ;$$

- n is even; denote the corresponding number of up-down permutations by e_n and introduce the exponential generating function

$$\mathcal{E}(y) = 1 + \frac{e_1}{1!}y + \frac{e_2}{2!}y^2 + \cdots = 1 + \frac{1}{2!}y^2 + \frac{5}{4!}y^4 + \dots$$

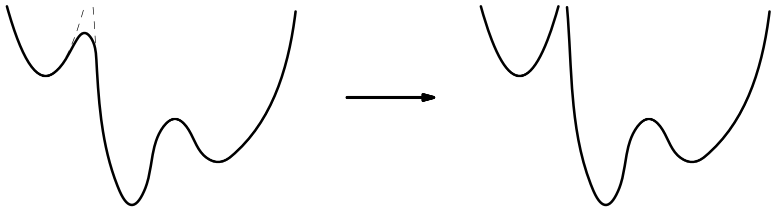


Figure 7. Associating two new types to a type of a polynomial

Now let us deduce a recurrence relation for the numbers of up-down sequences. To do this, let us take the global maximum of the polynomial and tend it to infinity (see Fig. 7). As a result, we have associated to a type of a polynomial two new types, and if the initial polynomial has $n+1$ critical points, then the new polynomials have k and $n-k$ critical points, with $n-k$ odd.

For odd n we obtain the following recurrence relation for the numbers b_n:

$$b_{n+1} = \sum_{k \text{ odd}} \binom{n}{k} b_k b_{n-k}. \tag{5.2}$$

The binomial coefficients arise because we must shuffle the sets of critical values of the left and of the right polynomials, i.e., choose k critical values of the left polynomial among the n critical values (all, but the greatest one).

Recalling that for exponential generating functions the right-hand side of Eq. (5.2) corresponds to the square of the generating function $\mathcal{B}(x)$ and the left-hand side corresponds to its derivative we can rewrite Eq. (5.2) in the form

$$\mathcal{B}'(x) = \mathcal{B}^2(x) + 1. \tag{5.3}$$

The last equation is an ordinary differential equation with separable variables. Solving it we obtain

$$\begin{aligned} d\mathcal{B} &= (\mathcal{B}^2+1)\,dx, \\ \int \frac{d\mathcal{B}}{\mathcal{B}^2+1} &= \int dx, \\ \arctan \mathcal{B} &= x, \\ \mathcal{B}(x) &= \tan x. \end{aligned}$$

Hence, the Bernoulli side gives the expansion of the tangent,

$$\mathcal{B}(x) = \tan x = x + 2\frac{x^3}{3!} + 16\frac{x^5}{5!} + 272\frac{x^7}{7!} + \dots$$

The coefficients b_n in this expansion are also called the *tangential numbers*. Pay attention to the fact that the vertex element 1 of the triangle is not included in the Bernoulli side.

In the case of even n, the recurrence relation has the form

$$e_{n+1} = \sum_{k \text{ odd}} \binom{n}{k} e_k b_{n-k}, \tag{5.4}$$

and the corresponding equation for the exponential generating functions looks like

$$\mathcal{E}'(y) = \mathcal{E}(y)B(y). \tag{5.5}$$

Solving the latter we obtain

$$\begin{aligned} \frac{\mathcal{E}'(y)}{\mathcal{E}(y)} &= B(y), \\ (\ln \mathcal{E}(y))' &= \tan y, \\ \ln \mathcal{E}(y) &= \int \tan y, \\ \mathcal{E}(y) &= \frac{1}{\cos y}, \end{aligned}$$

and we conclude that the Euler side determines the expansion of the inverse cosine. The coefficients e_n of this expansion are called the *Euler numbers* [2].

[2] Usually, the term "Euler numbers" refers to the numbers on the Euler side in the alternated triangle (that is, to the same sequence, but with alternating signs). However, we will not stress this difference.

Using the substitutions

$$\sin x = \left(e^{ix} - e^{-ix}\right)/2i, \quad \cos x = \left(e^{ix} + e^{-ix}\right)/2,$$

we can rewrite the generating functions for the alternated triangle in the form

$$\widetilde{\mathcal{B}}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad \widetilde{\mathcal{E}}(y) = \frac{2}{e^y + e^{-y}}.$$

Now the preparatory work for writing out the exponential generating function for the Bernoulli–Euler triangle is completed. Denote by $\mathrm{be}_{k,l}$ the entry of the triangle having the coordinate k along the Bernoulli side and the coordinate l along the Euler side.

Theorem 5.3. *The exponential generating function for the alternated Bernoulli-Euler triangle is*

$$\mathcal{BE}(x,y) = \sum_{k,l=0}^{\infty} \mathrm{be}_{k,l} \frac{x^k}{k!}\frac{y^l}{l!} = \frac{2e^x}{e^{x+y} + e^{-(x+y)}}. \tag{5.6}$$

Proof. Let us prove that the exponential generating function for the alternated Bernoulli–Euler triangle satisfies the differential equation

$$\mathcal{BE}(x,y) + \frac{\partial \mathcal{BE}(x,y)}{\partial y} = \frac{\partial \mathcal{BE}(x,y)}{\partial x}.$$

This equation is nothing but the defining rule for the alternated triangle. Indeed, consider a half-line in the triangle parallel to the Euler side. The differentiation of the exponential generating function of this half-line over y is nothing but the shift by one along the Euler side. Adding the initial function to the derivative we obtain the neighboring line (since $\mathrm{be}_{k,m} = \mathrm{be}_{k-1,m} + \mathrm{be}_{k-1,m+1}$), i.e., the derivative of the initial exponential generating function with respect to x.

Hence, the function $\mathcal{BE}(x,y)$ is uniquely determined by the initial condition

$$\mathcal{BE}(0,y) = \widetilde{\mathcal{E}}(y) = \frac{2}{e^y + e^{-y}}$$

and the partial differential equation. Now a straightforward verification shows that the function (5.6) satisfies the above differential equation, which completes the proof of the theorem.

5.5. Representing generating functions as continued fractions

The generating function for the Catalan numbers satisfies the quadratic equation (2.13):

$$s^2 \operatorname{Cat}^2(s) - \operatorname{Cat}(s) + 1 = 0.$$

Let us rewrite this equation in the form

$$\operatorname{Cat}(s) - s^2 \operatorname{Cat}^2(s) = 1,$$

or

$$\operatorname{Cat}(s) = \frac{1}{1 - s^2 \operatorname{Cat}(s)}. \tag{5.7}$$

Substituting this expression for Cat(s) into the right-hand side of the same Eq. (5.7) we obtain

$$\operatorname{Cat}(s) = \frac{1}{1 - \frac{s^2}{1 - s^2 \operatorname{Cat}(s)}}.$$

Iterating the process of substitution of the expression (5.7) for Cat(s) in the resulting identity we finally obtain the expression for the Catalan function in the form of a *continued fraction*:

$$\operatorname{Cat}(s) = \frac{1}{1 - \frac{s^2}{1 - \frac{s^2}{1 - \dots}}}. \tag{5.8}$$

The right-hand side of this equation should be interpreted as follows. If we break the continued fraction at the nth level (replacing it thus with a finite ratio, a *convergent* to the continued fraction), then the coefficients of the expansion of the resulting rational function in powers of s will coincide with the coefficients of the expansion of Cat(s) up to the term s^{2n}. Note that since the numerator of the ratio added at the $(n + 1)$th step is divisible by s^2, the increasing of the number of terms in the convergent does not change the first coefficients of the expansion. For example,

$$\frac{1}{1-s^2} = \mathbf{1 + s^2} + s^4 + s^6 + s^8 + \dots,$$

$$\frac{1}{1-\frac{s^2}{1-s^2}} = \mathbf{1 + s^2 + 2s^4} + 4s^6 + 8s^8 + \dots,$$

$$\frac{1}{1-\frac{s^2}{1-\frac{s^2}{1-s^2}}} = \mathbf{1 + s^2 + 2s^4 + 5s^6} + 13s^8 + \dots,$$

$$\frac{1}{1-\frac{s^2}{1-\frac{s^2}{1-\frac{s^2}{1-s^2}}}} = \mathbf{1 + s^2 + 2s^4 + 5s^6 + 14s^8} + \dots$$

The stabilizing part of the expansion is shown in bold.

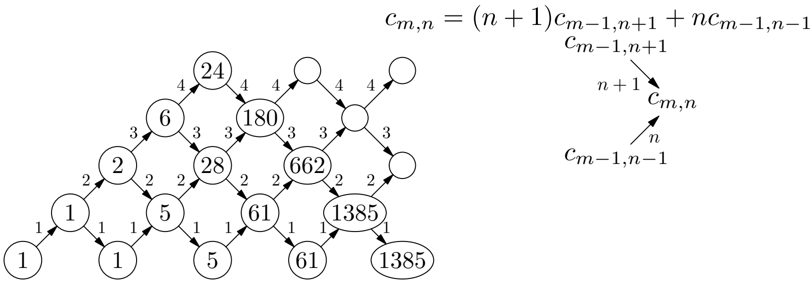
The representation of the Catalan function as a continued fraction is closely related with the two ways of deducing this function, namely, by enumerating the Dyck paths (Sec. 2.5) and by using a generating grammar (Sec. 3.2). Other functions enumerating paths of various kinds also possess similar representations.

Let us modify slightly the Dyck triangle (see Fig. 1) by assigning some numbers to the arrows. To be more precise, we assign to each arrow the number of the row to which it belongs (see Fig. 8 a)). We will interpret the number at an arrow as its multiplicity, i.e., as the number of "distinct" arrows in the same direction. As a result, each path in the Dyck triangle corresponds to several "distinct" paths in the triangle with multiplicities. The number of these paths is equal to the product of all multiplicities of all arrows in the path.

The numbers on the lower row of the triangle in Fig. 8 a) resemble the already well-known sequence of Euler numbers studied in Sec. 5.4. We postpone the proof of the fact that these two sequences indeed coincide with the next section. Now we only construct the continued fraction expansion of the corresponding generating function.

Theorem 5.4. *The generating function*

$$F_0(s) = 1 + s^2 + 5s^4 + 61s^6 + 1385s^8 + \dots$$



a) b)

Figure 8. Dyck triangles with multiplicities

for the lower side of the triangle shown in Fig. 8 a) *admits the following representation as a continued fraction:*

$$F_0(s) = \cfrac{1}{1 - \cfrac{1^2 s^2}{1 - \cfrac{2^2 s^2}{1 - \cfrac{3^2 s^2}{1 - \dots}}}}.$$

Proof. The generating function $F_0(s)$ enumerates distinct paths starting and ending at the level 0. Denote by $F_i(s)$ the generating function enumerating paths starting and ending at the level i and not going below this level, with respect to their lengths. Then

$$F_0(s) = \frac{1}{1 - s^2 F_1(s)}.$$

Indeed, each path starting and ending at level 0 admits a unique decomposition into (broken) segments such that

1) the ends of each segment are on the level 0;

2) the height of each internal point of the segment is positive.

If we erase the first and the last vector in such a segment, then we obtain a path starting and ending at the level 1. This proves the statement.

Similarly,

$$F_1(s) = \frac{1}{1 - 4s^2 F_2(s)}.$$

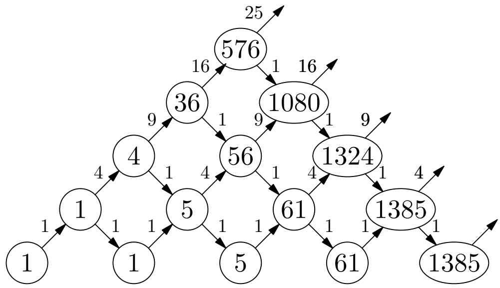


Figure 9. Another distribution of multiplicities in the Dyck triangle

The coefficient 4 at s^2 appears because there are four "distinct" ways to add the first and the last vector to a path on the level 2 in order to make it into a path on the level 1.

Repeating this procedure we conclude that

$$F_k(s) = \frac{1}{1-(k+1)^2 s^2 F_{k+1}(s)},$$

and the way to write out the continued fraction becomes obvious:

$$\begin{aligned} F_0(s) &= \frac{1}{1-s^2F_1(s)} \\ &= \frac{1}{1-\frac{s^2}{1-4s^2F_2(s)}} \\ &= \dots \\ &= \frac{1}{1-\frac{s^2}{1-\frac{4s^2}{1-\frac{9s^2}{1-\dots}}}}. \end{aligned}$$

The proof of the theorem immediately implies that the distribution of the multiplicities among the ascending and descending vectors inside each layer is of no importance. Only their product (which must be constant) makes sense. For example, the triangle shown in Fig. 9 produces the same generating functions for paths starting and ending at the level 0 as the one shown in Fig. 8 a). Note that the same is true for paths on other levels as well.

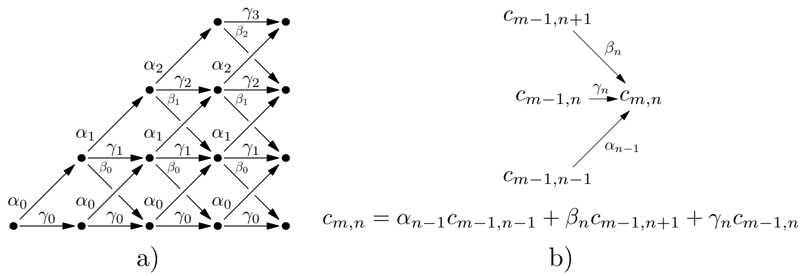


Figure 10. Motzkin triangle with multiplicities

Of course, the proof of the theorem can be immediately rephrased for arbitrary distribution of multiplicities. Moreover, it has an obvious generalization for the Motzkin triangle with multiplicities (see Fig. 10).

Theorem 5.5. *Let* $\alpha_i, \beta_i, \gamma_i$ *denote the multiplicities of the vectors* $(1,1)$, $(1,-1)$, *and* $(1,0)$, *respectively, in the ith layer of the weighted Motzkin triangle. Then the generating function* $F_k(s)$ *for the paths starting and ending at the height k and not going below this height admits the following continued fraction representation:*

$$F_k(s) = \cfrac{1}{1-\gamma_k s - \cfrac{\alpha_k\beta_k s^2}{1-\gamma_{k+1}s-\cfrac{\alpha_{k+1}\beta_{k+1}s^2}{1-\dots}}}.$$

Proof. Of course, this theorem can be proved in the same vein as its special case Theorem 5.4. However, we prefer to express the same proof in the language of formal grammars (Chapter 3). Associate to the vectors $(1,1)$, $(1,-1)$, $(1,0)$ in the ith layer the letters a_i, b_i, c_i, respectively. Consider the languages $\mathcal{F}_0, \mathcal{F}_1, \dots$ over the inifinite alphabet $\{a_0, b_0, c_0, a_1, b_1, c_1, \dots,\}$. The language $\mathcal{F}_k$ consists of the words corresponding to the paths starting and ending at the height k and not going below this height.

The grammar

$$\begin{aligned} r_0 &\longrightarrow \lambda, \\ r_0 &\longrightarrow c_0 r_0, \\ r_0 &\longrightarrow a_0 r_1 b_0 r_0, \\ r_1 &\longrightarrow \lambda, \\ r_1 &\longrightarrow c_1 r_1, \\ r_1 &\longrightarrow a_1 r_2 b_1 r_1, \\ &\dots \end{aligned}$$

is unambiguous. The letter r_k, $k = 0, 1, 2, \dots$ generates the language $\mathcal{F}_k$. Therefore, the non-commutative generating functions $\mathcal{F}_k$ satisfy the equations

$$\begin{aligned} \mathcal{F}_0 &= \lambda + c_0\mathcal{F}_0 + a_0\mathcal{F}_1 b_0 \mathcal{F}_0, \\ \mathcal{F}_1 &= \lambda + c_1\mathcal{F}_1 + a_1\mathcal{F}_2 b_1 \mathcal{F}_1, \\ &\dots\dots\dots\dots\dots\dots\dots\dots \end{aligned}$$

Making the substitution $\lambda = 1$, $a_i = \alpha_i s$, $b_i = \beta_i s$, $c_i = \gamma_i s$, we obtain the following system of equations for the commutative generating functions:

$$\begin{aligned} F_0(s) &= 1 + \gamma_0 s F_0(s) + \alpha_0\beta_0 s^2 F_0(s) F_1(s), \\ F_1(s) &= 1 + \gamma_1 s F_1(s) + \alpha_1\beta_1 s^2 F_1(s) F_2(s), \\ &\dots\dots\dots\dots\dots\dots\dots\dots \end{aligned}$$

whence

$$\begin{aligned} F_0(s) &= \frac{1}{1 - \gamma_0 s - \alpha_0\beta_0 s^2 F_1(s)} \\ &= \frac{1}{1 - \gamma_0 s - \frac{\alpha_0\beta_0 s^2}{1 - \gamma_1 s - \alpha_1\beta_1 s^2 F_2(s)}} \\ &= \dots \end{aligned}$$

For other F_k, the derivation is similar, which completes the proof.

