Analytic Combinatorics

Marni Mishna

Department of Mathematics
Simon Fraser University
Burnaby Canada

EJCIM • “Bordeaux”, France

Motivating question #1

In a given regular language over $\{0, 1\}$, what is the proportion of words of length n that have the same number of 0s and 1s? (for n large) What about an alphabet of size k? What about for a context-free language?

Motivating question #2

Which proportion of sequences of n "king" chess moves on $\mathbb{Z}^2$ start and end at the origin, and stay in $\mathbb{N}^2$? (3-D version? arbitrary dimension? other possible moves? regions?)

A Course on Analytic Combinatorics

Objective

Develop tools for analysing the large scale behaviour of combinatorial classes in a **systematic** manner.

Strategy

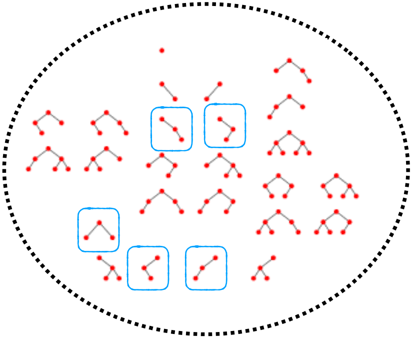
Examine singularities of multivariable combinatorial generating functions and develop a geometric understanding of its singular structure and then deduce the asymptotic expressions for counting sequences.

Organization

- I. Combinatorial Framework
  - Combinatorial Calculus
  - Parameters and Extracted Classes
- II. Introductory Singularity Analysis & Asymptotic Expressions
  - Univariate case
  - Multivariate case

I. Combinatorial Functional Equations

Combinatorial Classes



tree $\mapsto z^{\#\text{nodes}}$

$C(z) = z + 2z^2 + 5z^3 + 14z^4 + 42z^5$

A **class** is a set $\mathcal{C}$, and size $|\cdot|$. The number of elements of a given size is finite.

$$C(z) := \sum_{\gamma \in \mathcal{C}} z^{|\gamma|}$$

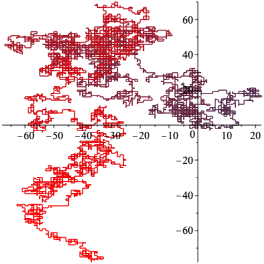
$$= \sum_{n \geq 0} c_n z^n$$

$c_n = \#$ objects of size n

$C(z)$ is the ordinary generating function (OGF) for $\mathcal{C}$

Immediate Objectives

- $c_n = ?$ Enumerate objects exactly or asymptotically
- Understand the large scale behaviour of the objects in a class
- Interpret functional equations combinatorially
  - Combinatorial understanding of solutions to linear differential equations?
    See study of holonomic functions
  - Determine analytic criteria for combinatorial hierarchies (eg. structure of recursively enumerable languages)



Everything is non-holonomic unless it is holonomic by design.

Flajolet, Gerhold and Salvy

Combinatorial Calculus

	$\mathcal{C}$	Notes	$C(z) = \sum z^{\lvert\gamma\rvert}$
Epsilon	$\{\epsilon\}$	$\lvert\epsilon\rvert = 0$	1
Atom	$\{\circ\}$	$\lvert\circ\rvert = 1$	z
Disjoint Union	$\mathcal{A} + \mathcal{B}$	$\gamma \times \epsilon_{\mathcal{A}}, \gamma \times \epsilon_{\mathcal{B}}$	$A(z) + B(z)$
Cartesian Product	$\mathcal{A} \times \mathcal{B}$	$(\alpha, \beta), \alpha \in \mathcal{A}, \beta \in \mathcal{B}$	$A(z)B(z)$
Power	$\mathcal{A}^k$	$(\alpha_1, \ldots, \alpha_k), \alpha_i \in \mathcal{A}$	$A(z)^k$
Sequence	$\mathrm{Seq}(\mathcal{A}) = \mathcal{A}^*$	$\epsilon + \mathcal{A} + \mathcal{A}^2 + \mathcal{A}^3 + \ldots$	$\frac{1}{1-A(z)}$

