L1: Symbolic Method for Unlabelled Structures & Ordinary Generating Functions

Analytic Combinatorics

Yun Kuen Cheung, School of Computing, ANU

This advanced course presents the strong connections between *combinatorial* counting problems and the corresponding *analytic* tool called *generating functions*. To proceed, students should already have background knowledge in enumerative combinatorics, discrete probability theory, calculus, linear algebra and real analysis. While complex analysis is essential in establishing these connections, it is not a prerequisite. Instead, we cover the necessary complex analysis just before it is needed.

Unlabelled Combinatorial Classes and Ordinary Generating Functions

Enumerative combinatorics asks: "How many objects of size n are there in a certain set?" For example, how many binary strings of length n are there? How many complete binary trees with n internal nodes are there? Analytic combinatorics systematically approaches these questions by studying the structural definitions of these objects and translating them into equations involving functions. We begin with some motivating examples before formally defining our main objects of study.

Example 1. Consider the set of all binary strings, denoted by $\mathcal{B} = \{\epsilon, 0, 1, 00, 01, 10, 11, \ldots\}$, where ϵ is the empty string. The "size" of a string is naturally its length. Let $b_n = 2^n$ be the number of binary strings of size n. Instead of treating the sequence $(b_0, b_1, b_2, \ldots)$ as a collection of isolated numbers, we encapsulate it in a single power series:

$$B(z) \;=\; \sum_{n=0}^{\infty} b_n z^n \;=\; 1 + 2z + 4z^2 + 8z^3 + \cdots \;=\; \frac{1}{1-2z},$$

where the latter equality holds if $|z| < \frac{1}{2}$. Thus, the closed form $\frac{1}{1-2z}$ encodes the entire counting sequence $(b_n)_{n\geq 0}$ in some open neighborhood around $z = 0$.

A complete binary tree is either a single leaf or an internal node with two ordered subtrees. Let t_n be the number of complete binary trees with n internal nodes. A classical combinatorial approach using the *reflection principle* shows that t_n is the *Catalan number*: $t_n = \frac{1}{n+1}\binom{2n}{n}$. Later in this lecture, we shall use the *combinatorial specification* of complete binary trees to derive that

$$T(z) \;=\; \sum_{n=0}^{\infty} t_n z^n \;=\; \frac{1-\sqrt{1-4z}}{2z}$$

if $|z| < \frac{1}{4}$. Thus, the closed form $\frac{1-\sqrt{1-4z}}{2z}$ encodes the enumeration of complete binary trees, just as $B(z)$ does for binary strings.

We now formalize these concepts.

Definition 2. An unlabelled combinatorial class is a pair $(\mathcal{A}, |\cdot|)$, where $\mathcal{A}$ is a finite or countably infinite set and $|\cdot| : \mathcal{A} \to \mathbb{N} \cup \{0\}$ is a size function, such that the number of elements in $\mathcal{A}$ of size n is finite for any integer $n \geq 0$. Let a_n denote the number of elements in $\mathcal{A}$ of size n.

Definition 3. The ordinary generating function (OGF) of a combinatorial class $\mathcal{A}$ is the formal power series

$$A(z) \;=\; \sum_{n=0}^{\infty} a_n z^n.$$

Equivalently, it can also be written as a sum over all objects in the class: $A(z) = \sum_{\alpha \in \mathcal{A}} z^{|\alpha|}$.

Throughout this course, we use calligraphic symbols like $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{T}$ to denote combinatorial classes, their non-calligraphic counterparts like $A(z), B(z), C(z), T(z)$ to denote generating functions, and a_n, b_n, c_n, t_n to denote the respective coefficients of z^n in these functions.

The Symbolic Method

The symbolic method is a powerful framework that directly translates operations on combinatorial classes into algebraic operations on their corresponding generating functions. This translation allows us to bypass recurrence relations and construct the generating function directly from the specification of the class. We start with two fundamental building blocks.

Definition 4. The neutral class, denoted $\mathcal{E}$, consists of a single element of size 0. Its OGF is $E(z) = 1$. The atomic class, denoted $\mathcal{Z}$, consists of a single element of size 1. Its OGF is $Z(z) = z$.

In the context of binary strings from Example 1, the empty string ϵ is an element of size 0 and forms a neutral class. The single symbols 0 and 1 can each be viewed as forming atomic classes.

We build more complex classes from simpler ones using *admissible constructions*. An operation on combinatorial classes is *admissible* if the number of objects of size n in the resulting class is finite for any integer $n \geq 0$, and this number depends only on the counting sequences of the constituent classes. In this and the next lectures, we will discuss a series of admissible constructions and their applications.