5.6. The Euler numbers in the triangle with multiplicities

Statement 5.6. *The base side of the triangle in Fig.* 8 a) *is formed by the Euler numbers.*

Proof. We will prove that the number of distinct paths of length $2n$ in the Dyck triangle with multiplicities coincides with the number of up-down permutations of the set $\{1, \dots, 2n-1\}$ or, what is the same, with the number of types of Morse polynomials of degree $2n$. Let us add to the underlying set of the permutations the additional element $2n$ which we will treat as the last element of each up-down permutation. (Recall that we take into consideration only those up-down permutations that remain up-down even after adding the maximal element as the last one.)

Associate to each up-down permutation a path in the Dyck triangle in the following way. Each element i of the permutation is either a (local) maximum, or a (local) minimum in it. We choose the ith vector going up if i is a minimal element of the permutation, and going down otherwise. In Fig. 11 a permutation and the corresponding Dyck path are shown. It is clear that such a path is indeed a Dyck path: the number of maximal elements in an up-down permutation coincides with that of minimal elements, and not more than half of the elements are maximal among the first k elements $1, \dots, k$ for arbitrary k.

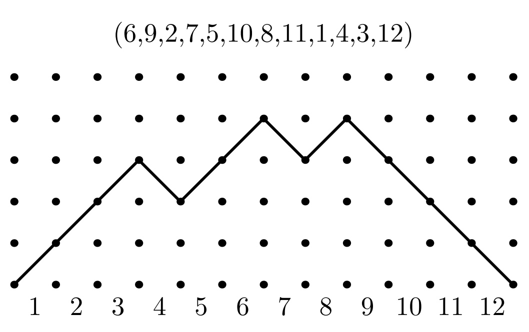


Figure 11. The Dyck path corresponding to an up-down permutation

Let us count the number of permutations corresponding to a given path. Suppose that the path corresponding to the first m elements of the permutation ends at the level k. And suppose also that the last element m is a local maximum (that is, the last vector of the path descends). Which number can be the local minimum on the right of m? This minimum is among the first $m-1$ elements of the permutation, and there are $k+1$ different ways to choose it. Indeed, if a right neighboring minimum is already assigned to each of the maxima among the first elements $1, \dots, m-1$, then there are exactly $k+1$ free minima.

The reasoning for the case, where the last element m is a minimum, proceeds similarly, but we must choose the right neighboring maximum and move along the permutation from right to left. The statement is proved.

5.7. Congruences in integer sequences

This section is devoted to properties of integer sequences reduced with respect to various moduli.

Consider, for example, the sequence of Euler numbers

$$1, 1, 5, 61, 1385, \dots$$

The remainders of these integers divided by 4 form the new sequence

$$1, 1, 1, 1, 1, \dots$$

One can check that all other elements of this sequence also are ones. The same sequence considered modulo 3 looks as follows:

$$1, 1, 2, 1, 2, 1, 2, \dots$$

The periodicity of the sequence hints that it is given by a rational generating function. Indeed, let N be a period length of a sequence a_i, i.e., $a_{k+N} = a_k$ for all sufficiently large k. This means that the sequence is given by a linear recurrence relation with constant coefficients, and hence, by Theorem 2.1, the corresponding generating function is rational.

It is easy to find this rational function for the Euler numbers. Indeed, consider the continued fraction expansion of the generating function for them:

$$E(s) = \cfrac{1}{1 - \cfrac{1^2 s^2}{1 - \cfrac{2^2 s^2}{1 - \cfrac{3^2 s^2}{1 - \dots}}}}.$$

When the sequence is reduced modulo 4, the second term of this continued fraction vanishes, and the fraction acquires the form

$$E_4(s) \equiv \frac{1}{1 - s^2} \bmod 4.$$

(Two power series with integer coefficients are congruent if the corresponding coefficients of the series are congruent with respect to the given module.) When reduced modulo 3, the third term vanishes and the entire fraction becomes

$$E_3(s) \equiv \cfrac{1}{1 - \cfrac{s^2}{1-4s^2}} \bmod 3 \equiv \frac{1 - s^2}{1 + s^2} \bmod 3.$$

More generally, reducing the Euler sequence modulo p we obtain a finite fraction

$$E_p(s) \equiv \cfrac{1}{1 - \cfrac{1^2 s^2}{1 - \dots \cfrac{(p-2)^2 s^2}{1-(p-1)^2 s^2}}} \bmod p,$$

since the coefficient p^2 of the next term is zero modulo p, and hence the whole tail of the fraction is zero. It is clear how to extend this argument to an arbitrary continued fraction.

Theorem 5.7. *Suppose the generating function $A(s)$ is represented as a continued fraction*

$$A(s) = \cfrac{1}{1 - c_1 s - \cfrac{p_1 s^2}{1 - c_2 s - \cfrac{p_2 s^2}{1 - c_3 s - \cfrac{p_3 s^2}{1 - \dots}}}}.$$

Then the function $A_p(s) \equiv A(s) \bmod p$ is rational for an arbitrary number p which is a divisor of one of the products $p_1, p_1 p_2, p_1 p_2 p_3, \dots$ If p divides the product $p_1 \dots p_k$, then

$$A_p(s) \equiv \cfrac{1}{1 - c_1 s - \cfrac{p_1 s^2}{1 - \dots \cfrac{p_{k-2} s^2}{1 - c_{k-1} s - p_{k-1} s^2}}} \bmod p.$$

Thus, Theorem 5.7 allows one to find rational generating functions for sequences corresponding to weighted Dyck triangles, reduced with respect to some (and, sometimes, to all as in the case of the Euler numbers) moduli.

Another approach to the study of arithmetic properties of integer sequences is based on using various combinatorial interpretations for them. Here is the simplest example of such an argument. The number

$$\frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!n!}$$

is integer for arbitrary n, which is not obvious from the formula. However, we know that this number enumerates regular bracket structures with n pairs of brackets and therefore cannot be a non-integer.

The representation of the Catalan numbers as triangulations of regular $(n+2)$-gons leads to the following statement.

Statement 5.8. *If $n+2$ is a power of a prime number, $n+2 = p^k$ and $n > 1$, then the Catalan number c_n is divisible by p.*

For example,

$$c_2 = 2 \equiv 0 \bmod 2; \quad c_5 = 42 \equiv 0 \bmod 7; \quad c_7 = 429 \equiv 0 \bmod 3.$$

Proof. The group $\mathbb{Z}_{n+2}$ of residues modulo $n+2$ acts by rotations on the set of triangulations of a regular $(n+2)$-gon. If $n > 1$, then the action has no fixed points, and the length of each orbit, being, by the Lagrange theorem from the group theory, a divisor of $n+2$, is divisible by p. Hence the total number of triangulations is divisible by p.

Similarly, interpreting the Catalan numbers in terms of regular bracket structures we obtain one more property.

Statement 5.9. *If n is a power of a prime number, $n = p^k$, then $c_n \equiv 2 \bmod p$.*

For example,

$$c_2 = 2 \equiv 2 \bmod 2; \quad c_3 = 5 \equiv 2 \bmod 3; \quad c_5 = 42 \equiv 2 \bmod 5.$$

Proof. The group $\mathbb{Z}_{2n}$ of residues modulo $2n$ acts on the set of regular bracket structures with n pairs of brackets in the following way. The generator of the group is represented by the minimal cyclic shift of the structure. Under such a shift

1) the leftmost bracket is deleted;

2) instead, the rightmost bracket is added;

3) the right bracket corresponding to the deleted leftmost bracket is replaced with the left bracket (which now corresponds to the rightmost bracket). All other brackets remain the same.

This transformation possesses no fixed points if $n > 1$. Exactly one of the orbits of this action has length 2. It consists of the bracket structures

$$\underbrace{()()\dots()}_{n \text{ pairs}} \quad \text{and} \quad (\underbrace{()\dots()}_{n-1 \text{ pairs}}).$$

The lengths of all other orbits are divisible by p, and the statement follows.

5.8. How to solve ordinary differential equations in generating functions

When deriving generating functions for the Bernoulli and Euler sides of the Bernoulli–Euler triangle we had to solve ordinary differential equations satisfied by these functions. The theorem below solves the problem of existence and uniqueness of a solution for a large class of ordinary differential equations containing both Equations (5.3) and (5.5).

Theorem 5.10. *Consider the ordinary differential equation*

$$f'(s) = F(s, f(s)) \tag{5.9}$$

with respect to the generating function $f(s)$, where $F = F(s,t)$ is a generating function in two variables, polynomial in t (i.e., having finite degree in t). Then for each f_0 Eq. (5.9) *possesses a unique solution with the initial condition $f(0) = f_0$.*

For Eq. (5.3), the function F is

$$F(s,t) = t^2 + 1,$$

while for Eq. (5.5) it is

$$F(s,t) = B(s)t.$$

Proof of the theorem. The proof follows our usual pattern of finding the coefficients of the unknown function f one by one. Let n be the degree of F with respect to t and let

$$\begin{aligned} F(s,t) = \quad & (F_{00} + F_{01}s + F_{02}s^2 + \dots) \\ & + (F_{01} + F_{11}s + F_{21}s^2 + \dots)t \\ & + \dots + \\ & + (F_{0n} + F_{1n}s + F_{2n}s^2 + \dots)t^n, \\ f(s) = f_0 + f_1 s + f_2 s^2 + \dots \end{aligned}$$

Equating the coefficients of s^0 on the left- and on the right-hand sides of Eq. (5.9) we obtain

$$f_1 = F_{00} + F_{01}f_0 + \dots + F_{0n}f_0^n.$$

Similarly, the equation for the coefficients of s^1 yields

$$2f_2 = F_{10} + F_{01}f_1 + F_{11}f_0 + \dots + F_{0n}f_0^{n-1}f_1 + F_{1n}f_0^n.$$

More generally, f_n is the root of the equation

$$nf_n = \dots, \tag{5.10}$$

where dots denote a polynomial in coefficients of F and the coefficients $f_0, f_1, \dots, f_{n-1}$ of f. For each $n > 0$ Eq. (5.10) has a unique solution, and the theorem follows.

5.9. Problems

5.1. The Chebyshev polynomial T_n is defined by the equality

$$\cos n\varphi = T_n(\cos\varphi).$$

Here are the first few Chebyshev polynomials:

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1, \quad T_3(x) = 4x^3 - 3x.$$

Prove the recurrence relation $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$ and deduce from it the identity

$$\sum_{n\geq 0} T_n(x)t^n = \frac{1-tx}{1-2tx+t^2}.$$

5.2. Prove that

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

5.3. Suppose the sequence $\{a_n\}$ starting with $1, 1, 2, 5, 17, 73, \ldots$ is defined by the conditions

$$a_0 = a_1 = 1,\ a_2 = 2, \quad a_{n+1} = (n+1)a_n - \binom{n}{2} a_{n-2},\ n > 2.$$

Prove that the exponential generating function $A(s)$ for this sequence satisfies the ordinary differential equation

$$(1-s)A'(s) = \left(1 - \frac{1}{2}s^2\right) A(s)$$

and is given by the formula

$$A(s) = (1-s)^{-1/2} e^{s/2+s^2/4}.$$

5.4. Prove that

a) the sum and the product of two Hurwitz functions are Hurwitz functions;

b) the derivative and the integral of a Hurwitz function are Hurwitz functions;

c) the result of substitution of a Hurwitz function into a Hurwitz function is a Hurwitz function;

d) if, in the assumptions of Theorem 5.10, the right-hand side F of the equation is a Hurwitz function and the number f_0 is an integer, then the solution f of this equation satisfying the initial condition $f(0) = f_0$ is a Hurwitz function.

5.5. Denote by $a_{n,k}$ the number of Dyck paths of length n bounding area k; $a_{2,1} = 1$, $a_{2,k} = 0$ for k even. Prove that

$$A(s,t) = \sum a_{n,k} s^n t^k = \frac{1}{1 - \frac{s^2 t}{1 - \frac{s^2 t^3}{1 - \frac{s^2 t^5}{1 - \dots}}}}.$$

5.6. Prove the following continued fractions expansions:

a)

$$B(s) = \frac{s}{1 - \frac{1 \cdot 2 s^2}{1 - \frac{2 \cdot 3 \cdot s^2}{1 - \dots - \frac{k(k+1)s^2}{1 - \dots}}}},$$

where $B(s)$ is the ordinary generating function for the Bernoulli side of the Bernoulli–Euler triangle;

b)

$$\sum_{n=0}^{\infty} (2n-1)!! s^{2n} = \frac{1}{1 - \frac{s^2}{1 - \frac{2s^2}{1 - \frac{3s^2}{1 - \dots}}}},$$

where $(2n-1)!! = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n-1)$;

c)

$$\sum_{n=0}^{\infty} I_n s^n = \frac{1}{1 - s - \frac{s^2}{1 - s - \frac{2s^2}{1 - s - \frac{3s^2}{1 - \dots}}}},$$

where I_n is the number of involutions (that is, permutations whose squares are the identity permutation) on a set of n elements, $I_1 = 1$, $I_2 = 2$, $I_3 = 4$, $I_4 = 10$, $\dots$;

d)

$$\sum_{n=0}^{\infty} (n+1)! s^n = \frac{1}{1 - 2s - \frac{1 \cdot 2 s^2}{1 - 4s - \frac{2 \cdot 3 s^2}{1 - 6s - \frac{3 \cdot 4 s^2}{1 - \dots}}}};$$

e)

$$\sum_{n=0}^{\infty} n! s^n = \frac{1}{1 - s - \frac{1^2 s^2}{1 - 3s - \frac{2^2 s^2}{1 - 5s - \frac{3^2 s^2}{1 - \dots}}}}.$$

5.7. Consider the hypergeometric function

$$h(s) = 1 + \left(\frac{1}{2}\right)^2 s + \left(\frac{1 \cdot 3}{2 \cdot 4}\right)^2 s^2 + \left(\frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}\right)^2 s^3 + \dots$$

a) Prove that

$$s(1-s)h''(s) + (1-2s)h'(s) - \frac{1}{4}h(s) = 0.$$

b) Find the asymptotics of the coefficients of h.

5.8. Prove that the power series

$$y(s) = 1 + \frac{2s}{1!} + \frac{6s^2}{2!} + \frac{20s^3}{3!} + \dots + \binom{2n}{n}\frac{s^n}{n!} + \dots$$

satisfies the differential equation

$$sy'' + (1-4s)y' - 2y = 0.$$

5.9. Find the first five coefficients of the (unique) solution $y = y(x)$ of the following differential equation:

$$y' = 2 + 3x - 2y + x^2 + x^2 y.$$

5.10. Prove that the function $y(s) = \frac{\arcsin s}{(1-s^2)^{1/2}}$ satisfies the differential equation

$$(1-s^2)y' - sy = 1$$

and find the sequence of its coefficients.

5.11. Write out a differential equation satisfied by the function

$$e^{s^2} \int e^{-s^2/2}$$

and find the sequence of its coefficients.

5.12. Write out the Bernoulli–Euler triangle modulo 2.

5.13. Prove that the number of indecomposable meanders (see Problem 4.5) of order p^m, where p is a prime number and $m \geq 1$, is congruent to 2 modulo p.

Chapter 6

Partitions and Decompositions

6.1. Partitions and decompositions

In the process of solving the lucky tickets problem in Chapter 1 we have already studied the question about the number of representations of a positive integer n as a sum of k integers. Put aside the restriction on the value of a summand (in the lucky tickets problem, the summands were figures, and they could not be greater than 9). Let us find the number of ways to represent n as a sum of non-negative integers.

We consider two representations

$$n = a_1 + \cdots + a_k = b_1 + \cdots + b_k$$

as distinct if $a_i \neq b_i$ at least for one i, $1 \leq i \leq k$. We call such a representation of n a *decomposition*.

Statement 6.1. *The number of distinct decompositions of n into a sum of k non-negative integers is $\binom{n+k-1}{k-1}$.*

Proof. Let us think of n as of a tuple of n indistinguishable balls on the line. Associate to each decomposition of n with k summands a distribution of $k-1$ sticks on the intervals between the balls. The element a_i of the decomposition coincides with the number of balls between the $(i-1)$th and the ith sticks. Together there are $n+k-1$ objects — balls and sticks. Conversely, having $n+k-1$ objects, there are $\binom{n+k-1}{k-1}$ possibilities to appoint $k-1$ of them sticks. This proves the proposition.

It is also easy to deduce the generating function for the numbers of decompositions. In fact, we have already done this in Chapter 1.

Statement 6.2. *The generating function for the numbers of decompositions into k summands is*

$$B_k(s) = (1-s)^{-k}.$$

The computation of partitions of n is more complicated. A *partition* is an equivalence class of decompositions without zero summands. Two decompositions are considered to be equivalent if one of them can be obtained from the other one by a permutation of the summands.

Here are all partitions of small numbers:

$$\begin{array}{ll}
n=1 & 1 \\
n=2 & 2 = 1+1 \\
n=3 & 3 = 2+1 = 1+1+1 \\
n=4 & 4 = 3+1 = 2+2 = 2+1+1 = 1+1+1+1 \\
n=5 & 5 = 4+1 = 3+2 = 3+1+1 \\
 & \quad = 2+2+1 = 2+1+1+1 = 1+1+1+1+1
\end{array}$$

Pay attention to the fact that each partition is written in the decreasing order of the parts: the comparison of two partitions written in this form is an easy task.