Binary Words $\{\epsilon, \circ, \bullet, \circ\circ, \circ\bullet, \bullet\circ, \bullet\bullet, \circ\circ\circ, \ldots\}$

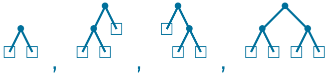
$$
\begin{array}{cccccccccc}
\{\circ\} & \{\bullet\} & & \mathcal{A} & = & \{\circ, \bullet\} & \mathcal{C} & = & \mathcal{A}^* \\
\downarrow & \downarrow & & \downarrow & & \downarrow & \downarrow & & \downarrow \\
z & z & & A(z) & = & 2z & C(z) & = & \frac{1}{1-A(z)}
\end{array}
$$

$$\Longrightarrow \quad C(z) = \frac{1}{1-2z}$$

$$\Longrightarrow \quad C(z) = \sum 2^n z^n \quad \Longrightarrow \quad c_n = 2^n$$

Combinatorial Calculus

	$\mathcal{C}$	Notes	$C(z) = \sum z^{\lvert\gamma\rvert}$
Epsilon	$\{\epsilon\}$	$\lvert\epsilon\rvert = 0$	1
Atom	$\{\circ\}$	$\lvert\circ\rvert = 1$	z
Disjoint Union	$\mathcal{A} + \mathcal{B}$	$\gamma \times \epsilon_{\mathcal{A}}, \gamma \times \epsilon_{\mathcal{B}}$	$A(z) + B(z)$
Cartesian Product	$\mathcal{A} \times \mathcal{B}$	$(\alpha, \beta), \alpha \in \mathcal{A}, \beta \in \mathcal{B}$	$A(z)B(z)$
Power	$\mathcal{A}^k$	$(\alpha_1, \dots, \alpha_k), \alpha_i \in \mathcal{A}$	$A(z)^k$
Sequence	$\mathrm{Seq}(\mathcal{A}) = \mathcal{A}^*$	$\epsilon + \mathcal{A} + \mathcal{A}^2 + \mathcal{A}^3 + \dots$	$\frac{1}{1-A(z)}$

Binary Trees $\mathcal{B} := \{\square,$  $, \dots\}$

$$\begin{array}{ccccccccccc}
\{\bullet\} & \{\square\} & & \mathcal{B} & \equiv & \square & + & \bullet & \times & \mathcal{B}^2 \\
\downarrow & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
1 & z & & B(z) & = & z & + & 1 & \cdot & B(z)^2
\end{array}$$

$$\Longrightarrow \quad B(z) = \frac{1-\sqrt{1-4z}}{2z}$$

Specifications

Generically we specify a combinatorial class by a set of combinatorial equations (like we have just seen):

$$\begin{aligned} \mathcal{C}_1 &= \Phi_1(\mathcal{Z}, \mathcal{C}_1, \ldots, \mathcal{C}_r) \\ &\vdots \\ \mathcal{C}_r &= \Phi_r(\mathcal{Z}, \mathcal{C}_1, \ldots, \mathcal{C}_r). \end{aligned} \tag{1}$$

.... and deduce a system of functional equations satisfied by the generating functions:

$$\begin{aligned} C_1(z) &= \Phi_1(z, C_1(z), \ldots, C_r(z)) \\ &\vdots \\ C_r(z) &= \Phi_r(z, C_1(z), \ldots, C_r(z)). \end{aligned} \tag{2}$$

Cyclic dependencies change the nature of the generating function.