Definition 5. The disjoint union of two combinatorial classes $\mathcal{A}$ and $\mathcal{B}$, denoted $\mathcal{A} + \mathcal{B}$, is the combinatorial class $\mathcal{A} \cup \mathcal{B}$ where $\mathcal{A} \cap \mathcal{B} = \emptyset$. The size of an element is its size in its original class.

The Cartesian product of two combinatorial classes $\mathcal{A}$ and $\mathcal{B}$, denoted $\mathcal{A} \times \mathcal{B}$, is the class of ordered pairs (α, β) where $\alpha \in \mathcal{A}$ and $\beta \in \mathcal{B}$. The size function is additive: $|(\alpha, \beta)| = |\alpha| + |\beta|$.

Theorem 6.
(a) If $\mathcal{C} = \mathcal{A} + \mathcal{B}$, then the OGF of $\mathcal{C}$ is $C(z) = A(z) + B(z)$.
(b) If $\mathcal{D} = \mathcal{A} \times \mathcal{B}$, then the OGF of $\mathcal{D}$ is $D(z) = A(z)B(z)$.

Proof: By definition, the number of objects of size n in $\mathcal{C}$ is $c_n = a_n + b_n$. Multiplying by z^n and summing over all $n \geq 0$ yields $C(z) = A(z) + B(z)$.

Using the object-centric definition of the OGF from Definition 3, we have

$$D(z) \;=\; \sum_{\gamma \in \mathcal{D}} z^{|\gamma|} \;=\; \sum_{(\alpha,\beta) \in \mathcal{A} \times \mathcal{B}} z^{|\alpha|+|\beta|} \;=\; \left(\sum_{\alpha \in \mathcal{A}} z^{|\alpha|}\right)\left(\sum_{\beta \in \mathcal{B}} z^{|\beta|}\right) \;=\; A(z)B(z). \qquad \square$$

The next construction captures the formation of finite sequences or lists.

Definition 7. The sequence class of a combinatorial class $\mathcal{A}$, denoted $\mathrm{SEQ}(\mathcal{A})$, is the class of all finite sequences (tuples) of elements from $\mathcal{A}$. Formally,

$$\mathrm{SEQ}(\mathcal{A}) \;=\; \mathcal{E} + \mathcal{A} + (\mathcal{A} \times \mathcal{A}) + (\mathcal{A} \times \mathcal{A} \times \mathcal{A}) + \cdots$$

The size of a sequence $(\alpha_1, \ldots, \alpha_k)$ is the sum of the sizes of its components: $\sum_{i=1}^{k} |\alpha_i|$.

For this construction to be admissible, we require that $\mathcal{A}$ contains no element of size 0 (i.e., $a_0 = 0$). Otherwise, we could form infinitely many sequences of size 0 (such as (α_0), (α_0, α_0), and so on), violating the finiteness condition in Definition 2.

Theorem 8. If $\mathcal{S} = \mathrm{SEQ}(\mathcal{A})$ and $a_0 = 0$, then the OGF of $\mathcal{S}$ is

$$S(z) \;=\; \frac{1}{1 - A(z)}.$$

Proof: Applying Theorem 6 to the definition of the sequence class, we get

$$S(z) \;=\; 1 + A(z) + A(z)^2 + A(z)^3 + \cdots$$

Since $a_0 = 0$, the lowest degree of z in $A(z)^k$ is at least k. Therefore, for any given n, the coefficient of z^n in the infinite sum depends only on the terms up to $A(z)^n$. Thus, the infinite geometric series is well-defined as a formal power series and evaluates to $S(z) = \frac{1}{1-A(z)}$. $\square$

Example 9. An *integer composition* of n is a sequence of positive integers $(p_1, p_2, \ldots, p_k)$ such that $p_1 + p_2 + \cdots + p_k = n$. Let $\mathcal{C}$ be the combinatorial class of all integer compositions, where the size of a composition is the sum of its parts.

A composition is simply a sequence of positive integers. We define the class of a single part, denoted $\mathcal{I}$, as the set of all positive integers. The size of a part k is k itself. Thus, $\mathcal{I} = \mathcal{Z}_1 + \mathcal{Z}_2 + \mathcal{Z}_3 + \cdots$, where $\mathcal{Z}_k$ represents the integer k with size k. The OGF for $\mathcal{I}$ is a geometric series:

$$I(z) \;=\; z^1 + z^2 + z^3 + \cdots \;=\; \frac{z}{1-z}.$$

Since a composition is an ordered sequence of these parts, the class of all compositions is $\mathcal{C} = \mathrm{SEQ}(\mathcal{I})$. Using the sequence construction rule, its OGF is:

$$C(z) \;=\; \frac{1}{1 - I(z)} \;=\; \frac{1}{1 - \frac{z}{1-z}} \;=\; \frac{1-z}{1-2z}.$$

We expand this to find the coefficients c_n:

$$C(z) \;=\; (1-z)\sum_{n=0}^{\infty}(2z)^n \;=\; 1 + \sum_{n=1}^{\infty} 2^{n-1} z^n.$$

Thus, for $n \geq 1$, the number of compositions of n is 2^{n-1}. This simple formula can also be established directly by considering a sequence of n balls, with each consecutive pair of balls separated by one possible divider; each of the $n-1$ dividers can either be present or absent.