Denote the number of partitions of n by p_n; then we obtain the following table of the first few elements of the sequence p_n:

n	0	1	2	3	4	5	6	7	8
p_n	1	1	2	3	5	7	11	15	22

The problem at hand now is to find the generating function for the sequence p_n. To do this, let us first write out the generating functions for the numbers of partitions with restrictions on the size of the parts. Let $P_1(s)$ denote the generating function for the number of partitions of n into parts equal to 1. Obviously, there is exactly one such partition for each n and we have

$$P_1(s) = 1 + s + s^2 + s^3 + \cdots = \frac{1}{1-s}.$$

The number of partitions of n into parts equal to 2 is one for even n and 0 for odd ones; therefore,

$$P_2(s) = 1 + s^2 + s^4 + s^6 + \cdots = \frac{1}{1-s^2}.$$

Hence, the number of partitions of n into parts not exceeding 2 is described by the generating function

$$P_1(s)P_2(s) = \frac{1}{(1-s)(1-s^2)}.$$

Similarly, the number of partitions of n into parts equal to 3 is described by the generating function $P_3(s) = 1/(1-s^3)$, while partitions into parts not greater than 3 are enumerated by the generating function

$$P_1(s)P_2(s)P_3(s) = \frac{1}{(1-s)(1-s^2)(1-s^3)}.$$

Repeating this argument we arrive at the following statement.

Theorem 6.3 (Euler). *The generating function for the number of partitions of n has the form*

$$P(s) = \frac{1}{(1-s)(1-s^2)(1-s^3)(1-s^4)\ldots}. \tag{6.1}$$

This theorem makes sense if we can interpret the infinite product in the denominator of the right-hand side of Eq. (6.1). This product must be a formal power series

$$Q(s) = q_0 + q_1 s + q_2 s^2 + \cdots = (1-s)(1-s^2)(1-s^3)\ldots \tag{6.2}$$

In order to say what are the coefficients $q_0, q_1, q_2, \ldots$ of this infinite product, let us first look at the finite products:

$$
\begin{aligned}
1-s &= \mathbf{1-s} \\
(1-s)(1-s^2) &= \mathbf{1-s-s^2} + s^3 \\
(1-s)(1-s^2)(1-s^3) &= \mathbf{1-s-s^2} \qquad + s^4 + \ s^5 - s^6 \\
(1-s)\dots(1-s^4) &= \mathbf{1-s-s^2} \qquad\qquad + 2s^5 + \dots \\
(1-s)\dots(1-s^5) &= \mathbf{1-s-s^2} \qquad\qquad + \ \boldsymbol{s^5} + \dots
\end{aligned}
$$

We see that the coefficients in these finite sequences "stabilize", i.e., they remain unchanged starting from some moment (the stabilized terms of the expansions are shown in bold). There is nothing strange in this fact: the multiplication by $1-s^k$ produces no changes in the coefficients of the polynomail at degrees less than k. Therefore, we may simply set q_k to be the coefficient of s^k in the polynomial $(1-s)(1-s^2)\dots(1-s^k)$.

Now we are able to write out the generating functions for partitions satisfying various additional restrictions.

For example, the number of partitions into distinct parts is given by the generating function

$$(1+s)(1+s^2)(1+s^3)\dots,$$

partitions into distinct odd parts are described by the generating function

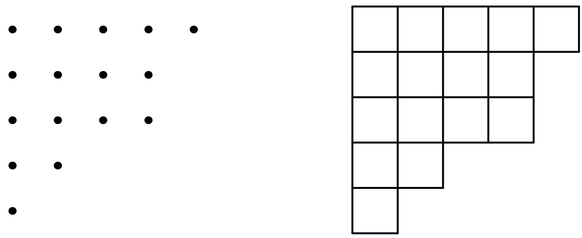
$$(1+s)(1+s^3)(1+s^5)\dots,$$

while partitions into arbitrary odd parts are enumerated by the generating function

$$\frac{1}{(1-s)(1-s^3)(1-s^5)\dots},$$

and so on.

Partitions are closely related to the algebra $\mathbb{C}[x_1, x_2, x_3, \dots]$ of polynomials in the infinite number of variables. We assign to the variable x_i the weight i and assume that the weight of a monomial is the sum of the weights of the variables. Let us count the number of monomials of weight n, i.e., the dimension of the space of homogeneous polynomials of weight n.



a) b)

Figure 1. a) Ferrer's and b) Young's diagrams

For $n = 1$ there is one such monomial, namely, x_1. For $n = 2$ there are two monomials of weight n, namely, x_1^2 and x_2. The number of monomials of weight 3 is three: x_1^3, x_1x_2 x_3. More generally, the number of monomials of weight n is p_n. Indeed, one can associate to a monomial of weight n a partition of n according to the following rule: the number of parts i in the partition coincides with the power of x_i in the monomial. Clearly, this correspondence is one-to-one.

Here is a useful geometric interpretation of partitions. It is convenient to represent a partition as a *Ferrer diagram* or a *Young diagram* (see Fig. 1). The diagrams in this figure correspond to the partition $5+4+4+2+1$ of 16. The number of elements in the ith row of the diagram coincides with the ith part of the partition.

Ferrer's and Young's diagrams provide a convenient tool for proving various properties of partitions. For example, there is a natural involution on the set of Young diagrams, the reflection with respect to the diagonal. Some diagrams remain fixed under this involution (see Fig. 2). We call such diagrams (and the corresponding partitions) *symmetric*.

Let us prove the following property of symmetric partitions.

Statement 6.4. *The number of symmetric partitions of n coincides with the number of its partitions into pairwise distinct odd parts.*

Proof. To prove the statement, let us associate to each symmetric Young diagram the diagram consisting of the "central hooks" in it (see Fig. 2 b)). The number of cells in each central hook of a symmetric

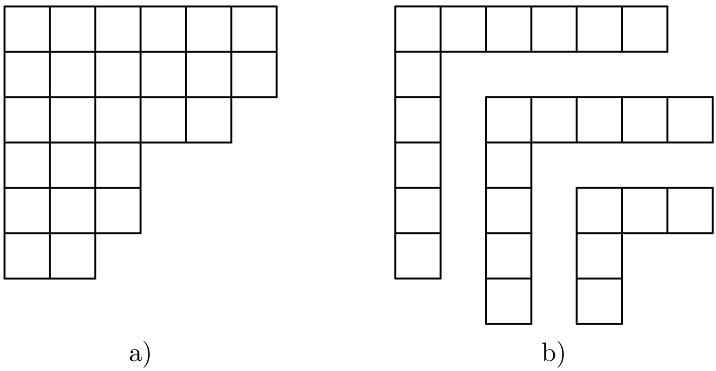


Figure 2. a) A symmetric Young diagram; b) the central hooks in the diagram

diagram is odd, and these numbers are pairwise distinct. Conversely, given a diagram consisting of distinct odd rows, we may take each row, "break" it at the middle and construct a new diagram of the resulting hooks.

6.2. The Euler identity

The generating function Q defined by Eq. (6.2) is a very interesting one. Euler continued counting its coefficients and obtained

$$Q(s) = 1 - s - s^2 + s^5 + s^7 - s^{12} - s^{15} + s^{22} + s^{26} - s^{35} - s^{40} + s^{51} + s^{57} - s^{70} - s^{77} + s^{92} + s^{100} - \dots$$

We see that the coefficients on the right are only ones or negative ones and zeroes. The non-zero coefficients are situated at rather specific places, and the signs at ones alternate in pairs. These observations led Euler to a conjecture, which we state here as a theorem.

Theorem 6.5 (the Euler identity)**.**

$$Q(s) = 1 + \sum_{k=1}^{\infty} (-1)^k \left(s^{\frac{3k^2-k}{2}} + s^{\frac{3k^2+k}{2}} \right).$$

Proof. After removing parentheses, the series

$$Q(s) = (1-s)(1-s^2)(1-s^3)\dots$$

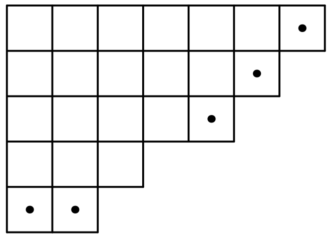


Figure 3. The lower row and the side diagonal of a diagram

contains the same terms as the generating series for the number of partitions into distinct parts

$$(1+s)(1+s^2)(1+s^3)\dots$$

However, some terms enter this series with the positive sign, while the others with the negative sign. The positive terms correspond to partitions into an even number of parts, while the negative terms correspond to partitions into an odd number of parts. We are going to prove that the number of partitions of n into an odd number of parts coincides with that into an even number of parts for all values of n but some special ones.

Let us represent each partition by its Young diagram. The lower row and the "side diagonal" of the diagram (see Fig. 3) will play an essential role in the proof.

Let l be the length of the lower row, d the length of the side diagonal, and let k be the number of rows in the diagram, that is, the number of parts in the partition. Define a mapping from the set of diagrams with rows of pairwise distinct length into itself in the following way:

- if $l < d$, then we cut off the lower row and glue it to the diagram along the side diagonal;
- if $l = d < k$, then we do the same thing;
- if $l > d$ and $k > l$, then, conversely, we cut off the side diagonal and glue it below the lower side.

We do nothing with all other (exceptional) tables.

This mapping switches the parity of the number of rows in the diagram, that is, the number of parts in the partitions for all partitions but the exceptional ones. Therefore, if there are no exceptional diagrams with n cells, then the coefficient of s^n in $Q(s)$ is zero.

The exceptional diagrams are selected by the conditions

$$\text{either } k = l = d \text{ or } k = d,\ l = k + 1.$$

In the first case we have

$$n = k + (k + 1) + (k + 2) + \cdots + (2k - 1) = \frac{3k^2 - k}{2};$$

and in the second

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

In both cases the exceptional diagram is unique. The Euler identity is proved.

Statement 6.6. *We have*

$$p_n = p_{n-1} + p_{n-2} - p_{n-5} - p_{n-7} + p_{n-12} + p_{n-15} - \dots$$

Indeed, this is an immediate corollary of the identity

$$P(s)Q(s) = 1.$$

The recurrence relation of Statement 6.6 found the following elegant interpretation as “D. B. Fuchs’ ruler”[1]:

> This formula allows one to generate effectively a rather long table of the numbers p_n. Here is a practical approach. Take a sheet of graph paper. Cut off a strip 3–4 squares wide along its longer side. Lay the strip on the table vertically and insert some mark, say a star, in the lowest cell. Then, moving up along the strip, insert the sign + in the first and second rows, the sign − in the fifth and seventh rows, the sign + in the twelfth and fifteenth rows, and so on, up to the upper side of the strip. Place the remaining part of the sheet vertically also and draw a vertical line from the upper

[1]Kvant. 1981. no. 8. p. 15.

to the lower side of the sheet, at the distance 10–15 cells from the left. Inscribe in the cells to the left of the line the values of p_n you already know top-down, starting with p_0: 1, 1, 2, 3, 5, 7. In order to find the next value, place the strip against the line so as to put the star against the first empty cell. Then subtract the sum of numbers against the $-$ signs from the sum of those against the $+$ signs. Inscribe the result in the next cell. This is the next value of the function p_n. Shift the strip one square down and repeat the procedure, and so on. In several minutes you will obtain a column of the numbers p_n of height equal to that of your sheet.

6.3. Set partitions and continued fractions

We have seen in Sec. 2.5 that by fixing the vertices of a polygon we can simplify enumeration of its triangulations. Similarly, partitions of sets are enumerated easier than partitions of numbers.

Consider the set $N_n = \{1, 2, \dots, n\}$ of positive integers between 1 and n. A *partition* of N_n is a representation of this set as a union of non-empty disjoint subsets. For example, the set N_3 admits five partitions:

$$\{\{1,2,3\}\},\ \{\{1,2\},\{3\}\},\ \{\{1,3\},\{2\}\},\ \{\{2,3\},\{1\}\},\ \{\{1\},\{2\},\{3\}\}.$$

Denote the number of partitions of N_n by $\tilde{p}_n$. We are going to study the generating function

$$\widetilde{P}(s) = \tilde{p}_0 + \tilde{p}_1 s + \tilde{p}_2 s^2 + \dots$$

(we set, by definition, $\tilde{p}_0 = 1$).

There is a natural way to associate to each partition of N_n a partition of the number n. In order to do this, it suffices to represent n as the sum of the cardinalities of the parts in the given partition of N_n. It is also easy to count the number of partitions of N_n corresponding to a given partition

$$n = n_1 + n_2 + \dots + n_k$$

of n. There are $\binom{n}{n_1}$ ways to choose the elements of the first part, $\binom{n-n_1}{n_2}$ ways to choose the elements of the second part after the first part is already chosen, and so on. All in all, there are

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\dots\binom{n-n_1-\dots-n_{k-1}}{n_k}$$
$$= \frac{n!}{n_1!(n-n_1)!}\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\dots\frac{(n-n_1-\dots-n_{k-1})!}{n_k!0!}$$
$$= \frac{n!}{n_1!n_2!\dots n_k!} = \binom{n}{n_1\ n_2\ \dots\ n_k}$$

ways to split the elements of N_n into parts having n_1, n_2, ..., n_k elements. The resulting expression is called the *multinomial coefficient.* This notion generalizes that of a binomial coefficient. It is easy to see that the multinomial coefficient is the coefficient of the monomial $x_1^{n_1}x_2^{n_2}\dots x_k^{n_k}$ in the expansion of $(x_1+\dots+x_k)^n$:

$$(x_1+\dots+x_k)^n = \sum_{\substack{n_1,\dots,n_k=0\\ n_1+\dots+n_k=n}}^{n} \binom{n}{n_1\ \dots\ n_k} x_1^{n_1}\dots x_k^{n_k}.$$

However, the number of partitions of N_n corresponding to a given partition of n is not exactly the multinomial coefficient. The reason is that the parts of the partition having the same number of elements can be permuted. Therefore, the correct answer is

$$\frac{1}{m_1!\dots m_n!}\binom{n}{n_1\ \dots\ n_k},$$

where m_i is the number of parts equal to i.

Now associate to each partition N_n into subsets a path in the Motzkin triangle according to the following rule. Take the part in the partition containing the element i. The ith vector in the path is horizontal if either the corresponding part consists of the single element i, or i is neither the minimal, nor the maximal element in this part. The ith vector of the path is the raising vector $(1,1)$ if i is the minimal, and it is the descending vector $(1,-1)$ if i is the maximal element in this set. The starting point of the path is, as usual, the origin $(0,0)$. A partition of the set N_{10} and the corresponding path in the Motzkin triangle are shown in Fig. 4.

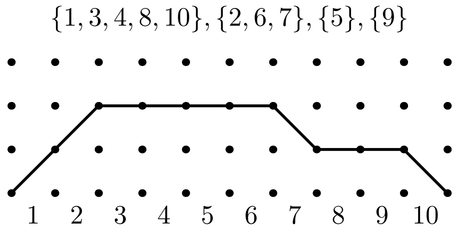


Figure 4. The path in the Motzkin triangle corresponding to a given partition

It is clear that the path associated to a partition is indeed a Motzkin path: it lies in the positive quadrant and ends at the height 0. Indeed, for each m the number of maximal elements among the first m elements of N_n cannot exceed the number of minimal elements among them, and for $m = n$ the numbers of minimal and maximal elements coincide.

Let us count the number of partitions corresponding to a given path. Suppose the beginning i segments end at the height j and suppose the first j elements of N_n are already split into subsets. If the $(j+1)$th vector of the path is the raising vector $(1, 1)$, then $j+1$ is the minimum in the new part of the partition, and there are no other possibilities. Therefore, the multiplicity of the corresponding edge in the triangle is 1. If it is the horizontal vector, then the corresponding element can either enter one of the existing subsets (there are exactly j possibilities since the maximal element is not yet fixed in j subsets), or form a subset by itself. Therefore, the multiplicity of a horizontal vector at the height j is $j+1$. Finally, the multiplicity of a descending vector $(1, -1)$ is j since the corresponding element can be the maximal element in one of the j subsets. This distribution of multiplicities in the Motzkin triangle is shown in Fig. 5.

Hence, the following statement is true.

Theorem 6.7. *The number $\tilde{p}_n$ of partitions of the set N_n into non-empty subsets is equal to the number of paths of length n in the Motzkin triangle with multiplicities shown in Fig. 5.*

This theorem together with Theorem 5.5 immediately imply the following statement.