Acyclic Dependencies: S-regular classes

Combinatorial classes specified using $+, \times, *$, Atoms, and Epsilons with **no** cyclic dependencies are **S-regular classes.**

$$\begin{aligned}\mathcal{L} &= \mathrm{Seq}(\{0\} + (\{1\} \times \mathrm{Seq}(\{0\} \times \mathrm{Seq}(\{1\}) \times \{0\}) \times \{1\})) \\ &= (0 + (1(01^*0)^*1))^* = \{\epsilon, 0, 00, 11, 000, 011, 1001, 10101, \ldots\}\end{aligned}$$

$$L(z) = \frac{1}{1 - \left(z + z\frac{1}{1 - z\frac{1}{1-z}z}\right)}$$

Theorem

The generating function of an S-regular class is a **rational** *function.*

Remark: Not all rational functions Taylor series in $\mathbb{N}[[z]]$ arise this way. ($\exists$ singularity criteria)

Cyclic Dependencies: Algebraic Classes

Well defined combinatorial classes specified using $+$, $\times$, Atoms, and Epsilons (using possibly cyclic dependencies) are **algebraic classes.**

$$\text{eg. } \mathcal{B} \equiv \square + \bullet \times \mathcal{B} \times \mathcal{B}$$

Theorem

The generating function of an algebraic class is an **algebraic** *function. (It satisfies a system of polynomial relations)*

$$P(x,y) = x + xy^2 - y \implies P(x, B(x)) = 0$$

Criterion: If a class has a transcendental OGF, it is **not** an algebraic class.

Remark: Not all algebraic functions with series in $\mathbb{N}[[z]]$ arise this way. ($\exists$ asymptotic criteria)

Derivation Tree

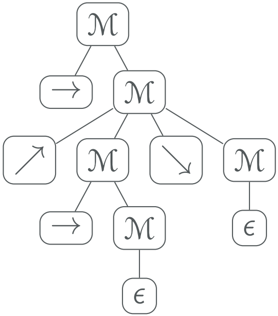
The history of rules expanded is encoded in derivation tree.
We identify derivation trees and elements

Motzkin Paths $\mathcal{M}$

Walks with steps $\{\nearrow, \searrow, \rightarrow\}$ confined to the upper half plane.

$$\mathcal{M} \equiv \epsilon + \rightarrow \mathcal{M} + \nearrow \mathcal{M} \searrow \mathcal{M}.$$

$$\rightarrow \nearrow \rightarrow \searrow \in \mathcal{M}$$



This is just the start – we can define other combinatorial operators using sets, cycles, labellings, ... and determine similar calculus. Covers permutation classes, functional graphs, three connected labelled planar graphs. The book Analytic Combinatorics of Flajolet and Sedgewick is your best resource for a deep study.

We will go another way, however.

Combinatorial Parameters

Combinatorial Parameters

A **parameter** of a class is a map $\chi : \mathcal{C} \to \mathbb{Z}$
e.g. $\# \to$ steps ; # leaves in a tree ; end position of a walk

$$C(u,z) := \sum_{\gamma\in\mathcal{C}} u^{\chi(\gamma)} x^{|\gamma|} = \sum_{n\geq 0} \left(\sum_{k\in\mathbb{Z}} c_{k,n} u^k \right) z^n.$$

$c_{k,n} = \#$ objects of size n with parameter value k.
$C(u,z) \in \mathbb{N}[u,u^{-1}][[z]]$ Power series with Laurent polynomial coefficients

Example

$\chi(w) = |w|_\circ = \#\ \circ$s a word in $\{\circ, \bullet\}^*$: $\chi(\circ\bullet\bullet\circ\bullet) = 2$

$$C(u,z) = 1 + (u+1)\,z + \left(u^2 + 2\,u + 1\right) z^2 + \left(u^3 + 3u^2 + 3u + 1\right) z^3 + \dots.$$

$$C(u,z) = \left(\frac{1}{1-(z+uz)} \right). \tag{3}$$

Treat it as a 2-dimensional parameter: $w \mapsto (\chi(w), |w|)$.

Inherited parameters

The d-dimensional parameter χ is inherited from ξ and ζ if, and only if ...