Exercise 10.

(a) Derive the OGF for the combinatorial class of all integer compositions in which each part is odd.

(b) Derive the OGF for the combinatorial class of all integer compositions in which each part k can be colored in k distinct ways.

Applications of the Symbolic Method: Words, Languages and Finite Automata

With the above symbolic method operations, we can re-evaluate our motivating example and tackle more complex structures.

Example 11. Let $\mathcal{Z}_0$ and $\mathcal{Z}_1$ be the atomic classes corresponding to the symbols 0 and 1, respectively. Their OGFs are $Z_0(z) = z$ and $Z_1(z) = z$. A single bit can be viewed as an element from the class $\mathcal{A} = \mathcal{Z}_0 + \mathcal{Z}_1$. By Theorem 6(a), the OGF for a single bit is $A(z) = Z_0(z) + Z_1(z) = 2z$.

A binary string is precisely a finite sequence of bits. Thus, the class of binary strings is $\mathcal{B} = \mathrm{SEQ}(\mathcal{A})$. By Theorem 8, its OGF is

$$B(z) \;=\; \frac{1}{1 - A(z)} \;=\; \frac{1}{1 - 2z},$$

which recovers the result from Example 1 instantly.

Example 12. Consider all words over the alphabet $\{a, b, c, d\}$. Let $\mathcal{S}$ be the class of *Smirnov words*, in which no two adjacent letters are identical. For example, *abadc* is a Smirnov word, but *dabbcd* is not. Elementary counting gives $s_0 = 1$ and $s_n = 4 \cdot 3^{n-1}$ for $n \geq 1$. Hence, the OGF is

$$S(z) \;=\; 1 + \sum_{n=1}^{\infty} (4 \cdot 3^{n-1}) z^n \;=\; 1 + \frac{4}{3} \sum_{n=1}^{\infty} (3z)^n \;=\; 1 + \frac{4}{3} \cdot \frac{3z}{1 - 3z} \;=\; \frac{1+z}{1-3z}.$$

Here, we use a direct approach to derive the OGF. We will see a symbolic method approach in Example 26.

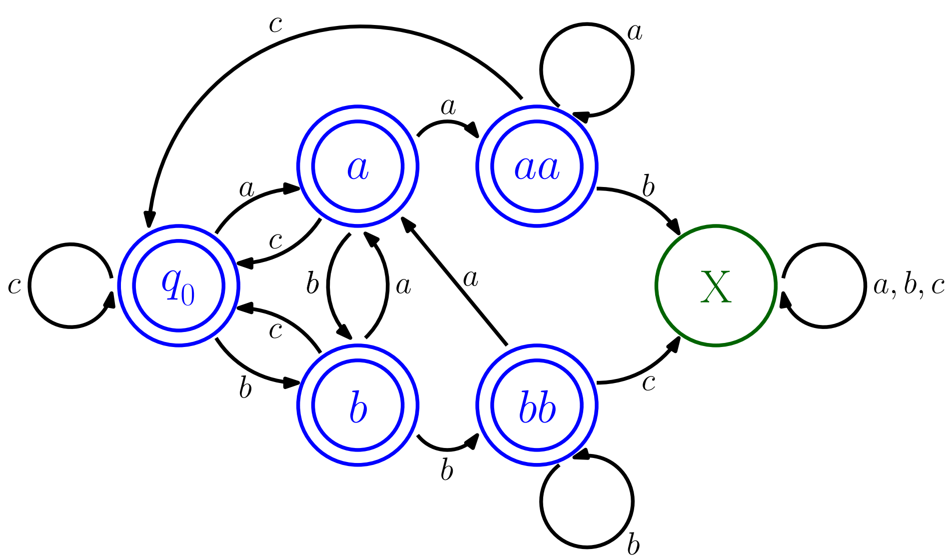
Example 12 is a classic example of words constructed under a certain constraint. More generally, a *language* is any class of words over a given alphabet. A particularly well-behaved family of languages consists of those recognized by a deterministic finite automaton.