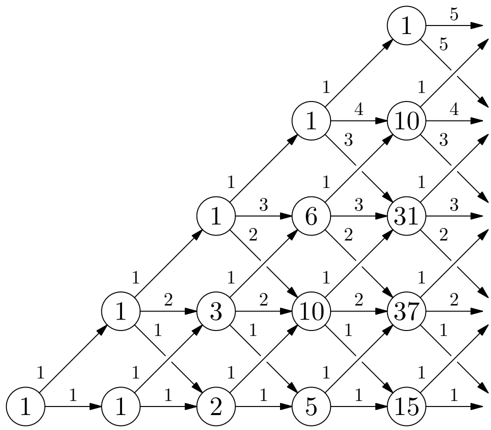


Figure 5. The distribution of multiplicities corresponding to the generating function for set partitions

Corollary 6.8. *The generating function* $\widetilde{P}_n(s)$ *for the number of partitions of* N_n *admits the following continued fraction expansion:*

$$\widetilde{P}_n(s) = \cfrac{1}{1 - s - \cfrac{s^2}{1-2s-\cfrac{2s^2}{1-3s-\cfrac{3s^2}{1-\dots}}}}.$$

6.4. Problems

6.1. In how many ways can one change a rouble (=100 copecks) with 1, 5, 10, and 50 kopecks' coins?

6.2. Count the number of ways to measure 78 grams having weights of 1, 1, 2, 5, 10, 10, 20, 50 grams by means of

a) a one-pan balance;

b) two-pan balance?

(Applying two distinct weights of the same size we obtain two distinct ways of measuring.)

6.3. Each positive integer has a unique presentation in decimal notation. Deduce from this fact the following identity:

$$(1 + s + s^2 + \dots + s^9)(1 + s^{10} + \dots + s^{90}) \dots = \frac{1}{1-s}.$$

6.4. Prove that

$$(1+s)(1+s^2)(1+s^3)\cdots = \frac{1}{(1-s)(1-s^3)(1-s^5)\ldots}.$$

6.5. Prove that each positive integer has as many partitions into distinct positive summands as into odd (may be coinciding) summands.

6.6. Prove that the number of partitions of n such that only odd parts are allowed to be repeated coincides with the number of partitions of n such that each part is repeated not more than 3 times.

6.7. Prove that there are $2^{n-1}-1$ ways to represent a positive integer n as a sum of smaller positive integers, if we consider two representations with different order of summands as distinct. For example, $n=4$ has seven representations:

$$\begin{aligned} 4 &= 3+1=1+3=2+2=2+1+1 \\ &= 1+2+1=1+1+2=1+1+1+1. \end{aligned}$$

6.8. Find the generating function for the number of symmetric partitions.

6.9. Consider the ring of polynomials in an infinite number of weighted variables, and suppose the number of variables with given weight i is finite for each i. Denote this number by q_i. Write out the generating function for the sequence of dimensions of spaces of homogeneous polynomials of weight n.

6.10. Denote by σ_n the sum of divisors of a positive integer n (including 1 and n itself); for example, $\sigma_6 = 1+2+3+6 = 12$. Let $\Sigma(s)$ be the generating function for the sequence σ_n,

$$\Sigma(s) = s + 3s^2 + 4s^3 + 7s^4 + 6s^5 + 12s^6 + \ldots$$

a) Prove that

$$\Sigma(s)P(s) = sP'(s),$$

where $P(s)$ is the generating function for the number of partitions.

b) Deduce the recurrence relation for σ_n from this identity.

6.11. Prove that the sequence p_n is increasing and estimate the rate of its growth.

6.12. Show that the generating function for the sequence of numbers of Young diagrams with perimeter $2n$ is $\frac{x^2}{1-2x}$.

6.13. Prove that

$$\sum_{n_1,n_2,\dots,n_k\geq 1} \min(n_1,n_2,\dots,n_k)t_1^{n_1}t_2^{n_2}\dots t_k^{n_k} = \frac{t_1t_2\dots t_k}{(1-t_1)(1-t_2)\dots(1-t_k)(1-t_1t_2\dots t_k)}.$$

6.14. Prove that

$$\Big(1+\sum_{n=1}^{\infty}\frac{t^n}{(1-q)\dots(1-q^n)}\Big)^{-1} = 1+\sum_{n=1}^{\infty}\frac{(-1)^n q^{\binom{n}{2}}t^n}{(1-q)\dots(1-q^n)}.$$

Chapter 7

Dirichlet Generating Functions and the Inclusion-Exclusion Principle

7.1. The inclusion-exclusion principle

We start with a very simple general theorem of formal logic. Let B be a finite set each of whose elements can possess some of the properties $c_1, \dots, c_m$. Denote by $N(c_i)$, $1 \le i \le m$, the number of elements of the set B possessing the property c_i, by $N(c_i, c_j)$, $i \ne j$, the number of elements of the set B possessing both properties c_i, c_j, and so on. Also, let $N(1)$ denote the total number of elements in B.

Theorem 7.1 (inclusion-exclusion principle)**.** *The number of elements in B possessing none of the properties c_i, $i = 1, \dots, m$, is*

$$N(1) - N(c_1) - \dots - N(c_m) + N(c_1, c_2) + \dots - N(c_1, c_2, c_3) - \dots$$

Proof. Split all elements in B into disjoint groups: $B = B_0 \sqcup B_1 \sqcup \dots \sqcup B_m$, where B_l is the subset of elements possessing exactly l

properties. Consider the expressions

$$N(1),$$
$$N(1) - N(c_1) - \dots - N(c_m),$$
$$N(1) - N(c_1) - \dots - N(c_m) + N(c_1, c_2) + \dots + N(c_{m-1}, c_m),$$
$$\dots\dots\dots\dots\dots$$

one by one. We associate to each of these expressions a distribution of integers on the sets B_l. The first expression associates to each subset the number 1: it counts each of the elements exactly once. The second expression associates to the set B_l the number $1-l$ since each element of B_l has been taken into account, when subtracted, exactly l times. The third expression assigns to the set B_l the multiplicity $1-l+\binom{l}{2}$, and so on. Hence, the transition from the lth expression to the $l+1$th expression does not change the multiplicities of the sets $B_0, \dots, B_l$. These stable multiplicities are

$$\binom{l}{0} - \binom{l}{1} + \binom{l}{2} - \dots + (-1)^l \binom{l}{l},$$

which is zero for all l but $l = 0$, and the theorem follows.

The following simple mnemonics allows one to remember the inclusion-exclusion principle easily. Associate 1 to the objects possessing all properties, then $1 - c_i$ will denote objects without property c_i. Then the expression for the objects possessing none of the properties $c_1, \dots, c_m$ will be

$$(1 - c_1)(1 - c_2) \dots (1 - c_m),$$

which after erasing the brackets gives

$$(1-c_1)(1-c_2)\dots(1-c_m) = 1-c_1-\dots-c_m+c_1c_2+\dots-c_1c_2c_3-\dots$$

Now let us apply the inclusion-exclusion principle to the lucky tickets problem of Sec. 1.1. Note first that the number of lucky tickets coincides with the number of tickets having the sum of digits 27. Indeed, suppose a ticket $a_1b_1c_1a_2b_2c_2$ is lucky. Then the sum of the digits in the ticket $a_1b_1c_1(9-a_2)(9-b_2)(9-c_2)$ is 27. Obviously, this correspondence is one-to-one.

Now consider the set of all distributions of non-negative integers with sum 27 in six positions and introduce the following six properties of such distributions. The property c_i states that the number at the ith position is at least 10. The number of lucky tickets is exactly the number of distributions possessing none of the properties $c_1, \dots, c_6$.

The number $N(1)$ of distributions of non-negative integers with sum 27 at six positions is $\binom{32}{5}$. The number $N(c_i)$ of distributions possessing the property c_i is the same for all $i = 1, \dots, 6$ and equals $\binom{22}{5}$. Indeed, we may fix the number 10 at the ith position and then distribute the complementary sum 17 among the six positions arbitrarily.

Similarly, the number of distributions possessing simultaneously two properties c_i and c_j is $\binom{12}{5}$: we may fix the number 10 at the ith and jth positions and distribute the complementary sum 7 arbitrarily at the six positions. The number of distributions possessing simultaneously three or more properties is zero since the total sum of the numbers is less than 30. Hence, the total number of distributions possessing none of the properties c_i is given by the following proposition.

Statement 7.2. *The number of lucky tickets is*

$$\binom{32}{5} - 6\binom{22}{5} + 15\binom{12}{5}.$$

Using the inclusion-exclusion principle, we solve one more problem having many applications.

A permutation π of the elements of the set $\{1, 2, \dots, n\}$ is called a *disorder* if $\pi(k) \neq k$ for all $k = 1, \dots, n$. Let d_n denote the number of disorders on the n-element set. Here is the beginning of the table of the numbers of disorders:

n	0	1	2	3	4
d_n	1	0	1	2	9

To enumerate disorders introduce n properties of permutations on the n-element set. The property c_i states that the permutation fixes the element i. The total number of permutations is $n!$. The number of permutations possessing the property c_i is $(n-1)!$: the ith element of the set is fixed, while other $n-1$ elements undergo an arbitrary permutation. The number of elements possessing two properties c_i and c_j is $(n-2)!$: two elements of the set are fixed, while the permutation of the other $n-2$ elements is arbitrary. More generally, the number of permutations possessing m properties equals $(n-m)!$. Hence, we arrive at the following formula.

Statement 7.3. *The number of disorders on an n-element set is*

$$\begin{aligned} d_n &= \binom{n}{0}n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! - \dots \\ &= n!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots + (-1)^n\frac{1}{n!}\right). \end{aligned}$$

As is well known, as $n \to \infty$ this expression tends to e^{-1}. Thus, disorders form approximately $1/e$-part of all permutations.

7.2. Dirichlet generating functions and operations with them

All generating functions we considered up to now are power series. However, in multiplicative number theory another type of series, Dirichlet functions, are also useful. The most important of them is the *Riemann zeta function*

$$\zeta(s) = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \dots \tag{7.1}$$

A general *Dirichlet generating function* corresponding to a sequence $a_1, a_2, a_3, \dots$, has the form

$$\frac{a_1}{1^s} + \frac{a_2}{2^s} + \frac{a_3}{3^s} + \dots$$

The Riemann zeta function corresponds to the sequence $1, 1, 1, \dots$. It plays the same role for the Dirichlet generating functions as the geometric series for the ordinary and the exponent for the exponential generating functions. Pay attention to the fact that the numbering of coefficients in Dirichlet generating functions starts with 1, not with 0, as for ordinary or exponential ones.

Dirichlet generating functions are introduced because of their behavior under the multiplication: the product of two functions $A(s) =$

$\sum a_n n^{-s}$ and $B(s) = \sum b_n n^{-s}$ is the function

$$A(s)B(s) = \frac{a_1b_1}{1^s} + \frac{a_1b_2 + a_2b_1}{2^s} + \frac{a_1b_3 + a_3b_1}{3^s} + \frac{a_1b_4 + a_2b_2 + a_4b_1}{4^s} + \dots = \sum_n \frac{\sum\limits_{kl=n} a_k b_l}{n^s},$$

where the internal summation is carried over all decompositions of n into a product of two ordered factors. Hence, Dirichlet generating functions reflect the multiplicative structure of integers. Note that the addition of two such functions corresponds to the usual termwise addition of sequences.

The function $1 = 1^{-s}$ plays the role of the unit under the multiplication of Dirichlet generating functions. Any Dirichlet generating function $A(s)$ with a non-zero constant term $a_1 \neq 0$, is invertible: there is a function $B(s)$ such that $A(s)B(s) = 1$. Let us construct the inverse of the Riemann zeta function.

Theorem 7.4. *The inverse function of the Riemann zeta function is*

$$M(s) = \frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu_n}{n^s},$$

where

$$\mu_n = \begin{cases} (-1)^{t_n} & \text{where } t_n \text{ is the number of prime divisors of } n, \\ & \text{if there are no repeating divisors} \\ & \text{in the prime factorization of } n; \\ 0 & \text{otherwise.} \end{cases}$$

The sequence μ_n is called the *Möbius sequence*, and the function $M(s)$ the *Möbius function.*

Proof. To prove the theorem, let us multiply $\zeta(s)$ by $M(s)$. The coefficient of n^{-s}, $n > 1$, in the product is

$$\binom{t_n}{0} - \binom{t_n}{1} + \dots + (-1)^{t_n}\binom{t_n}{t_n} = 0.$$

Indeed, suppose the factorization of n into the product of positive powers of distinct prime numbers has the form $n = p_1^{k_1} \dots p_t^{k_t}$, where

$t = t_n$. Then the coefficient of m^{-s} in $M(s)$ contributes to the coefficient of n^{-s} in the product if and only if m is a product of prime numbers forming a subset of the set $p_1, \dots, p_t$. The number of such subsets containing k elements is $\binom{t}{k}$, and the sign of the coefficient of m^{-s} is $(-1)^k$.

The theorem is proved.

This theorem immediately implies the following statement.

Corollary 7.5. *Let f_n, g_n be two sequences such that*

$$f_n = \sum_{t|n} g_t, \tag{7.2}$$

where the summation is carried over all divisors t of n. Then the elements g_n can be expressed in terms of the elements f_n in the following way:

$$g_n = \sum_{t|n} \mu_{n/t} f_t. \tag{7.3}$$

Proof. Indeed, Eq. (7.2) means that

$$F(s) = \zeta(s)G(s),$$

where $F(s)$ (resp., $G(s)$) is the Dirichlet generating function for the sequence f_n (resp., g_n). Multiplying both parts of the last equation by $M(s)$ we obtain

$$M(s)F(s) = M(s)\zeta(s)G(s) = G(s),$$

which is exactly Eq. (7.3). The corollary is proved.

Since any positive integer admits a unique factorization into a product of powers of distinct prime numbers, we obtain a representation of the zeta function as an infinite product (and, therefore, one more representation of the Möbius function):

Statement 7.6. *We have*

$$\zeta(s) = \frac{1}{1-2^{-s}} \frac{1}{1-3^{-s}} \frac{1}{1-5^{-s}} \frac{1}{1-7^{-s}} \dots,$$
$$M(s) = \left(1-2^{-s}\right)\left(1-3^{-s}\right)\left(1-5^{-s}\right)\left(1-7^{-s}\right)\dots,$$

where the product is taken over all prime numbers.

7.3. Möbius inversion

The formula for the sum of a geometric series, the inclusion-exclusion formula of Theorem 7.1, Theorem 7.4 and the Euler inversion (6.1) all are manifestations of a simple general principle, called the Möbius inversion principle. This principle allows one to find the inverse function to a zeta function in a variety of situations.

Namely, let $s_1, s_2, \dots$ be a set (may be infinite) of variables, and suppose we consider the algebra of formal power series in these variables. Define the *zeta function* of this algebra as the sum of all monomials in it, taken with coefficient 1. Thus, the geometric series

$$1 + s + s^2 + s^3 + \dots$$

is the zeta function of the algebra of power series in a single variable s. The function inverse to the zeta function is the *Möbius function* of the algebra. It is a sum of some monomials taken with some coefficients, which we have already computed in Theorem 7.4.

Theorem 7.7. *The coefficient of the monomial* $s_{i_1}^{n_1} \dots s_{i_m}^{n_m}$ *in the Möbius function of the algebra of formal power series in variables* $s_1, s_2, \dots$ *is* 0 *if any of the variables enters the monomial with the degree at least* 2 (*i.e.,* $n_i > 1$ *for some* i) *and it equals* $(-1)^m$ *if all* m *variables in the monomial have degree* 1*.*

Proof. One may prove the theorem repeating the proof of Theorem 7.4 almost word for word. We choose another way, however. In fact, we know the inverse function for the zeta function explicitly. Indeed, the zeta function itself is the product of zeta functions in each of the variables $s_1, s_2, \dots$. Therefore, it is the product of the geometric series

$$(1 + s_1 + s_1^2 + s_1^3 + \dots)(1 + s_2 + s_2^2 + s_2^3 + \dots) \dots$$

(Note that the coefficient of each monomial in this product is a sum of *finitely* many *finite* products.) Therefore, the Möbius function, which is inverse to the zeta function, is simply the product

$$(1 - s_1)(1 - s_2)(1 - s_3) \dots,$$

and the statement of the theorem follows immediately.

Let us look at applications of this theorem. In order to deduce from it Theorem 7.4 consider the set of variables $s_2, s_3, s_5, s_7, \dots$ (to each prime number there corresponds a single variable, indexed by this number). Then the algebra of Dirichlet generating functions is isomorphic to the algebra of power series in the chosen set of variables: under this isomorphism the element n^{-s}, with $n = p_1^{k_1} \dots p_m^{k_m}$ being the factorization into the prime factors, corresponds to the monomial $s_{p_1}^{k_1} \dots s_{p_m}^{k_m}$ of the algebra of power series. It is easy to see that this mapping indeed extends linearly to an algebra isomorphism. Now Theorem 7.4 follows immediately from Theorem 7.7 (cf. Statement 7.6).

The inclusion-exclusion principle can be deduced from Theorem 7.7 in the following way. Consider the algebra of polynomials in variables $s_1, \dots, s_n$ (each variable corresponds to a property under study) *truncated at degree two.* This means that each monomial containing a variable of degree at least two is considered to be 0. A monomial $s_{i_1} \dots s_{i_m}$ in this algebra is identified with the subset $\{i_1, \dots, i_m\}$ of the set $\{1, \dots, n\}$. The inclusion-exclusion principle is nothing but the formula for the Möbius function in this algebra.