$\mathcal{C} = \mathcal{A} + \mathcal{B}$

$$\chi(\gamma) = \begin{cases} \xi(\gamma) & \gamma \in \mathcal{A} \\ \zeta(\gamma) & \gamma \in \mathcal{B} \end{cases}$$

$$\implies C(z_1, \ldots, z_d) = C_\chi(\mathbf{z}) = A_\xi(\mathbf{z}) + B_\zeta(\mathbf{z})$$

$\mathcal{C} = \mathcal{A} \times \mathcal{B}$

$$\chi(\alpha, \beta) = \xi(\alpha) + \zeta(\beta).$$

$$\implies C(z_1, z_2, \ldots, z_d) = C_\chi(\mathbf{z}) = A_\xi(\mathbf{z}) B_\zeta(\mathbf{z})$$

e.g. $\mathcal{C} = \mathcal{A}^*; \mathcal{A} = \{\circ, \bullet\}; \chi(w) = (|w|_\circ, |w|)$.
Systematic translation of structural parameters to OGF

Derived Classes

Derived Classes

Given a class $\mathcal{C}$, multidimensional inherited parameter $\chi : \mathcal{C} \to \mathbb{Z}^d$, and $d-1$-dimensional vector r, define a derived class of $\mathcal{C}$ as a class

$$\mathcal{C}^{\chi,r} = \bigcup_n \{\gamma \in \mathcal{C} \mid \chi(\gamma) = (r_1 n, \ldots, r_{d-1} n, n)\}$$

Fixed Value

If $r = (0, 0, \ldots, 0)$, then $\mathcal{C}^{\chi,r}$ is the subset of objects where the parameter is always zero. Its OGF is the constant term of $C(\mathbf{z})$ with respect to $z_1, \ldots, z_{d-1}$.

Balanced Subclasses

For χ counts occurrences of subobjects, consider $r = (1, 1, \ldots, 1)$: The subobjects occur equally.

After two examples, we consider to how find the generating functions of derived classes.

Balanced word classes

$$\mathcal{L} = \{\text{binary expansions of } n \mid n \equiv 0 \mod 3.\}$$ Size = length of string

$$\mathcal{L} = \{\epsilon, \overset{0}{0}, 00, 000, \ldots, \overset{3}{11}, 011, 0011, \ldots, \overset{6}{110}, 0110, 00110, \ldots,$$
$$\overset{9}{1001}, 01001, \overset{12}{1100}, 01100, \ldots, \overset{15}{1111}, 01111, \ldots\}$$

S-regular specification: $\mathcal{L} = (0 + (1(01^*0)^*1))^*$

Parameter: $\chi(w) = (|w|_0, |w|_1, |w|) = (\#0s \text{ in } w, \#1s \text{ in } w, |w|)$

Balanced sub-class:

$$\begin{aligned}
\mathcal{L}_= &= \{w \in \mathcal{L} \mid \chi(w) = (n, n, 2n), n \geq 0\} \\
&= \{w \in \mathcal{L} \mid \#0s = \#1s\} \\
&= \{1001, 0011, 0110, 1100, 010101, 101010, 11100001, 10011001, \\
&\qquad 10000111, 00101101, 01011010, 00111001, 00100111, \ldots\}
\end{aligned}$$

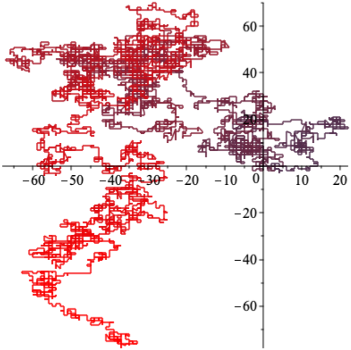
more interesting: $\mathcal{L} \subseteq \{a_1, a_2, \ldots, a_d\}^*$ with $\chi_i(w) = \#$ of a_i in w.

Excursions

$S = \{\uparrow, \downarrow, \leftarrow, \rightarrow\} = \maltese$ is a set of steps. Consider walks starting at $(0,0)$ taking steps from S. Unrestricted walks are S-regular:

$$\{\uparrow, \downarrow, \leftarrow, \rightarrow\}^*$$

Define parameter $\chi(w) := $ (endpoint of w, # of steps).