Definition 13. A deterministic finite automaton (DFA) is a 5-tuple $(Q, \Sigma, \delta, q_0, F)$, where:

- Q is a finite set of states,
- Σ is a finite alphabet,
- $\delta : Q \times \Sigma \to Q$ is the transition function,
- $q_0 \in Q$ is the start state,
- $F \subseteq Q$ is the set of accepting states.

A word $w = w_1 w_2 \dots w_n$ is accepted by the DFA if the sequence of states $r_0, r_1, \dots, r_n$ defined by $r_0 = q_0$ and $r_i = \delta(r_{i-1}, w_i)$ terminates in an accepting state, i.e., $r_n \in F$. The language recognized by the DFA is the class of all words it accepts.

Example 14. Let $\Sigma = \{a, b, c\}$ and let $\mathcal{L}$ be the language of words with no occurrence of the consecutive substrings "aab" or "bbc". A DFA for $\mathcal{L}$ is given in the figure below. The state labeled q_0 is the start state. All accepting states are marked blue, while the green state X is the only non-accepting state. The four middle states are labeled by their suffixes.



The symbolic method provides a powerful way to derive the generating function of any language recognized by a DFA. We can translate the automaton directly into a system of algebraic equations.

Theorem 15. Let $\mathcal{L}$ be a language recognized by a DFA $(Q, \Sigma, \delta, q_0, F)$. For each state $q \in Q$, let $L_q(z)$ be the OGF of the class of words that take the automaton from state q to an accepting state. These generating functions satisfy the system of linear equations:

$$L_q(z) \;=\; f_q + z \sum_{a \in \Sigma} L_{\delta(q,a)}(z), \tag{1}$$

where $f_q = 1$ if $q \in F$ and 0 otherwise. The OGF of the language $\mathcal{L}$ is $L_{q_0}(z)$, which is a rational function.

Proof: Let $\mathcal{L}_q$ be the class of words that take the automaton from state q to an accepting state. A word in $\mathcal{L}_q$ is either the empty string ϵ (which is only valid if $q \in F$, contributing $\mathcal{E}$), or it begins with some letter $a \in \Sigma$, forcing the DFA to transition to the state $p = \delta(q, a)$. The remainder of the word must then take the automaton from p to an accepting state, which is precisely an element of $\mathcal{L}_p$.

Symbolically, this recursive structure is expressed as a disjoint union:

$$\mathcal{L}_q \;=\; (\text{if } q \in F \text{ then } \mathcal{E} \text{ else } \emptyset) + \sum_{a \in \Sigma} \mathcal{Z}_a \times \mathcal{L}_{\delta(q,a)}.$$

Translating this to generating functions using Theorem 6 yields the linear system depicted by Equation (1). Let $\mathbf{v} = (L_q(z))_{q \in Q}$, let $\mathbf{f} = (f_q)_{q \in Q}$, and let $\mathbf{M} = (M_{q,p})_{q,p \in Q}$ with $M_{q,p} = |\{a \in \Sigma : \delta(q, a) = p\}|$. Then the linear system can be rewritten as $(\mathbf{I} - z\mathbf{M})\mathbf{v} = \mathbf{f}$. Observe that the matrix $(\mathbf{I} - z\mathbf{M})$ has constant term equal to the identity matrix. (See Example 16 below for a concrete example.) Hence its determinant has constant term 1 and is nonzero as a formal power series. Cramer's rule therefore guarantees that every component of the solution, including $L_{q_0}(z)$, is a rational function of z. $\square$

Example 16. We apply Theorem 15 to the DFA in Example 14. The accepting states are $F = \{q_0, a, aa, b, bb\}$. Since the "dead" state X is not an accepting state and cannot be exited, its generating function is $L_X(z) = 0$ and we omit it from our equations. The transitions between the valid states give the following system of equations:

$$\begin{aligned}
L_{q_0}(z) &= 1 + zL_{q_0}(z) + zL_a(z) + zL_b(z) \\
L_a(z) &= 1 + zL_{q_0}(z) + zL_{aa}(z) + zL_b(z) \\
L_{aa}(z) &= 1 + zL_{q_0}(z) + zL_{aa}(z) \\
L_b(z) &= 1 + zL_{q_0}(z) + zL_a(z) + zL_{bb}(z) \\
L_{bb}(z) &= 1 + zL_a(z) + zL_{bb}(z)
\end{aligned}$$

This system can be rewritten as

$$\left(\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} - z \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} L_{q_0}(z) \\ L_a(z) \\ L_{aa}(z) \\ L_b(z) \\ L_{bb}(z) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

We solve it using Cramer's rule:

$$L_{q_0}(z) = \frac{1}{1 - 3z + 2z^3 - z^5}.$$

Exercise 17. Construct a DFA that recognizes the Smirnov words specified in Example 12. By using Theorem 15, confirm that the OGF is $\frac{1+z}{1-3z}$.