The inversion formula for the generating function enumerating partitions also can be deduced easily. Associate to each partition $n = n_1 + \dots + n_m$ the monomial $s_{n_1} s_{n_2} \dots s_{n_m}$ (if some parts in the partition occur more than once, then the degree of the corresponding variable in the monomial is equal to the number of the parts). As we already know, the Möbius function in this algebra is

$$(1 - s_1)(1 - s_2)(1 - s_3) \dots$$

Making the substitution $s_n = s^n$ we transform the zeta function of the algebra into the generating function enumerating partitions (indeed, the coefficient of s^n after the substitution is exactly the number of all partitions of n). The inverse function is transformed into the function

$$(1 - s)(1 - s^2)(1 - s^3) \dots,$$

and Eq. (6.2) follows.

7.4. Multiplicative sequences

There are other Dirichlet generating functions, different from the zeta and the Möbius function, that also play important roles in number theory. The most useful of them are the functions corresponding to multiplicative number sequences.

Definition 7.8. A sequence $a_1,\ a_2,\ a_3,\dots$ is called *multiplicative* if for any coprime positive integers m, n the equality $a_m a_n = a_{m\cdot n}$ holds.

Note that if $a_1 = 0$ in a multiplicative sequence, then this sequence consists of zeroes. Indeed, $a_n = a_{1\cdot n} = a_1 a_n = 0$ for any positive integer n. The same argument shows that if $a_1 \neq 0$, then $a_1 = 1$. In what follows we will consider only non-zero multiplicative sequences.

The sequence $1,\ 0,\ 0,\ 0,\dots$ is multiplicative. The sequence consisting only of 1's is also multiplicative. The Möbius sequence is also multiplicative, which follows, for example, from Theorem 7.4. Let us give several examples more.

Example 7.9. Denote by τ_n the number of divisors of n. Obviously, the Dirichlet generating function for the sequence τ_n is

$$\tau(s) = \frac{\tau_1}{1^s} + \frac{\tau_2}{2^s} + \cdots = \zeta^2(s).$$

If the numbers m and n are coprime, then the number of divisors of their product mn is $\tau_m\tau_n$: if p is a divisor of m and q is a divisor of n, then pq is a divisor of mn, and each divisor of mn can be represented as the product of divisors of m and n in a unique way. Therefore, the sequence τ_n is multiplicative.

Example 7.10. Denote by ν_n the number of distinct prime factors of n. Then the sequence $a_n = a^{\nu_n}$ is multiplicative for any real a.

Each multiplicative sequence is uniquely determined by its elements whose indices are powers of prime numbers. In other words, the following analogue of Statement 7.6 takes place for multiplicative sequences.

Statement 7.11. *A sequence* $\{a_i\}$ *is multiplicative if and only if the corresponding Dirichlet generating function admits the following representation:*
(7.4)
$$\left(\frac{1}{1^s}+\frac{a_2}{2^s}+\frac{a_4}{4^s}+\dots\right)\left(\frac{1}{1^s}+\frac{a_3}{3^s}+\frac{a_9}{9^s}+\dots\right)\left(\frac{1}{1^s}+\frac{a_5}{5^s}+\dots\right)\dots,$$
where the product is taken over all prime numbers.

This statement immediately implies the following remarkable property of multiplicative sequences, which generalizes the fact that the Möbius sequence is multiplicative.

Corollary 7.12. *If Dirichlet generating functions* $A(s)$ *and* $B(s)$ *correspond to multiplicative sequences, then the sequences corresponding both to their product* $A(s)B(s)$ *and their ratio* $A(s)/B(s)$ *are multiplicative. In other words, Dirichlet generating functions corresponding to non-zero multiplicative number sequences form a group with respect to the multiplication.*

Indeed, if each of the functions $A(s), B(s)$ possess the representation (7.4), then both their product and their quotient possess this representation. Statement 7.11 follows immediately from the definition of a multiplicative sequence.

7.5. Problems

7.1. Using the inclusion-exclusion formula find the area of the spherical triangle on the unit sphere, having the angles α, β, γ.

7.2. Using the inclusion-exclusion formula find the number of lucky tickets with $2p$ digits in the number system to the base q.

7.3. Let $a_1, a_2, \dots, a_k$ be all distinct prime factors of a number $n = a_1^{p_1} \dots a_k^{p_k}$. Prove that the number φ_n of numbers smaller than n and prime to n is given by the formula
$$\varphi_n = n\left(1-\frac{1}{a_1}\right)\left(1-\frac{1}{a_2}\right)\dots\left(1-\frac{1}{a_k}\right).$$

7.4. Make use of the previous problem to show that the sequence φ_n is multiplicative.

7.5. Show that the number of distinct regular closed n-gons (including self-intersecting ones) inscribed in the unit circle is $\varphi_n/2$.

7.6. Count the number of uncancellable fractions among the n^2 fractions

$$\begin{array}{lllll} 1/1, & 1/2, & 1/3, & \ldots, & 1/n \\ 2/1, & 2/2, & 2/3, & \ldots, & 2/n \\ \ldots & & & & \\ n/1, & n/2, & n/3, & \ldots, & n/n. \end{array}$$

7.7. Show that the number of disorders in an n-element set is the integer closest to $n!/e$.

7.8. Suppose the diagonal elements of an $n \times n$-matrix are zero, while all other entries are non-zero. Count the number of non-zero products in the expansion of the determinant of this matrix.

7.9. Prove that the exponential generating function for the numbers of disorders is $D(s) = e^{-s}/(1-s)$.

7.10. Let μ_n denote the Möbius sequence. Show that

$$\prod_{n=0}^{\infty} (1 - x^n)^{-\mu_n/n} = e^x.$$

7.11. Describe all ideals in the algebra of Dirichlet generating functions.

7.12. Prove the identity

$$\begin{aligned} \max(a_1, \ldots, a_n) = a_1 + \cdots + a_n - \min(a_1, a_2) - \cdots - \min(a_{n-1}, a_n) \\ + \min(a_1, a_2, a_3) + \cdots + (-1)^{n-1} \min(a_1, \ldots, a_n). \end{aligned}$$

7.13. Set $\lambda_n = (-1)^k$, where k is the number of prime factors of n (taking multiplicities into account). Show that the sequence λ_n is multiplicative.

7.14. Find the Dirichlet function $\zeta(s)\lambda(s)$, where the coefficients of the function $\lambda(s)$ are defined in the previous problem.

7.15. Denote by $\sigma_\alpha(n)$ the sum of the divisors of n taken to the degree α, $\sigma_\alpha(n) = \sum_{t|n} t^\alpha$ (α is a non-negative integer). Prove that the

Dirichlet generating function for the sequence $\sigma_\alpha(n)$ is

$$\Sigma_\alpha(s) = \sum_{n=1}^{\infty} \frac{\sigma_\alpha(n)}{n^s} = \zeta(s)\zeta(s-\alpha).$$

Chapter 8

Enumeration of Embedded Graphs

The present chapter is devoted to some geometric aspects of combinatorics, including those undergoing active development during the last decades. These problems are related to graphs and their embeddings into two-dimensional surfaces. In this chapter we extensively use the notation $[s^k]f(s)$ for the coefficient f_k in a generating function $f(s) = f_0 + f_1 s + f_2 s^2 + \dots$

8.1. Enumeration of marked trees

Many difficulties in enumeration problems are due to the fact that objects under enumeration have distinct *symmetries*. Thus, if we consider equal those diagonal triangulations of regular polygons that are taken to each other by a rotation of the polygon (see Sec. 2.5), then receiving an exact formula for the number of triangulations would become a complicated problem, and the resulting formula would say nothing essentially new about triangulations. This happens because distinct triangulations possess non-isomorphic symmetry groups. For example, all six rotations of the triangulation of a hexagon shown in Fig. 1 a) produce distinct results, while the rotations of the triangulation shown in Fig. 1 b) lead to only two new triangulations, and

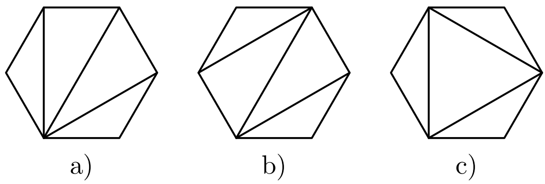


Figure 1. Three diagonal triangulations of the hexagon, with distinct symmetries

the rotations of the triangulation in Fig. 1 c) give only one additional picture.

On the other hand, the explicit formula expressing the number of diagonal triangulations of a polygon *with numbered vertices* in terms of the Catalan numbers, gives a good estimate for the asymptotics of the numbers of triangulations of polygons with non-numbered vertices. Indeed, the number of triangulations considered up to rotation of the $(n+2)$-gon is at most $c_n \sim \text{const} \cdot 4^n \cdot n^{-3/2}$, and it is not smaller than $c_n/(n+2) \sim \text{const} \cdot 4^n \cdot n^{-5/2}$. Hence, the break of symmetry (that is, the numbering of the vertices of a polygon) seriously simplified the problem and had only a minor impact on the precision of the answer. The same trick — marking — proves to be efficient in many other enumeration problems. We start with showing how it is used in enumeration of trees.

Definition 8.1. A *graph* is a triple $\Gamma = \{V, E, I\}$ consisting of a finite set of *vertices* V, a finite set of *edges* E and an *incidence mapping* $I \colon E \to V \times V$ assigning to each edge a pair of vertices, the *ends of the edge*, connected by the edge. An edge is called a *loop* if its ends coincide. The *valency* of a vertex in a graph is the number of edges having this vertex as an end (when computing valencies, each loop having both ends at the given vertex is counted twice).

Graphs are usually drawn in the plane. Vertices are shown as fat points, and edges as arcs connecting these points (see Fig. 2).

Remark 8.2. 1. A graph is more naturally understood not as the object defined above, but as an isomorphism class of such objects. Two triples $\Gamma_1 = \{V_1, E_1, I_1\}$ and $\Gamma_2 = \{V_2, E_2, I_2\}$ are *isomorphic* if

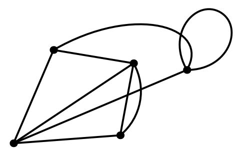


Figure 2. A plane picture of a graph. Fat points show the vertices and the edges of the graph are represented by arcs. The intersection points of the edges not marked with fat points are not vertices

there are one-to-one mappings $v\colon V_1 \to V_2$ and $e\colon E_1 \to E_2$ such that $I_2 \circ e = (v \times v) \circ I_1$. We shall use below mainly this definition.

2. The above definition of a graph admits various variants. For example, it is sometimes natural to require that at most one edge passes through each pair of vertices. Sometimes loops are forbidden, and so on. We will specify such restrictions at proper places.

3. From the topological point of view, a graph is a *one-dimensional complex.* If we introduce an *orientation* on each edge of a graph (that is, we choose one of the two directions of the edge), then the *boundary of the edge* is the formal difference between the end and the starting vertices of the edge.

Definition 8.3. Two vertices of a graph are said to be *adjacent* if there is an edge connecting them. A graph is said to be *connected* if for any pair $u, v \in V$ of its vertices there is a chain $v_0 = u, v_1, \dots, v_k = v \in V$ of the vertices of a graph such that the two vertices v_{i-1} and v_i are adjacent for each $i = 1, 2, \dots, k$. A *cycle* is a sequence $v_0, v_1, \dots, v_k \in V$ of graph vertices such that the vertices v_{i-1} and v_i are adjacent for each $i = 1, 2, \dots, k$, all vertices $v_0, v_1, \dots, v_{k-1}$ are distinct and $v_0 = v_k$. A *tree* is a connected graph without cycles.

All trees with n vertices ($n \leq 5$) are shown in Fig. 3.

Enumeration of trees with n vertices is a complicated problem since different trees have different symmetry. We will discuss a simpler problem, that of enumerating marked trees. Mark each of the vertices of a tree with one of the numbers in $\{1, 2, \dots, n\}$ in such a way that distinct vertices acquire distinct marks. All marked trees with $n \leq 4$

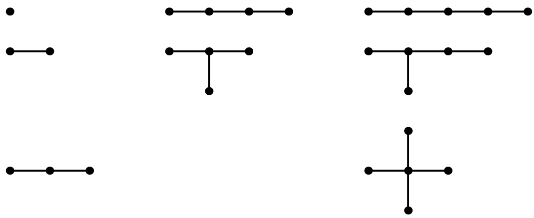


Figure 3. All trees with $n \le 5$ vertices

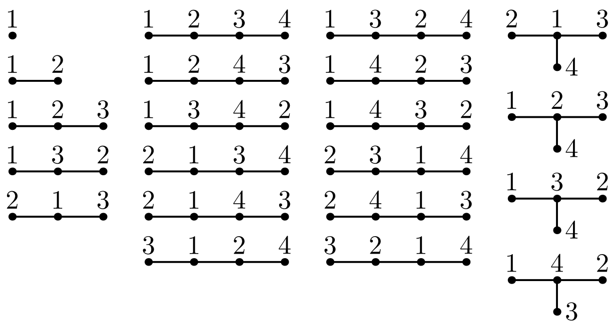


Figure 4. All marked trees with n vertices ($n \le 4$)

vertices are shown in Fig. 4. The sequence of numbers of marked trees with n vertices starts with the numbers $1, 1, 3, 16, \ldots$

Denote by T_n the number of rooted marked trees with n vertices, i.e., the number of marked trees with a distinguished vertex, called the *root* of the tree. Clearly, the number of rooted marked trees with n vertices is n times the number of marked trees with n vertices: there are n different choices of the root.

Let us find the exponential generating function

$$\mathcal{T}(s) = \sum_{n=1}^{\infty} \frac{1}{n!} T_n s^n = \frac{1}{1!}s + \frac{2}{2!}s^2 + \frac{9}{3!}s^3 + \frac{64}{4!}s^4 + \ldots$$

for the number of rooted marked trees. After we delete the root, the tree splits into several new trees; the number of these new trees coincides with the valency of the root. The new trees also can be considered to be marked: the only thing we have to do is to replace the existing marks $\{l_1, \dots, l_i\}$, $l_1 < \dots < l_i$, by the marks $\{1, \dots, i\}$, preserving their order. For the root of a new tree the vertex adjacent to the root of the initial tree is chosen. Hence, to each rooted marked tree with a root of valency k we have associated a (multi)set consisting of k rooted marked trees. We speak about multisets because some of the newly generated trees can coincide.

This description implies that trees with a root of valency k are enumerated by the exponential generating function $s\mathcal{T}^k(s)$. Indeed, exactly the elements

$$\frac{T_{l_1}}{l_1!} \cdots \frac{T_{l_k}}{l_k!} s^{l_1+\dots+l_k},$$

such that $l_1 + \dots + l_k = n$ contribute to the coefficient of s^{n+1} in $s\mathcal{T}^k(s)$. The set of marks of n vertices of k trees can be split into k disjoint subsets containing $l_1, \dots, l_k$ marks in $\binom{n}{l_1 \dots l_k} = \frac{n!}{l_1! \dots l_k!}$ ways. Therefore, the number of marked rooted trees having $n+1$ vertices and a root of valency k is

$$n![s^n]\mathcal{T}^k(s) = \sum_{l_1+\dots+l_k=n} \frac{n!}{l_1! \dots l_k!} T_{l_1} \dots T_{l_k}.$$

Now, summing the functions $\frac{1}{k!}\mathcal{T}^k$ over all k we obtain the following statement.

Theorem 8.4. *The exponential generating function $\mathcal{T}(s)$ for the number of marked rooted trees enumerating them with respect to the number of vertices satisfies the Lagrange equation*

$$\mathcal{T}(s) = se^{\mathcal{T}(s)}. \tag{8.1}$$

Now the Lagrange theorem allows us to compute easily first coefficients of the function $\mathcal{T}(s)$. For example, we get $T_5 = 625, T_6 = 7776$. However, it would be nice to have an explicit formula for the coefficients. To obtain such a formula we will need the following more precise version of the Lagrange theorem.

Theorem 8.5. *Suppose two functions $\varphi = \varphi(s)$ $(\varphi(0) = 0)$ and $\psi = \psi(t)$ are related by the Lagrange equation*

$$\varphi(s) = s\psi(\varphi(s)). \tag{8.2}$$

Then the coefficient of s^n in the function φ is

$$[s^n]\varphi(s) = \frac{1}{n}[t^{n-1}]\psi^n(t).$$

Let us apply this statement to Eq. (8.1) whose solution is the function $\mathcal{T}(s)$. We obtain

$$T_n = n![s^n]\mathcal{T}(s) = n!\frac{1}{n}[t^{n-1}]e^{nt} = (n-1)!\frac{n^{n-1}}{(n-1)!} = n^{n-1}.$$

Hence, we have proved the following result.

Theorem 8.6 (Cayley)**.** *The number of rooted marked trees with n vertices is $T_n = n^{n-1}$.*

Corollary 8.7. *The number of marked trees with n vertices is n^{n-2}.*

Proof of Theorem 8.5.