Endpoint is an inherited parameter

$$\sum walk_{\mathbb{Z}^2}^{\maltese}((0,0) \xrightarrow{n} (k,\ell))\, x^k y^\ell t^n = \frac{1}{1 - t(x + 1/x + y + 1/y)}$$

Excursions are a derived class

$$\mathcal{E} = \{w \in \{\uparrow, \downarrow, \leftarrow, \rightarrow\}^* \mid \chi(w) = (0,0,n)\}$$

Diagonals

The central diagonal maps series expansions to series expansions. e.g.

$$\Delta : K[[z_1, z_1^{-1}, \ldots, z_d, z_d^{-1}]][[t]] \to K[[t]].$$

defined as:

$$\Delta F(\mathbf{z}, t) = \Delta \sum_{k \geq 0} \sum_{\mathrm{n} \in \mathbb{Z}^d} f(\mathbf{n}, k)\, \mathbf{z}^{\mathrm{n}} t^k := \sum_{n \geq 0} f(n, n, \ldots, n)\, t^n. \tag{4}$$

$$\Delta(z_1^2 z_2 t + 3\mathbf{z}_1\mathbf{z}_2\mathbf{t} + 7z_1z_2t^2 + 5\mathbf{z}_1^2\mathbf{z}_2^2\mathbf{t}^2) = 3t + 5t^2$$

Defined for any series. Appear in many places in mathematics

We use diagonals to describe the generating functions of derived classes.

Example: Multinomials

Central diagonal

$$\Delta \frac{1}{1-x-y} = \Delta \sum_{n\geq 0} (x+y)^n = \Delta \sum_{\ell \geq 0} \sum_{k\geq 0} \binom{\ell + k}{k} x^k y^\ell = \sum_{n\geq 0} \binom{2n}{n,n} y^n.$$

Off center diagonals

$$\Delta^{(r,s)} \frac{1}{1-x-y} = \sum_{n\geq 0} \binom{rn+sn}{rn,sn} y^n.$$

This example generalizes naturally to arbitrary dimension, using multinomials:

$$\Delta^{\mathrm{r}} \frac{1}{1-(z_1+\cdots+z_d)} = \sum_{n\geq 0} \binom{n(r_1+\cdots+r_d)}{nr_1,\ldots,nr_d} z_d^n.$$

Balanced word classes

$$\mathcal{L} = \{\text{binary expansions of } n \mid n \equiv 0 \mod 3.\}$$

Size = length of string

$$\mathcal{L} = (0 + (1(01^*0)^*1))^*$$

Parameter $\chi(w) = (|w|_0, |w|_1) = (\#0s \text{ in } w, \#1s \text{ in } w)$

$$\mathcal{L}_= = \{w \in \mathcal{L} \mid \#0s = \#1s\}$$

$$= \{\epsilon, 1001, 0011, 0110, 1100, 010101, 101010, 11100001, 10011001, \dots\}$$

$$L(x,y) = \frac{1}{1 - \left(x + \frac{y^2}{1 - \frac{x^2}{1-y}}\right)} \qquad L_=(y) = \Delta L(x,y)$$

$$L_=(y) = \Delta\left(1 + x + .. + y^2(1 + 2x + 4x^2 + ..) + y^3(x^2 + 2x^3 + 5x^4 + ..) + \dots\right)$$

$$= 1 + 4y^2 + 2y^3 + 36y^4 + \dots$$

(size by half length)

Other subseries extraction as diagonal

$F(\mathbf{z}, t)$ with series $\in K[[z_1, z_1^{-1}, \dots, z_d, z_d^{-1}]][[t]]$:

$$\sum_{k\geq 0} \sum_{\mathrm{n}\in\mathbb{Z}^d} f(\mathbf{n}, k)\, \mathbf{z}^{\mathrm{n}} t^k$$

Constant Term

$$\begin{aligned} \mathrm{CT}\, F(\mathbf{z}, t) &= \sum_{n\geq 0} f(0, 0, \dots, 0, n)\, t^n \\ &= \Delta F\left(\frac{1}{z_1}, \dots, \frac{1}{z_d}, z_1 z_2 \dots z_d t\right) \end{aligned}$$