Not all structurally interesting languages are recognizable by a deterministic finite automaton. A classic example is the language of well-formed parentheses given below.

Example 18. Let $\mathcal{D}$ be the class of well-formed strings of matching parentheses. For example, `()`, `(())`, and `()()` are in $\mathcal{D}$, but `)(` and `(()` are not. Let the size of a string be the number of pairs of parentheses. The atomic class $\mathcal{Z}$ here corresponds to a single pair.

A non-empty matching parenthesis string can always be uniquely decomposed into two parts by finding the matching closing parenthesis for the very first opening parenthesis. Thus, any string in $\mathcal{D}$ is either empty, or it has the form '(u) v', where u and v are themselves (possibly empty) matching parenthesis strings. Symbolically, this translates to:

$$\mathcal{D} = \mathcal{E} + \mathcal{Z} \times \mathcal{D} \times \mathcal{D}.$$

Translating this into generating functions gives a quadratic equation:

$$D(z) = 1 + zD(z)^2.$$

Solving this quadratic equation for $D(z)$ using the quadratic formula yields two possible branches:

$$D(z) = \frac{1 \pm \sqrt{1-4z}}{2z}.$$

As a formal power series, $D(z)$ must evaluate to a finite constant d_0 at $z = 0$. The positive branch is singular at $z = 0$, whereas the negative branch yields $\lim_{z\to 0} \frac{1-\sqrt{1-4z}}{2z} = 1$ by L'Hôpital's rule. This matches the condition $D(0) = 1$ (since the empty string is the unique valid string of size 0), so we must choose the negative sign: $D(z) = \frac{1-\sqrt{1-4z}}{2z}$. By Theorem 15, any DFA for $\mathcal{D}$ would make the word-length OGF $D(u^2) = \frac{1-\sqrt{1-4u^2}}{2u^2}$ rational; it is not, so no such DFA exists.

To extract the coefficients of this generating function, we apply Newton's generalized binomial theorem, $(1+x)^r = \sum_{n=0}^{\infty} \binom{r}{n} x^n$, to $(1-4z)^{1/2}$:

$$(1-4z)^{1/2} \;=\; \sum_{n=0}^{\infty} \binom{1/2}{n} (-4z)^n.$$

For $n \geq 1$, the generalized binomial coefficient can be expanded as follows:

$$\binom{1/2}{n} \;=\; \frac{\frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\cdots\left(\frac{1}{2}-n+1\right)}{n!} \;=\; \frac{(-1)^{n-1} 1\cdot 3\cdots(2n-3)}{2^n n!}.$$

Multiplying by $(-4)^n = (-1)^n 2^{2n}$ gives:

$$\binom{1/2}{n}(-4)^n \;=\; -\frac{2^n\cdot 1\cdot 3\cdots(2n-3)}{n!} \;=\; -\frac{2^n n!\cdot 1\cdot 3\cdots(2n-3)}{n!n!} \;=\; -\frac{2}{n}\binom{2n-2}{n-1}.$$

Therefore, the coefficient of z^n in $D(z)$ is the n-th *Catalan number*:

$$[z^n]D(z) \;=\; [z^{n+1}]\frac{1-\sqrt{1-4z}}{2} \;=\; -\frac{1}{2}[z^{n+1}](1-4z)^{1/2} \;=\; \frac{1}{n+1}\binom{2n}{n}.$$

Applications of the Symbolic Method: Trees

Recursive structures, such as trees, are ideally suited for the symbolic method. A tree is typically defined as a root node attached to subtrees that belong to the same class. This self-referential definition translates directly into a functional equation for the generating function. We present several fundamental varieties of trees and their counting results below.

Example 19. Consider the class $\mathcal{T}$ of *complete binary trees*. A complete binary tree is either a single external node (a leaf), or an internal node with a left complete binary subtree and a right complete binary subtree. We define the size of the tree as the number of its *internal* nodes.

Under this size function, an external node has size 0, corresponding to the neutral class $\mathcal{E}$. An internal node has size 1, corresponding to the atomic class $\mathcal{Z}$. A tree that is not a leaf consists of an internal node and an ordered pair of subtrees (left and right), which corresponds to the Cartesian product $\mathcal{Z} \times \mathcal{T} \times \mathcal{T}$. Therefore, the symbolic specification of $\mathcal{T}$ is:

$$\mathcal{T} \;=\; \mathcal{E} + \mathcal{Z} \times \mathcal{T} \times \mathcal{T}.$$

This specification is exactly the same as the one in Example 18. By following the logic in that example, the counting sequence is again the Catalan numbers.