Lemma 8.8 (transformation of the residue under a variable change)**.** *For a function $g(t)$ such that $g(0) = 0, g'(0) \neq 0$, we have*

$$[s^{-1}]f(s) = [t^{-1}]f(g(t))g'(t).$$

Indeed, suppose $f(s) = f_{-N}s^{-N} + f_{-N+1}s^{-N+1} + \dots,\ g(t) = g_1t + g_2t^2 + \dots$ For $n \neq -1$ we have

$$[t^{-1}]g^n(t)g'(t) = [t^{-1}]\frac{1}{n+1}(g^{n+1}(t))' = 0,$$

since the residue of the derivative of an function is 0. For $n = -1$

$$[t^{-1}]f_{-1}\frac{1}{g(t)}g'(t) = f_{-1},$$

and the statement of the lemma follows.

The coefficient of s^n in the generating function φ has the form

$$[s^n]\varphi(s) = [s^{-1}]s^{n+1}\varphi(s).$$

Let us find the last residue using Lemma 8.8. To do this, rewrite the Lagrange equation (8.2) as the variable change

$$s = \frac{t}{\psi(t)},$$

where $t = \varphi(s)$. Then the lemma gives

$$\begin{aligned}
[s^{-1}]s^{-n-1}\varphi(s) &= [t^{-1}]\frac{\psi^{n+1}(t)}{t^n} \cdot \frac{\psi(t) - t\psi'(t)}{\psi^2(t)} \\
&= [t^{-1}]\left(\frac{\psi^n(t)}{t^n} - \frac{\psi^{n-1}(t)\psi'(t)}{t^{n-1}}\right) \\
&= [t^{n-1}]\psi^n(t) - \frac{1}{n}[t^{n-2}](\psi^n(t))' \\
&= \frac{1}{n}[t^{n-1}]\psi^n(t),
\end{aligned}$$

which completes he proof of the theorem.

8.2. Generating functions for non-marked, marked, ordered, and cyclically ordered objects

As we have seen, some sequences are better described in terms of ordinary generating functions, while others in terms of exponential ones. There are exceptions, however. For example, the exponential generating function for the number of up-down permutations is either the tangent or the secant (depending on the parity of the cardinality of the permuted set; see Sec. 5.4), while the corresponding ordinary generating functions admit remarkable representations as continued fractions (Sec. 5.5).

However, the general rule states that exponential generating functions better describe marked objects, while ordinary generating functions fit better for the description of non-marked objects. The following observation produces a base for this rule. Suppose we have a class of objects and we study finite ordered sequences of objects in this class and cyclically ordered sequences.

Statement 8.9. *Suppose the objects of the class are marked and* $\mathcal{A}(s) = \sum a_n s^n/n!$ *is the exponential generating function for their numbers; then the exponential generating function for sequences of these objects is* $1/(1-\mathcal{A}(s))$, *and the exponential generating function for cyclic sequences equals* $\ln(1/(1-\mathcal{A}(s)))$.

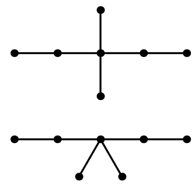


Figure 5. Two distinct embeddings of a tree into the plane

If the objects of the class are non-marked and $B(s) = \sum b_n s^n$ *is the ordinary generating function for them, then the generating function for sequences of these objects is* $1/(1-B(s))$, *and the generating function for cyclic sequences is* $\sum \frac{\varphi_k}{k} \ln(1/(1-B(s^k)))$, *where* φ_k *is the Euler function, that is, the number of integers between* 1 *and* k *prime to* k.

Here we suppose that an enumerating parameter, a weight, is assigned to each object (e.g., the number of vertices in a graph) in such a way that the weight of a complex object consisting of several simple parts is the sum of the weights of the parts.

Hence, associating to marked objects exponential generating functions and to unmarked objects ordinary generating functions we arrive at natural generating functions enumerating complex objects.

8.3. Enumeration of plane and binary trees

It is obvious that each tree can be drawn on the plane in such a way that its edges have no points of intersection and self-intersection other than the common vertices. (The edges can be even chosen as segments of straight lines, but we will not make use of this fact.) However, the same tree can admit distinct plane representations (see Fig. 5). The notion of embedding of a tree into the plane is formalized in the following definition.

Definition 8.10. Two embeddings of a tree into the plane are called equivalent if there is an orientation preserving homeomorphism of the plane taking the image of the first embedding to the second one. An equivalence class of embeddings of a tree is called a *plane tree.*

Hence, several plane trees may correspond to the same tree. Note that each tree with $n \leq 6$ vertices admits a single plane embedding.

Enumeration of plane trees remains a complicated problem, since distinct plane trees may have distinct symmetries. In order to destroy the symmetry, let us choose a *leaf* (that is, a vertex of valency one) for the root of the tree. A tree with such a root will be called *planted.* It is clear that the only transformation taking a planted plane tree into itself is the identity transformation. All planted plane trees having n vertices ($2 \leq n \leq 5$) are shown in Fig. 6. The trees are shown growing down from the root. The numbers $1, 1, 2, 5, \ldots$ of planted plane trees having $2, 3, 4, 5, \ldots$ vertices give us a hint that these trees are enumerated by the Catalan numbers.

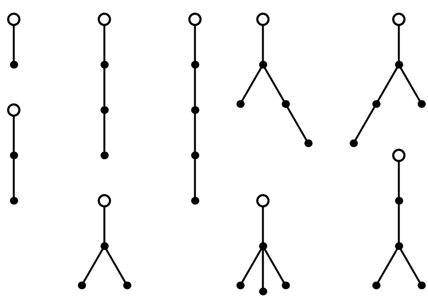


Figure 6. Planted plane trees having n edges ($1 \leq n \leq 4$)

Theorem 8.11. *The number of planted plane trees having $n + 2$ vertices is equal to the nth Catalan number c_n.*

Proof. One can associate to each vertex in a rooted tree a non-negative integer which is equal to the distance from this vertex to the root (the *level* of a vertex). The root itself has zero level, its neighbors have level one, and so on.

Denote the number of planted plane trees (in the course of the proof, we will refer to these objects simply as "trees") having $n + 2$ vertices by p_n. Then $p_0 = p_1 = 1$. For $n > 1$, associate to each tree with $n + 3$ vertices two trees in the following way. The first tree is the subtree of the initial tree growing from the leftmost edge issuing from the (unique) vertex of level one; the second tree is the remaining

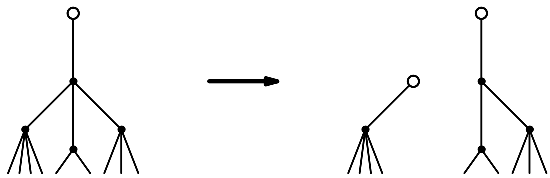


Figure 7. Associating to a planted plane tree a pair of trees of the same kind

part of the initial tree (see Fig. 7). The copy of the level one vertex becomes the root of the first tree in the pair. If the first tree has $k+2$ vertices, then the number of the vertices in the second tree is $n-k+2$. Conversely, having two trees with $k+2$ and $n-k+2$ vertices we can construct from them a tree with $n+3$ vertices by attaching the first tree at the level one vertex of the second tree on the left.

Hence,

$$p_{n+1} = p_0p_n + p_1p_{n-1} + \cdots + p_np_0,$$

and planted plane trees are enumerated by the Catalan numbers.

8.4. Graph embeddings into surfaces

Plane trees provide us with examples of embeddings of graphs into the plane. The same tree may admit distinct embeddings. On the other hand, we are going to show that not each graph admits an embedding into the plane. We will also touch on the problem of embedding graphs into arbitrary two-dimensional surfaces.

Giving no definition of a two-dimensional surface we shall make use of the classification theorem for such surfaces usually proved in standard elementary courses of topology.

Theorem 8.12. *Each* (*closed orientable two-dimensional*) *surface is homeomorphic to the sphere with finitely many handles attached.*

We will use the description given by this theorem for the definition of a surface.

Definition 8.13. A *surface* of *genus* g is the two-dimensional sphere with g handles glued to it.

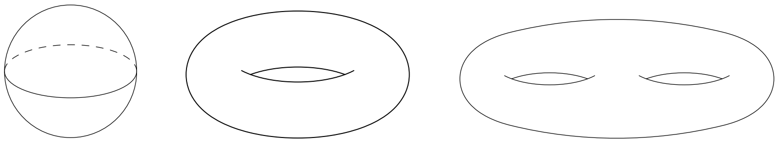


Figure 8. The surfaces of genus g for $g = 0, 1, 2$

In particular, the surface of genus 0 is the sphere itself, the surface of genus 1 is the torus, and so on (see Fig. 8).

In what follows, a graph is allowed to have loops and multiple edges (i.e., some pairs of its vertices may be connected by more than one edge). Graphs admitting an embedding into the plane (*planar* graphs) are exactly those graphs that admit an embedding into the sphere. Indeed, puncturing the sphere at a point we obtain a surface homeomorphic to the plane.

Definition 8.14. An *embedding* of a connected graph Γ into a surface M is a drawing of the graph on the surface such that

1) each vertex of the graph is represented by a point in M and distinct vertices are represented by distinct points;

2) each edge of the graph is represented by a non-selfintersecting curvilinear segment in M, with the ends of the segment coinciding with the vertices connected by the edge; no two segments intersect each other;

3) the complement to the image of Γ in M is a disjoint union of *cells* (two-dimensional domains homeomorphic to the disk).

The image of a graph under an embedding is called an *embedded graph* (or a *map*). As in the case of embedded trees, we make no difference between embeddings taken to each other by a homeomorphism of the ambient surface.

The first two requirements in the definition of an embedding coincide with those in the definition of an embedding of a graph into the plane. The third requirement is new. In fact, for embeddings in the sphere it is automatically satisfied: a connected graph can cut the sphere only into domains homeomorphic to the disk. For surfaces of higher genus this is not the case: the complement to a graph can

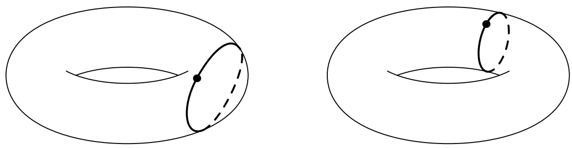


Figure 9. Drawings of a graph on the torus, which are not embeddings

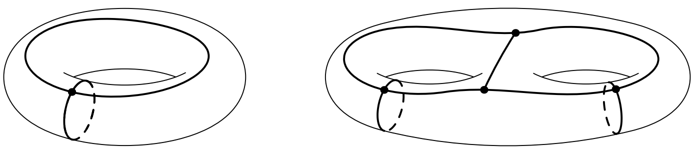


Figure 10. Graph embeddings into the torus ($g = 1$) and pretzel ($g = 2$)

have handles (see Fig. 9). The goal of the third requirement is to forbid such situations. We take for embeddings only those drawings that cut all handles of the surface.

In what follows we will, by some abuse of language, refer to the image of a graph on a surface as to the graph itself, and the images of the vertices and the edges will be referred to as simply vertices and edges.

In Fig. 10 graph embeddings into the torus and into the surface of genus 2 are shown. Already these examples demonstrate that the same graph can be embedded into different surfaces. For example, the "eight" graph, which consists of a single vertex with two loops, can be embedded both in the sphere and in the torus.

Suppose a graph Γ having V vertices and E edges is embedded into a surface M of genus g. Denote by F the number of cells in the complement to Γ in M. Then the numbers V, E, F and g are related by the following famous Euler's formula.

Theorem 8.15 (Euler)**.**

$$V - E + F = 2 - 2g.$$

The number $\chi_g = 2 - 2g$ is called the *Euler characteristic* of the surface M.

Euler's formula gives an immediate simple restriction on the maximal genus of the surface admitting an embedding of a given graph.

Corollary 8.16. *A graph with V vertices and E edges cannot be embedded in a surface of genus bigger than $(E - V + 1)/2$.*

Indeed, by Euler's formula,

$$g = (E - V - F + 2)/2 \leq (E - V + 1)/2,$$

since $F \geq 1$.

For example, the "eight" graph cannot be embedded into the surface of genus bigger than $(2 - 1 + 1)/2 = 1$.

Our closest goal is to show that any graph admits an embedding into some surface. Suppose a graph Γ is already embedded in a surface M. Consider a neighborhood of a vertex v of Γ and map it to the plane homeomorphically on the image preserving the orientation. The image of such a neighborhood is a disk containing the vertex of the graph and a set of *half-edges* issuing from it (see Fig. 11). (Some of these half-edges can belong to the same edge if there are loops in Γ attached to the chosen vertex. The number of half-edges issuing from a vertex coincides with its valency.) Let us define the following *cyclic order* in the set of half-edges: we say that a half-edge b follows immediately after a half-edge a if it follows immediately after a when moving along the boundary of the neighborhood in the counterclockwise direction. The half-edges in Fig. 11 have the following cyclic order:

$$\cdots \succ a \succ b \succ c \succ d \succ a \succ \ldots .$$

Hence, an embedding of a graph into a two-surface equips it with a cyclic order on each of the sets of half-edges issuing from each of its vertices.

Definition 8.17. A *graph with rotations* is a graph endowed with a cyclic order on each set of half-edges issuing from each of its vertices.

Example 8.18. Four half-edges issue from the only vertex of the "eight"-graph. Fig. 12 shows two distinct possibilities to define a cyclic order on this quadruple of half-edges (given by the orientation of the plane). There are no other possibilities. The first of them is

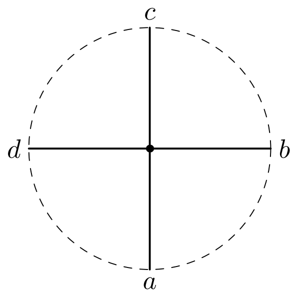


Figure 11. A neighborhood of a vertex of valency 4 in an embedded graph

realized by the embedding of the "eight"-graph in the plane, and the second one by the embedding into the torus from Fig. 10.

Definition 8.19. An *embedding of a graph with rotations* into a surface is an embedding of the graph which induces the same cyclic order as the given one on each set of half-edges issuing from each vertex.

It happens that not only an embedding of a graph into a surface determines the corresponding graph with rotations uniquely, but conversely the surface of embedding can be uniquely reconstructed from any graph with rotations.

Theorem 8.20. *For each graph with rotations there is a two-dimensional closed surface in which it can be embedded preserving the rotations, and this surface is unique.*

Remark 8.21. Before proving the theorem, let us remark that not only the surface of embedding, but the embedding itself is uniquely determined up to a homeomorphism of the surface. It is also possible, however, to consider another, more subtle, equivalence relation on the

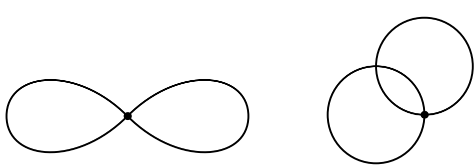


Figure 12. Two distinct cyclic orders on the set of half-edges issuing from the vertex of the "eight"-graph

set of embeddings: that up to *isotopy*. The two embeddings of the "eight"-graph into the torus shown in Fig. 13 cannot be transformed one into another by an isotopy, although they are homeomorphic. For the sphere, the equivalence relations with respect to homeomorphisms and isotopy coincide.

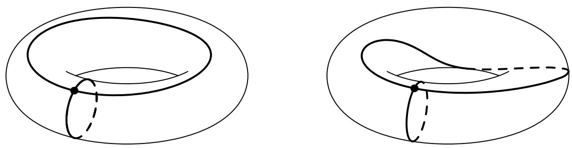


Figure 13. Two homemorphic but not isotopic embeddings of the "eight"-graph into the torus

Remark 8.22. Planar graphs are exactly those graphs that admit an embedding into the plane. However, two distinct embeddings of a graph into the plane may prove to be the same when considered as embeddings into the sphere. A simple example of such embeddings is shown in Fig. 14. The loop splits the plane into two domains, the internal and the external. These two domains cannot be exchanged by a homemomorphism of the plane. On the sphere, however, the internal and the external domains are indistinguishable.

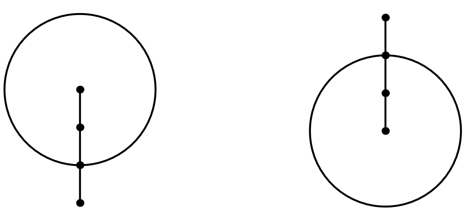


Figure 14. Two distinct plane graphs leading to the same spherical graph

Proof of Theorem 8.20. Let Γ be a graph with rotations. Draw the vertices of Γ and the outgoing half-edges as a set of "stars" on the plane (see Fig. 15). Now connect the ends of the half-edges in pairs so as to obtain the required graph Γ. Of course after connecting the half-edges, some of the resulting edges may intersect. However, we

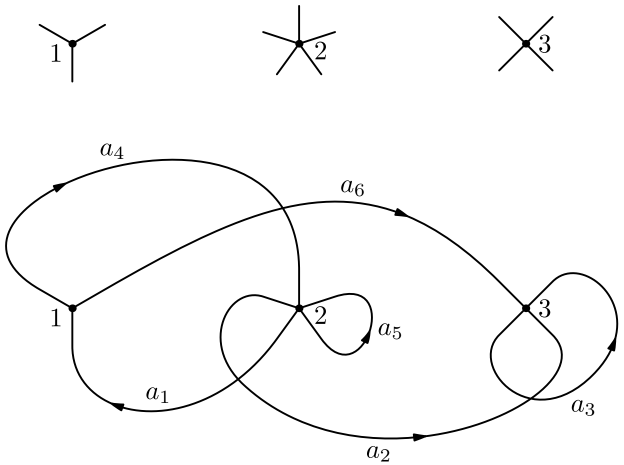


Figure 15. The graph obtained by connecting the half-edges of the stars (up) with given cyclic orders

require that the edges do not pass through the existing vertices, and points of their intersection are not considered as vertices. Put arrows on all the edges (in other words, "orient" the edges) and mark each edge with a letter, distinct edges having distinct marks. Using this data we will construct a surface M independent of the choice both of the orientations and marking.