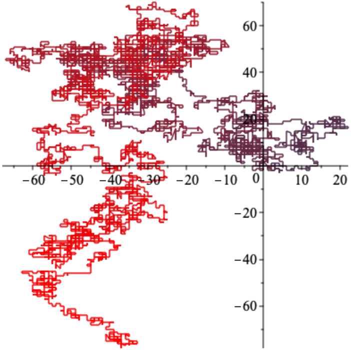
Positive Series

$$\begin{aligned} [z_1^{\geq 0} \dots z_k^{\geq 0}] F(\mathbf{x}, t) &= \sum_{\mathrm{n}\in\mathbb{N}^d} f(\mathbf{n}, k)\, \mathbf{z}^{\mathrm{n}} t^k \\ &= \Delta \left(\frac{F\left(\frac{1}{z_1}, \dots, \frac{1}{z_d}, z_1 z_2 \dots z_d t\right)}{(1 - z_1) \dots (1 - z_k)} \right) \end{aligned}$$

Excursions

Excursions: start and end at $(0,0)$ with steps from $S =$ ✥:

$$\mathcal{E} = \{w \in \{\uparrow, \downarrow, \leftarrow, \rightarrow\}^* \mid \chi(w) = (0,0)\}$$



OGF for excursions:

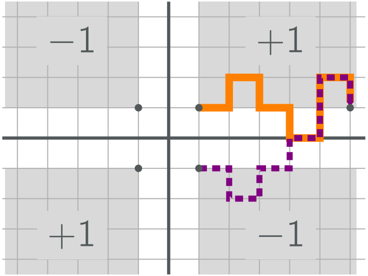
$$\sum walk_{\mathbb{Z}^2}^{\text{✥}}((0,0) \xrightarrow{n} (0,0))\, t^n = [x^0 y^0] \frac{1}{1 - t(x + 1/x + y + 1/y)}$$

$$= \Delta \frac{1}{1 - txy(1/x + x + 1/y + y))}$$

The set of combinatorial classes with OGF a diagonal of $\mathbb{N}$-rational is smaller than you'd like. (does not include Catalan!)
Differences of these classes are a wider class of series.

Walks confined to a quadrant - Reflection Principle

$$\sum_{n\geq 0} \text{walk}^{\scriptscriptstyle\text{✥}}_{\mathbb{N}^2}((0,0) \xrightarrow{n} (0,0))\, t^n$$



$$= [x^1y^1]\frac{xy - x/y + (xy)^{-1} + y/x}{(1 - t(x + 1/x + y + 1/y))}$$

$$= \mathrm{CT}\,\frac{\left(x - \frac{1}{x}\right)\left(y - \frac{1}{y}\right)}{xy(1 - t(x + 1/x + y + 1/y))}$$

$$= \Delta\frac{xy\left(x - \frac{1}{x}\right)\left(y - \frac{1}{y}\right)}{1 - txy(x + 1/x + y + 1/y)}$$

$$= \Delta\frac{(x^2 - 1)(y^2 - 1)}{1 - t(x^2y + y + xy^2 + x)}.$$

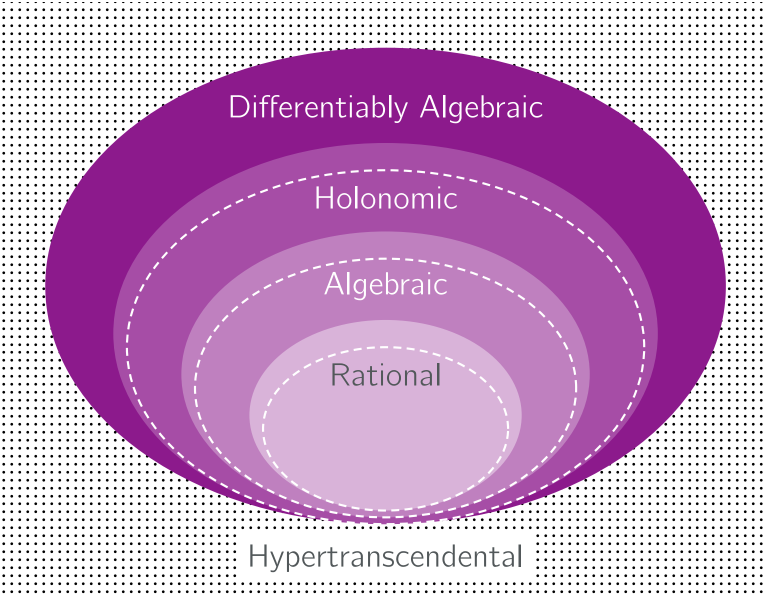
Diagonals and combinatorial generating functions