Example 20. A *rooted plane tree* (or *ordered tree*) is defined recursively: it consists of a single root node to which a finite ordered sequence of rooted plane trees is attached. The size of a tree is its number of nodes (including internal nodes and leaves). Let $\mathcal{T}$ be the class of rooted plane trees.

Because the sequence of subtrees could be empty, a single node is a valid tree. The root corresponds to the atomic class $\mathcal{Z}$. The sequence of subtrees belongs to the class $\mathrm{SEQ}(\mathcal{T})$. A tree is formed by taking a root and attaching this sequence, which corresponds to the Cartesian product. Therefore, the symbolic specification of $\mathcal{T}$ is:

$$\mathcal{T} \;=\; \mathcal{Z} \times \mathrm{SEQ}(\mathcal{T}).$$

Translating this specification to generating functions using our dictionary, we obtain the functional equation for the OGF $T(z)$:

$$T(z) \;=\; z \cdot \frac{1}{1 - T(z)}.$$

Rearranging gives the quadratic equation $T(z)^2 - T(z) + z = 0$. Solving for $T(z)$ and choosing the branch that satisfies $T(0) = 0$, we find

$$T(z) \;=\; \frac{1 - \sqrt{1 - 4z}}{2}.$$

Expanding as in Example 18 yields, for $n \geq 1$, $[z^n]T(z) = \frac{1}{n}\binom{2n-2}{n-1}$.

Example 21. A *unary-binary tree* (or *Motzkin tree*) is a rooted plane tree where each node has 0, 1 or 2 children. A unary node has a single child with no separate left/right choice, while a binary node has an ordered pair of children. Thus, each node is either a leaf, or it has one child that is a unary-binary tree, or two ordered children that are unary-binary trees. If we define the size as the total number of nodes, the symbolic specification is:

$$\mathcal{U} \;=\; \mathcal{Z} + (\mathcal{Z} \times \mathcal{U}) + (\mathcal{Z} \times \mathcal{U}^2).$$

This translates to the quadratic equation $U(z) \;=\; z + zU(z) + zU(z)^2$. Solving for $U(z)$ and choosing the branch that satisfies $U(0) = 0$, we find:

$$U(z) \;=\; \frac{1 - z - \sqrt{1 - 2z - 3z^2}}{2z}.$$

The coefficient $[z^{n+1}]U(z)$ is called the n-th *Motzkin number*.

Exercise 22. Consider the class of rooted plane trees.

(a) Suppose the size of a tree is its number of internal nodes. Apply the symbolic method directly and write down the apparent equation for the OGF. What goes wrong with this attempt?

(b) Now consider the number of leaves as the size instead. Apply the symbolic method directly and write down the apparent equation for the OGF. What goes wrong with this attempt?

(c) Keep the number of leaves as the size, but impose the restriction that every internal node has at least two children. Provide the symbolic specification, and solve it to derive the OGF.

Exercise 23. A *Schröder path* of size n is a lattice path from $(0,0)$ to (n,n) using horizontal steps $H=(1,0)$, vertical steps $V=(0,1)$ and diagonal steps $D=(1,1)$, that never goes above the diagonal line $y=x$. Let $\mathcal{S}$ be the class of all such paths.

(a) Show that its OGF satisfies

$$S(z) \;=\; 1 + zS(z) + zS(z)^2.$$

(b) Solve this equation for $S(z)$ and determine the first few coefficients.

Substitution

The substitution of two combinatorial classes allows us to replace the *atoms* of one class with objects from another.

Definition 24. Let $\mathcal{A}$ and $\mathcal{B}$ be combinatorial classes such that $b_0 = 0$. The substitution of $\mathcal{B}$ into $\mathcal{A}$, denoted $\mathcal{A}\circ\mathcal{B}$ or $\mathcal{A}(\mathcal{B})$, is the class obtained by replacing each atom of an object in $\mathcal{A}$ by an object from $\mathcal{B}$. The size of the resulting object is the sum of the sizes of the inserted $\mathcal{B}$-objects.

In this lecture, we use this operation only when the atoms of each object of $\mathcal{A}$ serve as distinguished replacement slots. For unlabelled structures with non-trivial symmetries among atoms, this simple substitution rule may fail.