The construction of the surface M splits into two stages. First of all we determine the set of cells, and then we glue the cells together. We describe the construction of the set of cells for the graph with rotations shown in Fig. 15. For the general case, it follows the same routine.

The cells are chosen as follows. Consider the edge a_1 of Γ. Moving along this edge we arrive at the vertex 1. A cyclic order of half-edges at vertex 1 is fixed, and we leave it along the half-edge which follows immediately after the half-edge of the edge a_1. Hence, we leave along the half-edge of the edge a_6 and arrive at vertex 3. We must leave vertex 3 along the half-edge a_3 which follows immediately after the edge a_6. The loop a_3 returns back to the same vertex 3, and we must leave it along the edge a_6, now in the direction opposite to that of the orientation of a_6. This time we return to 1 and leave this vertex in direction a_4.

This process is iterated until our path reaches *an edge of the graph which it had already passed in the same direction.*

Now we can write the path as a word, where each edge is denoted by the corresponding letter. An edge passed in the direction coinciding with its orientation is denoted just by the letter itself, while for edges passed in the opposite direction the letter is inverted. For the path in the example, the word looks like

$$a_1 a_6 a_3 a_6^{-1} a_4 a_2 a_3^{-1} a_2^{-1}.$$

On the next step, the first edge a_1 is repeated. Let us prove that this is an instance of a general situation.

Lemma 8.23. *The first repeated edge in each path is the first edge of this path.*

Indeed, suppose that this is not true and the first repeated edge is some edge x not coinciding with the first one. Let us look at the half-edge preceding the edge x at its starting vertex. This half-edge is defined uniquely by the rotation at this vertex. Therefore, the edge containing this half-edge must be repeated before x, and we arrive at a contradiction. The lemma is proved.

We associate to the path constructed above the octagon with edges marked by the same letters, in the same cyclic order. To construct the next cell, let us take one of the letters a_i^σ, $i = 1, \dots, 6$, $\sigma \in \{-1, +1\}$, not entering the already constructed cycle and construct the new cycle passing through the corresponding edge. In our example, we can start, say, with the edge a_5. A proof repeating that of the lemma almost literally shows that any two cycles constructed in this way either do not intersect (have no common edge passed in the same direction), or coincide. Therefore, all edges in the graph split into disjoint cycles. The result of this decomposition is a set of words such that each of the letters a_i^σ occurs in these words exactly once.

In our example, the set of cycles has the form

$$a_1 a_6 a_3 a_6^{-1} a_4 a_2 a_3^{-1} a_2^{-1}, \qquad a_5 a_1^{-1} a_4^{-1}, \qquad a_5^{-1}.$$

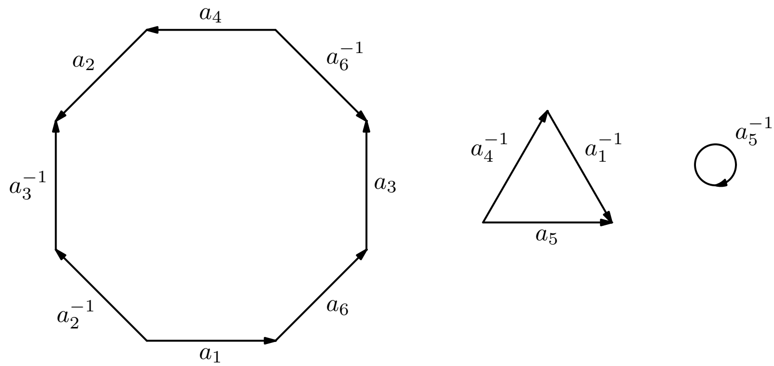


Figure 16. The three cells with marked edges, of which the embedding surface of the graph in Fig. 15 is glued

To this set of cycles, the set of cells shown in Fig. 16 is associated. The cells are glued together along the edges having the same marks but with opposite exponents. Clearly, the rotations induced by the resulting embedding coincide with the original rotations in the graph. This completes the proof of the theorem.

Remark 8.24. The above construction shows how to construct a homeomorphism of the surface taking one embedding to another embedding of the same graph with rotations. After cutting the surface along the edges of the graph, we obtain a set of cells. Each cell is a polygon with edges numbered by the edges of the graph (taking edge orientations into account). The set of polygons is independent of the embedding; it is constructed from the graph with rotations as in the proof of the theorem. To construct the required homeomorphism, it suffices to glue it from edge preserving homeomorphisms of the cells.

Remark 8.25. One can also study graph embeddings into non-orientable surfaces. Such an embedding defines not one, but two cyclic orders on the set of half-edges issuing from each vertex of the graph. These two cyclic orders are mutually inverse. In the non-orientable case, the correspondence between embeddings and rotations is not one-to-one. A graph equipped with a pair of inverse rotations for each set of half-edges issuing from each vertex can be embedded in different non-orientable surfaces.

For an embedded graph the dual embedded graph is well defined. The dual graph is embedded into the same surface. In Fig. 17 two pairs of mutually dual graphs in the sphere and in the torus are shown.

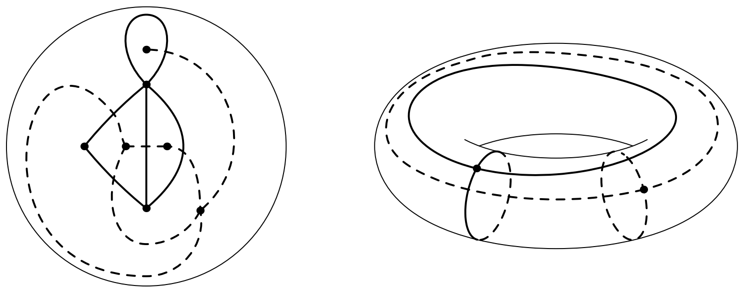


Figure 17. Pairs of mutually dual graphs on the sphere and torus

Definition 8.26. Suppose a graph Γ is embedded into a surface M. The *dual embedded graph* is the graph $\widetilde{\Gamma}$ embedded in M whose vertices are in one-to-one correspondence with the faces of Γ, edges are in one-to-one correspondence with the edges of Γ and each edge connects the vertices corresponding to the faces separated by the corresponding edge of Γ.

Example 8.27. As an illustration to the notion of dual embedded graph let us describe an embedding of the graph K_7, the complete graph with 7 vertices, into the torus. This embedding is used to give an example of a graph such that its vertices cannot be colored into six colors so as to make all neighboring vertices be of distinct colors. The dual embedding is shown in Fig. 18. The torus is obtained by gluing opposite sides of the hexagon in pairs. Each of the seven cells of the resulting partition of the torus is a neighbor of each of the other six cells.

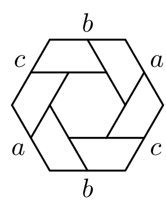


Figure 18. An embedding of a graph in the torus, with the dual embedded graph being the K_7 graph

8.5. On the number of gluings of a polygon

One can produce closed orientable surfaces by gluing sides of polygons in pairs. For example, gluing opposite sides of a square we obtain a torus; see Fig. 19).

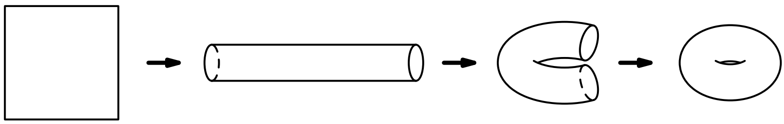


Figure 19. Gluing a torus from a square

Consider a regular $2n$-gon and split its sides into pairs in all possible ways. For each such splitting, glue together the sides belonging to the same pair (preserving the orientability of the surface). The result will be an orientable closed surface. We are interested in enumerating the number of ways to obtain a surface of genus 0, or genus 1, ..., or genus g.

Let us start, as usual, with examples.

Suppose $n = 2$ and we study gluings of the square. There are three ways to split the sides of the square into pairs. After gluing, the first two ways lead to the sphere, while the third way produces the torus (see Fig. 20). For $n = 3$ there are 15 ways to split the sides of the hexagon into pairs. It is easy to see that the five splittings in the first row of Fig. 20 produce the sphere, while the remaining ten splittings lead to the torus.

Note that we make a difference between a gluing and another one obtained from it by a rotation (or a reflection) of the polygon. This means that we used one of the ways to destroy the symmetry. We may assume, for example, that either the sides or the vertices of the polygon are marked. Or we may choose an initial side of the polygon. These marks are not shown in the picture in order to make it more transparent.

It happens that computing the genus of a gluing is an easy task. The image of the boundary of the polygon under the gluing is a graph embedded into the surface. Indeed, the complement to this graph is

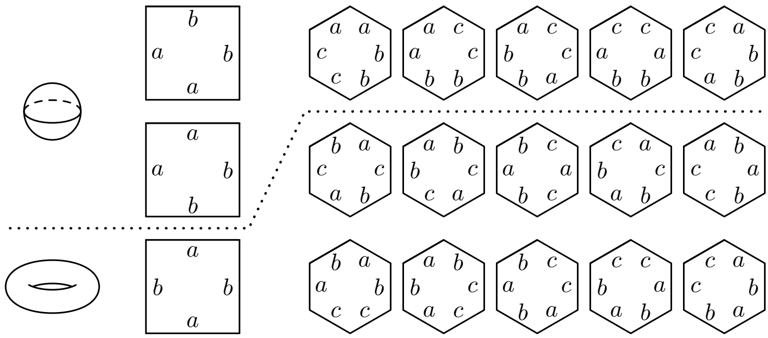


Figure 20. All gluings of the square and the torus. Two sides forming a pair are marked by the same letter

a single cell, the interior of the polygon. Hence we know the number of faces. The number of edges also is known: it is half the number of sides of the $2n$-gon, i.e., it equals n. What we need is only the number of vertices. To find it, let us mark the vertices of the polygon glued into a single vertex of the embedded graph by the same number (see Fig. 21). The number of distinct markings is exactly the number of vertices of the embedded graph. For example, the Euler characteristic for the surface obtained from the gluing in Fig. 21 is

$$2 - 2g = V - E + F = 2 - 3 + 1 = 0,$$

hence the surface is the torus.

Hence, the genus of the resulting surface is uniquely determined by the number of vertices in the graph glued from the sides of the polygon.

Our problem consists of enumerating gluings of a $2n$-gon leading to a surface of genus g. This problem admits a natural reformulation in terms of the dual graphs. In the dual setting, we replace the $2n$-gon with the $2n$-star on the plane. Each partition of the ends of the half-edges in this $2n$-star into pairs determines a graph with rotations having a single vertex. We ask how many of these graphs have genus g.

Associate to each number n a polynomial $T_n = T_n(N)$ which is the generating polynomial for the numbers of gluings of the sides of the polygon. The coefficient of N^V in the polynomial T_n equals

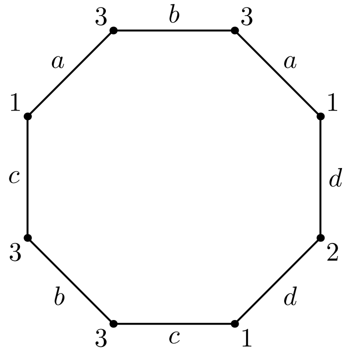


Figure 21. A marking of the vertices of a polygon. The vertices glued into a single vertex of the embedded graph have coinciding marks

the number of gluings of the $2n$-gon with V vertices in the resulting graph. This data uniquely determines the genus of the resulting surface. Hence, the number of ways to glue a surface of genus g from the $2n$-gon is

$$[N^{n-2g+1}]T_n(N).$$

Now let us write out several polynomials T_n for small n. For convenience, we set $T_0(N) = N$. Then, $T_1(N) = N^2$: the only gluing of the "regular 2-gon" gives a graph with two vertices and one edge on the sphere. We have already computed the next two polynomials:

$$T_2(N) = 2N^3 + N;$$
$$T_3(N) = 5N^4 + 10N^2.$$

By drawing all gluings of the regular octagon and decagon we can find two more polynomials:

$$T_4(N) = 14N^5 + 70N^3 + 21N,$$
$$T_5(N) = 42N^6 + 420N^4 + 483N^2,$$

but already these calculations are laborious.

It is clear that the sum of all coefficients of the polynomial T_n, i.e., its value $T_n(1)$ at $N = 1$, is $(2n-1)!! = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n-1)$. Indeed, the sum of all coefficients of T_n coincides with the number of splittings into pairs the sides of the $2n$-gon. It is easy to compute the number of such splittings. Indeed, choose a side of the polygon. Then the pairing side can be chosen in $2n-1$ ways. Taking one of the remaining $2n-2$ sides, we can find a pair to it among $2n-3$ sides not involved yet, and so on.

Now let us form the generating function in two variables N and s, which is an ordinary generating function in the first variable and exponential in the second one:

$$\begin{aligned} T(N;s) &= 1 + 2sT_0(N) + 2s^2\frac{T_1(N)}{1!!} + 2s^3\frac{T_2(N)}{3!!} + \dots \\ &= 1 + 2s\sum_{n=0}^{\infty} s^n \frac{T_n(N)}{(2n-1)!!}. \end{aligned}$$

In 1986, an unexpectedly simple and elegant expression for this generating function was discovered.

Theorem 8.28 (Harer, Zagier). *We have*

$$T(N;s) = \left(\frac{1+s}{1-s}\right)^N.$$

A proof (unfortunately, incomplete) of this theorem will be given in the next section.

Remark 8.29. Computation of the generating function $T(N;s)$ was not the main goal of the Harer and Zagier paper. The above formula serves there as a main tool in the calculation of the *virtual Euler characteristic of the moduli spaces* of smooth complex curves of given genus g whatever this means.

Let us check that the formula in the theorem indeed gives correct first coefficients of the generating function. We have

$$
\begin{aligned}
T(N;s) &= (1+s)^N(1-s)^{-N}\\
&= \left(1+\frac{N}{1!}s+\frac{N(N-1)}{2!}s^2+\frac{N(N-1)(N-2)}{3!}s^3+\dots\right)\\
&\quad\cdot\left(1+\frac{N}{1!}s+\frac{N(N+1)}{2!}s^2+\frac{N(N+1)(N+2)}{3!}s^3+\dots\right)\\
&= 1+\left(\frac{N}{1!}+\frac{N}{1!}\right)s+\left(\frac{N(N-1)}{2!}+\frac{N}{1!}\frac{N}{1!}+\frac{N(N+1)}{2!}\right)s^2\\
&\quad+\left(\frac{N(N-1)(N-2)}{3!}+\frac{N(N-1)}{2!}\frac{N}{1!}+\frac{N}{1!}\frac{N(N+1)}{2!}\right.\\
&\qquad\left.+\frac{N(N+1)(N+2)}{3!}\right)s^3+\dots\\
&= 1+2Ns+2N^2s^2+2\frac{2N^3+N}{3}s^3+\dots,
\end{aligned}
$$

and at least the first terms in the expansion coincide with those we have computed above.

8.6. Proof of the Harer–Zagier theorem

Let us first look at the sequence of the leading coefficients of the polynomials T_n. This sequence starts with the numbers $1, 1, 2, 5, 14, \dots$, which hints that it coincides with the Catalan sequence.

Lemma 8.30. *The degree of the polynomial T_n is $n+1$. The coefficient of N^{n+1} in the polynomial $T_n(N)$ is c_n, the nth Catalan number.*

Proof. The number of vertices of a graph embedded in a surface of genus g can be found from the Euler formula

$$2 - 2g = V - E + F.$$

For gluings of a polygon, the right-hand side is $V - n + 1$, and we obtain

$$V = n + 1 - 2g \le n + 1,$$

since the genus g is non-negative. The last inequality turns into equality if and only if $g = 0$, i.e., if the resulting surface is the sphere.