- Univariate algebraic functions are diagonals of bivariate rationals.
- Holonomic functions
  - Algebraic functions are Holonomic
  - Diagonals of Holonomic functions are Holonomic
  - OGFs of derived subclasses of algebraic and S-regular
  - Reflection principle – walks in Weyl chambers (from representation theory)

Lingering questions

Are combinatorial holonomic functions always diagonals? Are holomonic classes always (in bijection with) a derived class of an algebraic or regular? algebraic combination of derived classes?

Taxonomy of Generating Functions



Conclusion

Classic results of great utility to the combinatorialist

- Nature and type of singularities for series solutions of different equations types
- Behaviour near the singularities
- Asymptotic form of solutions for algebraic and linear ODE equations
- Results on series with positive coefficients (Pringsheim, Polya Carlson, Fatou,. . .)
  $F(z)$ converges inside the unit disc $\implies$ it is a rational function or transcendental over $\mathbb{Q}(z)$.

Transcendency

Transcendental OGF $\implies$ class has no algebraic specification.

Trancendancy criterion

$$[z^n]F(z) \sim C\mu^n n^s, \quad s \notin \mathbb{Q} \setminus \{-1, -2, \ldots\}$$

$$\mathcal{C} = \{u \in \{a, b, c\}^* \mid |u|_a \neq |u|_b \text{ or } |u|_a \neq |u|_c\}$$

$$\{a, b, c\}^* \setminus \mathcal{C} = \{u \in \{a, b, c\} \mid |u|_a = |u|_b = |u|_c\}$$

$$3^n - c_n = \binom{3n}{n, n, n}$$

$$\underbrace{\sum 3^n z^n}_{\text{rational}} - \sum c_n z^n = \underbrace{\sum \underbrace{\binom{3n}{n, n, n}}_{\sim C\, 27^n n^{-1}} z^n}_{\textit{transcendental}}$$

$$\implies \sum c_n z^n \text{ transcendental}$$

Conclusion

1. Dictionary between combinatorial specification and OGF functional equations
2. 3 families of combinatorial classes: S-regular, algebraic, derived subclasses
3. Use results on the nature of solutions to help sort objects and make effective computation
4. Diagonal operator is used to describe many combinatorial classes

$$\Delta F(\mathbf{z}, t) = \Delta \sum_{k \geq 0} \sum_{\mathrm{n} \in \mathbb{Z}^d} f(\mathbf{n}, k)\, \mathbf{z}^{\mathrm{n}} t^k := \sum_{n \geq 0} f(n, n, \ldots, n)\, t^n.$$

5. Christol's Conjecture: Every G-series is a diagonal of a rational function
6. Next: Given a multivariable rational function, determine the coefficient asymptotics of a diagonal.

Bibliography

- Analytic Combinatorics Philippe Flajolet, Robert Sedgewick
- Analytic Combinatorics in Several Variables (ACSV) Pemantle and Wilson
- Analytic Combinatorics: A Multivariable Approach, Marni Mishna
- Random Walk in a Weyl Chamber, Ira M. Gessel and Doron Zeilberger Proceedings of the American Mathematical Society Vol. 115, No. 1 (May, 1992), pp. 27-31
- R. Stanley. Enumerative Combinatorics, volume 2. Cambridge University Press, 1999.
- Analytic models and ambiguity of context-free languages Philippe Flajolet, Theoretical Computer Science Volume 49, Issues 23, 1987, Pages 283-309