Theorem 25. If $\mathcal{C} = \mathcal{A}\circ\mathcal{B}$ and $b_0 = 0$, then the OGF of $\mathcal{C}$ is

$$C(z) \;=\; A(B(z)).$$

Proof: Since the substitution replaces each distinguished slot in an object $\alpha \in \mathcal{A}$ (with $|\alpha| = k$) by an object from $\mathcal{B}$, this corresponds to forming k-tuples of elements from $\mathcal{B}$. The OGF of $\mathcal{B}^k$ is $B(z)^k$. Since $b_0 = 0$, $B(z)^k$ has degree at least k, so the composition is well-defined. Summing over all $\alpha \in \mathcal{A}$, we obtain

$$C(z) \;=\; \sum_{\alpha\in\mathcal{A}} B(z)^{|\alpha|} \;=\; \sum_{k=0}^{\infty} a_k B(z)^k \;=\; A(B(z)). \qquad \square$$

Example 26. We presented the OGF for Smirnov words in Example 12. Here, we present an alternative approach to deriving $S(z)$. Let $\mathcal{W} = \mathrm{SEQ}(\mathcal{Z}_a + \mathcal{Z}_b + \mathcal{Z}_c + \mathcal{Z}_d)$ be the class of all words, so $W(z) = \frac{1}{1-4z}$. Any word in $\mathcal{W}$ consists of a block of identical letters, followed by another block of a different letter, and so on. By collapsing each such block into just one letter, we get a Smirnov word. For instance, *bbdddacadddd* is collapsed into *bdacad*.

Conversely, any word in $\mathcal{W}$ can be formed by starting from a Smirnov word and substituting each single letter with a non-empty block of that same letter. For each fixed letter, this block class has OGF $\frac{z}{1-z}$; for example, $\mathrm{SEQ}_{\geq 1}(\mathcal{Z}_a) = \mathcal{Z}_a \times \mathrm{SEQ}(\mathcal{Z}_a)$. Thus, $W(z) = S\left(\frac{z}{1-z}\right)$. Set $u = \frac{z}{1-z}$, which implies $z = \frac{u}{1+u}$. Since $W(z) = \frac{1}{1-4z}$, we have

$$S(u) \;=\; \frac{1}{1 - 4\cdot\frac{u}{1+u}} \;=\; \frac{1+u}{1-3u},$$

recovering the OGF we derived in Example 12 by elementary means.

Example 27. Consider the class $\mathcal{W}^{[<4]}$ of words over the alphabet $\{a, b, c, d\}$ without any run of four consecutive identical letters (e.g., no $aaaa$). We can construct these words by starting with a Smirnov word over this alphabet and substituting each letter with a block of the same letter of length at most 3.

From Example 26, we know the OGF for the class $\mathcal{S}$ of Smirnov words over this four-letter alphabet is: $S(z) = \frac{1+z}{1-3z}$. We substitute each letter in the Smirnov word with a non-empty block of identical letters of length at most 3, represented by the class $\mathcal{B}_{\leq 3} = \mathcal{Z} + \mathcal{Z}^2 + \mathcal{Z}^3$ with OGF $B_{\leq 3}(z) = z + z^2 + z^3$. Since the substitution class has no elements of size 0, the operation is admissible. Thus, $\mathcal{W}^{[<4]} = \mathcal{S} \circ \mathcal{B}_{\leq 3}$. By Theorem 25, the OGF of $\mathcal{W}^{[<4]}$ is

$$W^{[<4]}(z) \;=\; S(B_{\leq 3}(z)) \;=\; \frac{1 + (z + z^2 + z^3)}{1 - 3(z + z^2 + z^3)} \;=\; \frac{1 + z + z^2 + z^3}{1 - 3z - 3z^2 - 3z^3}.$$

Exercise 28. Find the OGF of the class of words over the alphabet $\{a, b, c, d, e\}$ in which the length of each block (as described in Example 26) is odd.

Exercise 29. Show that the OGF for words over $\{a, b, c\}$ in which every run of a's or b's has length at most 3, while every run of c's has length at most 2, is

$$\frac{(1 + z + z^2 + z^3)(1 + z + z^2)}{1 - z - 3z^2 - 5z^3 - 4z^4 - 2z^5}.$$

One possible approach is via the bivariate OGF $S(x, y)$ for Smirnov words over $\{a, b, c\}$, where each occurrence of either a or b contributes a factor x, while each occurrence of c contributes a factor y.

Example 30. We revisit the class $\mathcal{U}$ of unary-binary trees from Example 21 to demonstrate the power of substitution.