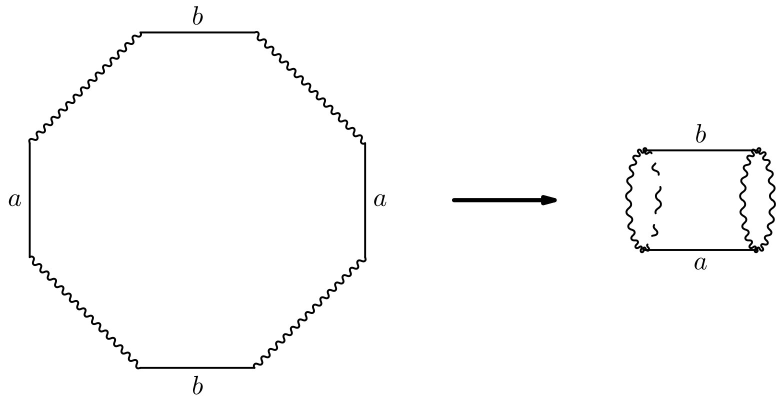


Figure 22. Gluing a handle from alternating pairs of sides

Call two pairs of polygon sides *alternating* if there is a side belonging to the second pair between the sides of the first pair, whatever order of the sides in the first pair we choose. In other words, two pairs of polygon sides alternate if the segment connecting the midpoints of the sides of the first pair intersects the segment connecting the midpoints of the sides of the second pair. If there are alternating pairs in a partition of the sides of a polygon into pairs, then after gluing these pairs we obtain a handle (see Fig. 22), and the resulting surface cannot be the sphere. Conversely, if there are no alternating pairs in a partition, then the resulting surface is the sphere. Indeed, one can always choose in such a partition a pair of adjacent sides that must be glued together. After gluing these two sides we obtain a polygon with less vertices and a partition of its sides into non-alternating pairs, and we can proceed by induction.

Partitions of the sides of a $2n$-gon into non-alternating pairs are in a one-to-one correspondence with the set of regular bracket structures of n pairs of brackets. Indeed, choose some of the vertices of the polygon for the initial vertex, and move along the boundary of the polygon in the counterclockwise direction starting at this vertex. We associate to each of the sides of the polygon we meet a left or a right bracket according to the following rule: if it is the first side in a pair, then the corresponding bracket is the left one, otherwise it is the

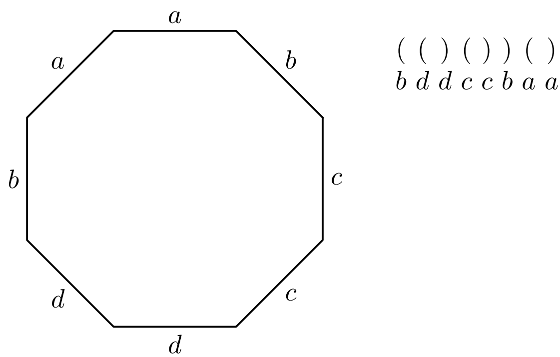


Figure 23. A partition of the sides of an octagon into non-alternating pairs and the corresponding regular bracket structure

right bracket (see Fig. 23). Clearly, the resulting bracket structure will be regular, and there is a unique way to get each regular bracket structure.

The proof of the lemma is completed.

The remaining part of the proof of the Harer–Zagier theorem splits into two stages. The first stage is the following lemma.

Lemma 8.31. *Consider the expression* $t(N,n) = \frac{T_n(N)}{(2n-1)!!}$ *as a function in* n*, for* N *fixed. Then* $t(N,n)$ *is a polynomial in* n *of degree* $N-1$.

The first proof of this fact, although not a complicated one, uses a non-trivial technique of integration over the space of Hermitian $N \times N$-matrices. Recently a new, purely combinatorial, proof due to B. Lass appeared. However, it also is laborious. I refer the interested reader to the original papers.

The remaining part of the proof is purely combinatorial.

Suppose the vertices of the $2n$-gon are colored in several colors. We call a gluing and a coloring *compatible* if only vertices of the same color are glued together.

Lemma 8.32. *The number $T_n(N)$ is exactly the number of gluings of the sides of a $2n$-gon compatible with its colorings in (not more than) N colors.*

Proof. Indeed, suppose that after the gluing, the sides of the polygon form a graph with V vertices. Color each vertex of the graph in one of N colors. Each such coloring produces a coloring of the vertices of the polygon compatible with the gluing. And there are exactly N^V ways to color the V vertices of the graph into N colors. Therefore, the total number of compatible colorings and gluings is the sum of the numbers N^V over all gluings, which coincides with the definition of the polynomial $T_n(N)$. The lemma is proved.

Introduce the function $\overline{T}_n(N)$, the number of gluings of the $2n$-gon compatible with colorings of its vertices in *exactly* N colors. Then

$$T_n(N) = \sum_{L=1}^{N} \binom{N}{L} \overline{T}_n(L). \tag{8.3}$$

Indeed, there are $\binom{N}{L}$ ways to choose L colors in a given set of N colors.

Now, we know that $\overline{T}_0(N) = \overline{T}_1(N) = \cdots = \overline{T}_{N-2}(N) = 0$, since, by Lemma 8.30, there are no gluings compatible with a coloring of the vertices of the $2n$-gon in more than $n+1$ colors. Hence, for a fixed positive integer N, we know $N-1$ roots of the function $\bar{t}(N,n) = \frac{\overline{T}_n(N)}{(2n-1)!!}$, which is a polynomial in n of degree $N-1$. These roots are $n = 0, 1, 2, \ldots, N-2$. Therefore, there exists a constant A_N such that

$$\begin{aligned} \bar{t}(N,n) &= A_N n(n-1)\ldots(n-N+2) \\ &= A_N \binom{n}{N-1}(N-1)!. \end{aligned}$$

Substituting this expression in Eq. (8.3) we obtain

$$T_n(N) = (2n-1)!! \sum_{L=1}^{N} A_L \binom{n}{L-1}\binom{N}{L}(L-1)!.$$

In particular, the leading coefficient of this polynomial in N is

$$(2n-1)!!\frac{A_{n+1}}{(n+1)}.$$

On the other hand, by Lemma 8.30, this coefficient coincides with the nth Catalan number,

$$(2n-1)!!\frac{A_{n+1}}{(n+1)} = c_n = \frac{(2n)!}{n!(n+1)!}.$$

This equation gives $A_{n+1} = \frac{2^n}{n!}$, and hence

$$T_n(N) = (2n-1)!!\sum_{L=1}^{N} 2^{L-1}\binom{N}{L}\binom{n}{L-1}.$$

The last formula gives exactly the coefficient of s^n in the expansion of $\left(\frac{1+s}{1-s}\right)^N$:

$$\begin{aligned}\left(\frac{1+s}{1-s}\right)^N &= (1+2s+2s^2+2s^3+\dots)^N \\ &= 1 + s\sum_{n=0}^{\infty}\sum_{L=1}^{N} 2^L\binom{N}{L}\binom{n}{L-1}s^n,\end{aligned}$$

which completes the proof of the Harer–Zagier theorem.

8.7. Problems

8.1. What is the genus of the surface into which the graph with rotations from Fig. 15 is embedded?

8.2. Prove that the number of distinct markings of the vertices of a given tree with n edges with the marks $\{1, 2, \dots, n+1\}$ is n times bigger than the number of markings of its edges with the marks $\{1, 2, \dots, n\}$.

8.3. A *binary* tree is a tree all of whose vertices have valency either 1 or 3. Prove that the number of vertices in each binary tree is even. Prove that the number of planted plane binary trees with $2n$ vertices coincides with the Catalan number c_{n-1}.

8.4. Prove that the number t_n of planted plane ternary trees (that is, trees all of whose vertices have valency either 4 or 1) with $2(n+1)$

leaves is $t_n = \frac{1}{2n+1}\binom{3n}{n}$. Derive the generating function for these numbers.

8.5. Establish a one-to-one correspondence between planted plane binary trees with $2n$ vertices and planted plane trees with $n+1$ vertices.

8.6. Enumerate rooted marked forests. (A *forest* is a graph all of whose connected components are trees. A forest is rooted if all its connected components are rooted.)

8.7. Give estimates for the minimal and the maximal genus of a surface, where the graph K_n, the complete graph with n vertices, can be embedded. (A *complete graph* is a graph with each pair of whose vertices are connected by an edge.)

8.8. Are there embeddings of the Petersen graph (see Fig. 24) into a) the torus; b) the surface of genus 2?

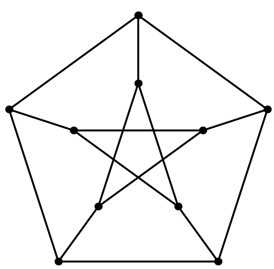


Figure 24. The Petersen graph

8.9. Is it possible to embed into the torus the graph K_2^4, the 1-skeleton of the 4-dimensional cube?

8.10. Using the Euler formula prove the *easy part of the Kuratovski theorem*: the graphs K_5 and $K_{3,3}$ admit no embeddings in the sphere. (The graph $K_{3,3}$ has six vertices split into triples, and each vertex of the first triple is connected by an edge with each vertex of the second triple, while there are no edges inside the triples.)

8.11. Is it possible to embed the graphs a) K_5; b) $K_{3,3}$; c) the Petersen graph into the projective plane? If yes, then describe the corresponding embeddings.

8.12. Let Γ be the graph formed by the edges of the icosahedron. The five edges at each vertex of the icosahedron have a natural cyclic order $\ldots 1 \succ 2 \succ 3 \succ 4 \succ 5 \succ 1 \ldots$ Let us replace this natural cyclic order at each vertex with the cyclic order $\ldots 1 \succ 3 \succ 5 \succ 2 \succ 4 \succ 1 \ldots$ (which leads to the "star icosahedron"). Find the genus of the surface into which the star icosahedron is embedded.

8.13. Prove that a gluing of the sides of a polygon produces the sphere if and only if the graph formed by the boundary of the polygon on the resulting surface is a tree.

8.14. Prove that the genus of a surface which can be glued from the $2n$-gon is at most $\left[\frac{n}{2}\right]$.

8.15. Write out the generating function for the numbers of gluings of $2n$-gons giving the torus.

8.16. Prove that the polynomial $T_n(N)$ is odd for n even and is even for n odd.

8.17. Describe all pairwise non-isotopic embeddings of the eight-graph into the torus.

Final and Bibliographical Remarks

There are many monographs treating generating functions as their main subject. We must mention first the two books [**GJ**] and [**S**]. Besides rich material which has serious overlap with that of the present book they also supply a lot of historical data and a huge bibliography. This is why we allowed ourselves to avoid historical remarks trying not to draw away the reader's attention. The bibliography below also is far from being complete; its main goal is to refer to publications never mentioned in monographies before.

In spite of the fact that eighty years have passed since the first edition of the book [**PS**] was published, it remains one of the best books in combinatorics and the method of generating functions. Many problems in the present book came from this one. Other problems come partly from the books [**GJ**] and [**S**] mentioned above, from other sources, or invented by myself. Enumerative problems in graph theory are discussed in [**HP**]. Among the books first published in Russian I would like to mention [**Sa**].

The approach to deducing equations for generating functions for languages generated by unambiguous formal grammars follows the paper [**DV**] (see also the references therein). The relationship between this approach and the Lagrange equation is described in detail in [**LZ1, LZ2**]. The Bernoulli–Euler triangle was introduced, and at great length, studied by V. I. Arnold [**A1, A2, A3, A4**] in connection with the investigation of various functional spaces.

All information about representation of generating functions in terms of continued fractions is taken from the papers [**F1, F2**] by P. Flajolet. The treating of the asymptotics of coefficients and its relation to the singularities of the functions follows [**FO**]. The proof of the Harer–Zagier theorem is the original one from [**HZ**], but with serious deletions. A purely combinatorial proof is given in [**L**] and a proof based on the theory of representations of finite groups can be found in [**Z**].

Bibliography

[A1] V. I. Arnold, *Bernoulli–Euler updown numbers associated with function singularities, their combinatorics and arithmetics*, Duke Math. J. **63** (1991), 537–555.

[A2] ———, *Springer numbers and morsification spaces*, J. Algebraic Geometry **1** (1992), 197–214.

[A3] ———, *Snake calculus and the combinatorics of the Bernoulli, Euler and Springer numbers of Coxeter groups*, Russian Math. Surveys **47** (1992), 1–51.

[A4] ———, *Congruences for Euler, Bernoulli and Springer numbers of Coxeter groups*, Russian Acad. Sci. Izv. Math. **41** (1993), 389–393.

[DV] M. P. Delest and G. Viennot, *Algebraic languages and polyominoes enumeration*, Theoretical Computer Science **34** (1984), 169–206.

[F1] P. Flajolet, *Combinatorial aspects of continued fractions*, Discrete Mathematics **32** (1980), 125–161.

[F2] ———, *On congruences and continued fractions for some classical combinatorial quantities*, Discrete Mathematics **41** (1982), 145–153.

[FO] P. Flajolet and A. Oldyzko, *Singularity analysis of generating functions*, SIAM J. Disc. Math. **3** (May 1990), 216–240.

[GJ] I. P. Goulden and D. M. Jackson, *Combinatorial enumeration*, John Wiley & Sons, Inc., New York, 1983.

[HP] F. Harary and E. M. Palmer, *Graphical enumeration*, Academic Press, New York–London, 1973.

[HZ] J. Harer and D. Zagier, *The Euler characteristic of the moduli space of curves*, Inv. Math. **86** (1986), 457–485.

[LZ1] S. K. Lando and A. K. Zvonkin, *Meanders*, Selecta Mathematica Sovietica **11** (1992), 117–144.

[LZ2] ———, *Plane and projective meanders*, Theoretical Computer Science **117** (1993), 227–241.

[L] B. Lass, *Démonstration combinatoire de la formule de Harer–Zagier*, C. R. Acad. Sci. Paris. Série I **333** (2001), 155–160.

[PS] G. Pólya and G. Szegö, *Problems and theorems in analysis.* I, II, Springer-Verlag, Berlin–New York, 1978, 1976.

[R] J. Riordan, *An introduction to combinatorial analysis*, Princeton University Press, Princeton, N.J., 1980.

[Sa] V. N. Sachkov, *Combinatorial methods in discrete mathematics*, Cambridge University Press, Cambridge, 1996.

[S] R. P. Stanley, *Enumerative combinatorics.* I, Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA, 1986.

[Z] D. Zagier, *On the distribution of the number of cycles of elements in symmetric groups*, Nieuw Arch. Wiskd., IV Ser. **13** (1995), 489–495.

Index

Titles in This Series

23 **S. K. Lando,** Lectures on generating functions, 2003

22 **Andreas Arvanitoyeorgos,** An introduction to Lie groups and the geometry of homogeneous spaces, 2003

21 **W. J. Kaczor and M. T. Nowak,** Problems in mathematical analysis III: Integration, 2003

20 **Klaus Hulek,** Elementary algebraic geometry, 2003

19 **A. Shen and N. K. Vereshchagin,** Computable functions, 2003

18 **V. V. Yaschenko, Editor,** Cryptography: An introduction, 2002

17 **A. Shen and N. K. Vereshchagin,** Basic set theory, 2002

16 **Wolfgang Kühnel,** Differential geometry: curves - surfaces - manifolds, 2002

15 **Gerd Fischer,** Plane algebraic curves, 2001

14 **V. A. Vassiliev,** Introduction to topology, 2001

13 **Frederick J. Almgren, Jr.,** Plateau's problem: An invitation to varifold geometry, 2001

12 **W. J. Kaczor and M. T. Nowak,** Problems in mathematical analysis II: Continuity and differentiation, 2001

11 **Michael Mesterton-Gibbons,** An introduction to game-theoretic modelling, 2000

10 **John Oprea,** The mathematics of soap films: Explorations with Maple®, 2000

9 **David E. Blair,** Inversion theory and conformal mapping, 2000

8 **Edward B. Burger,** Exploring the number jungle: A journey into diophantine analysis, 2000

7 **Judy L. Walker,** Codes and curves, 2000

6 **Gérald Tenenbaum and Michel Mendès France,** The prime numbers and their distribution, 2000

5 **Alexander Mehlmann,** The game's afoot! Game theory in myth and paradox, 2000

4 **W. J. Kaczor and M. T. Nowak,** Problems in mathematical analysis I: Real numbers, sequences and series, 2000

3 **Roger Knobel,** An introduction to the mathematical theory of waves, 2000

2 **Gregory F. Lawler and Lester N. Coyle,** Lectures on contemporary probability, 1999

1 **Charles Radin,** Miles of tiles, 1999

In combinatorics, one often considers the process of enumerating objects of a certain nature, which results in a sequence of positive integers. With each such sequence, one can associate a generating function, whose properties tell us a lot about the nature of the objects being enumerated. Nowadays, the language of generating functions is the main language of enumerative combinatorics.

This book is based on the course given by the author at the College of Mathematics of the Independent University of Moscow. It starts with definitions, simple properties, and numerous examples of generating functions. It then discusses various topics, such as formal grammars, generating functions in several variables, partitions and decompositions, and the exclusion-inclusion principle. In the final chapter, the author describes applications of generating functions to enumeration of trees, plane graphs, and graphs embedded in two-dimensional surfaces.

Throughout the book, the reader is motivated by interesting examples rather than by general theories. It also contains a lot of exercises to help the reader master the material. Little beyond the standard calculus course is necessary to understand the book. It can serve as a text for a one-semester undergraduate course in combinatorics.

ISBN 0-8218-3481-9

9 780821 834817

STML/23

For additional information and updates on this book, visit

www.ams.org/bookpages/stml-23