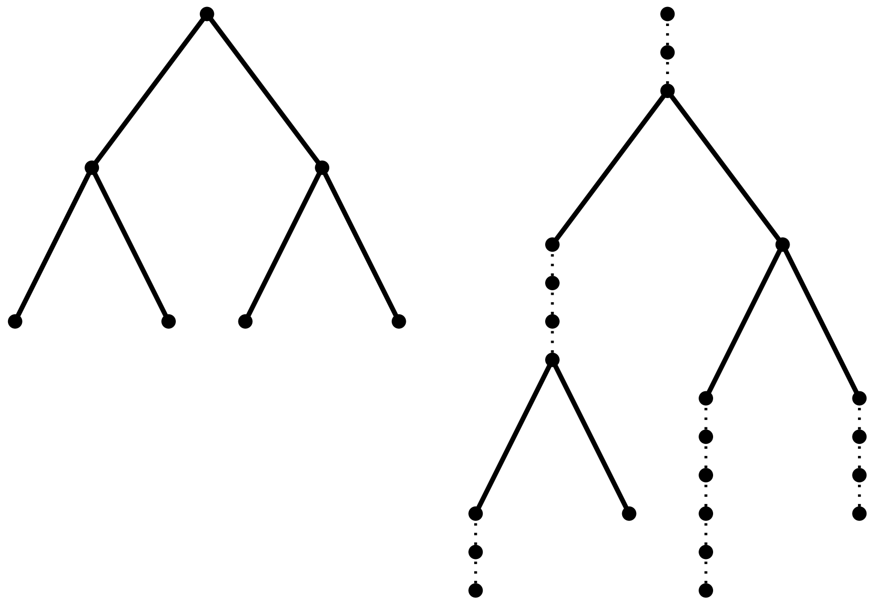
Consider the class $\mathcal{B}$ of complete binary trees, for which the size of a tree is its number of nodes. Recall that every node has exactly 0 or 2 children. Its symbolic specification is $\mathcal{B} = \mathcal{Z} + \mathcal{Z} \times \mathcal{B}^2$, which translates to $B(z) = z + zB(z)^2$. Solving this quadratic equation and choosing the branch with $B(0) = 0$ yields

$$B(z) \;=\; \frac{1 - \sqrt{1 - 4z^2}}{2z}.$$

A key observation is that any unary-binary tree can be uniquely formed from a complete binary tree by replacing each of its nodes with a downward chain containing at least one node. See the figure next page for an example. In each chain, all but the last node have exactly 1 child, and the last node adopts the children of the original node from the complete binary tree (i.e., 0 or 2 children). The class of such non-empty chains is $\mathcal{S} = \mathrm{SEQ}_{\geq 1}(\mathcal{Z})$, with OGF $S(z) = \frac{z}{1-z}$.

Substituting these chains into the complete binary tree gives $\mathcal{U} = \mathcal{B} \circ \mathcal{S}$. Since $\mathcal{S}$ contains no elements of size 0, the substitution is admissible. By Theorem 25, the OGF of $\mathcal{U}$ is:

$$U(z) \;=\; B(S(z)) \;=\; \frac{1 - \sqrt{1 - 4 \cdot S(z)^2}}{2 \cdot S(z)} \;=\; \frac{1 - \sqrt{1 - 4\left(\frac{z}{1-z}\right)^2}}{2 \cdot \frac{z}{1-z}} \;=\; \frac{1 - z - \sqrt{1 - 2z - 3z^2}}{2z}.$$



Example 31. Consider the class $\mathcal{T}$ of complete 2-3 trees in which all leaves are at the same depth, and the size of a tree is its number of leaves. A tree in $\mathcal{T}$ is either a single leaf (represented by the atomic class $\mathcal{Z}$) or it has depth $h \geq 1$. A tree of depth h can be uniquely constructed by taking a tree of depth $h-1$ and substituting each of its leaves with a root node connected to either 2 or 3 leaves.

Let $\mathcal{B} = \mathcal{Z}^2 + \mathcal{Z}^3$ be the class of a single root with 2 or 3 children. The operation of substituting $\mathcal{B}$ into the leaves of any tree in $\mathcal{T}$ uniformly increases the depth of all leaves by 1. Thus, the class of all such trees of depth at least 1 is exactly $\mathcal{T} \circ \mathcal{B}$. Including the base case of a single leaf, we obtain the symbolic specification:

$$\mathcal{T} \;=\; \mathcal{Z} + \mathcal{T} \circ (\mathcal{Z}^2 + \mathcal{Z}^3).$$

By Theorem 25, this specification translates to the functional equation for the OGF $T(z)$:

$$T(z) \;=\; z + T(z^2 + z^3).$$

Iterating this equation reveals that $T(z) = \sum_{h=0}^{\infty} P_h(z)$, where the polynomials $P_h(z)$ corresponding to trees of depth h are defined recursively by $P_0(z) = z$ and $P_h(z) = P_{h-1}(z^2 + z^3)$ for $h \geq 1$. Unwinding the first few levels gives

$$T(z) \;=\; z + z^2 + z^3 + z^4 + 2z^5 + 2z^6 + 3z^7 + 4z^8 + 5z^9 + 8z^{10} + 14z^{11} + 23z^{12} + \mathcal{O}(z^{13}).$$