Complexity of Combinatorial Sequences

Igor Pak, UCLA

ICM, Rio de Janeiro, August 7, 2018

What is a combinatorial sequence?

OEIS now has over 300,000 sequences!

> *Our policy has been to include all interesting sequences, no matter how obscure the reference.* [N.J.A. Sloane, S. Plouffe, EIS, 1995]

> [The EIS contains] *the unrelenting cascade of numbers,* [..] *lists Hard, Disallowed and Silly sequences.* [Richard Guy, 1997]

Selected integer sequences (from OEIS)

`A000001`: 1, 1, 1, 2, 1, 2, 1, 5, 2, 2, 1, 5, 1, 2, 1, 14, 1, 5, 1, 5, . . . $\leftarrow$ finite groups

`A000037`: 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, . . . $\leftarrow$ non-squares

`A000040`: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, . . . $\leftarrow$ primes

`A000041`: 1, 2, 3, 5, 7, 11, 15, 22, 30, 42, 56, 77, 101, 135, 176, . . . $\leftarrow p(n)$

`A000042`: 1, 11, 111, 1111, 11111, 111111, 1111111, 11111111, . . . $\leftarrow n$ in unary

`A000045`: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 232, 375, . . . $\leftarrow F_n$

`A000052`: 8, 5, 4, 9, 1, 7, 6, 3, 2, 0, 18, 80, 88, 85, 84, . . . $\leftarrow$ alphabetical ordering

`A000054`: 4, 14, 23, 34, 42, 50, 59, 72, 81, 86, 96, 103, 110, 116, . . . $\leftarrow$ NYC A line

`A000085`: 1, 2, 4, 10, 26, 76, 232, 764, 2620, 9496, . . . $\leftarrow$ # involutions in S_n

`A000088`: 1, 2, 4, 11, 34, 156, 1044, 12346, 274668, . . . $\leftarrow$ # graphs on n vertices

Combinatorial sequences

Let $\{a_n\}$ be a *combinatorial sequence*, e.g.

$$\begin{aligned}
a_n &= \#\text{ of triangulations of a convex } n\text{-gon} &= \tfrac{1}{n+1}\tbinom{2n}{n} \\
a_n &= \#\text{ of domino tilings of } [n \times n] &= \det M_n \\
a_n &= \#\text{ of connected labeled graphs on } n \text{ vertices} &\leftarrow \text{RR} \\
a_n &= \#\text{ of triangulations of a } n \times n \text{ grid} &= \gamma^{n^2(1+o(1))}
\end{aligned}$$

Note: Combinatorial sequences have $a_n \in \mathbb{N}$.

Main Questions

Question 1: Does $\mathcal{A}(t) = \sum_n a_n t^n$ have a *formula*?

Question 2: Can a_n be *computed efficiently*?

Conjecture [Wilf, 1982]:

Number of unlabeled graphs on n vertices is *hard* to compute.

Fibonacci Numbers:

$$(1) \qquad F_n = F_{n-1} + F_{n-2}$$

$$(2) \qquad F_n = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n-i}{i}$$

$$(3) \qquad F_n = \left(\phi^n + \phi^{-n}\right)/\sqrt{5} \qquad \text{where} \qquad \phi = \left(\sqrt{5}+1\right)/2$$

$$(4) \qquad \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix}$$

Note: (1) is a definition, (2) implies $\{F_n\}$ is $\mathbb{N}$-rational, (3) gives exact asymptotics, and (4) is good for fast computation.

More Examples:

(1) $$D_n = [[n!/e]], \quad \text{where } [[x]] \text{ denotes the nearest integer}$$

(2) $$C_n = [t^n]\frac{1-\sqrt{1-4t}}{2t}$$

(3) $$E_n = n! \cdot [t^n]y(t), \quad \text{where } 2y' = 1 + y^2, \; y(0) = 1$$

(4) $$T_n = (n-1)! \cdot [t^n]z(t), \quad \text{where } z = te^{te^{te^{te^{\cdots}}}}$$

(5) $$p(n) = [t^n]\prod_{i=1}^{\infty}\frac{1}{1-t^i}$$

(6) $$\pi(n) = \sum_{k=2}^{n}\left(\left\lfloor\frac{(k-1)!+1}{k}\right\rfloor - \left\lfloor\frac{(k-1)!}{k}\right\rfloor\right).$$

Complexity Approach:

Definition: Sequence $\{a_n\}$ can be *computed efficiently* if there is an algorithm which computes a_n in time poly(n).

Examples: Fibonacci numbers F_n, Catalan numbers C_n, derangement numbers D_n, partition numbers $p(n)$, etc.

Theorem: The following sequences can be computed efficiently:

connected 3-regular graphs on n vertices [Goulden–Jackson, Gessel]

non-isomorphic trees on n vertices [Goldberg]

partitions of $2^n - 1$ into $\{1, 2, 4, 8, \ldots\}$ (Cayley composition numbers) [P.–Yeliussizov]

Negative Results?

> *A theory is falsifiable if there exists at least one non-empty class of [..] basic statements which are forbidden by it.* [Karl Popper, 1934]

Note: Wilf's Conjecture is a potential example (or not?)
Some sequences take too long to write, e.g. $a_n = 2^{2^n}$.
Some sequences are essentially open problems in mathematics, e.g. $\{\text{prime } F_n\}$.

Conjecture: The number of self-avoiding walks in $\mathbb{Z}^2$ with n steps starting O cannot be computes in time poly(n).

What Gives?

> P *versus* NP — *a gift to mathematics from computer science.* [Stephen Smale]

Note: Sometimes a gift is a **Trojan Horse**.

Classes of combinatorial sequences

(1) ***rational*** g.f. $\mathcal{A}(t) = P(t)/Q(t)$, $P, Q \in \mathbb{Z}[t]$. E.g. $a_n := F_n$, $\mathcal{A}(t) = 1/(1-t-t^2)$.

(2) ***algebraic*** g.f. $c_0\mathcal{A}^k + c_1\mathcal{A}^{k-1} + \ldots + c_k = 0$, $c_i \in \mathbb{Z}[t]$. E.g. $a_n := C_n$, $\mathcal{A}(t) = (1-\sqrt{1-4t})/2t$.

(3) ***D-finite*** g.f. $c_0\mathcal{A} + c_1\mathcal{A}' + \ldots + c_k\mathcal{A}^{(k)} = 0$, $c_i \in \mathbb{Z}[t]$. E.g. $a_n :=$ # involutions in S_n, then $a_n = a_{n-1} + (n-1)a_{n-2}$. The sequences $\{a_n\}$ are called ***P-recursive***

(4) ***ADE*** (also ***D-algebraic***) g.f. $Q\big(t, \mathcal{A}, \mathcal{A}', \ldots, \mathcal{A}^{(k)}\big) = 0$, $Q \in \mathbb{Z}[t, x_0, x_1, \ldots, x_k]$
E.g. $a_n = \#\{\sigma(1) < \sigma(2) > \sigma(3) < \ldots \in S_n\}$, $\mathcal{A}'' = \mathcal{A} \cdot \mathcal{A}'$.
Also $p(n) =$ # integer partitions of n (Jacobi, Ramanujan). Then $F(t) = \sum_n p(n)t^n$ satisfies:

$$4F^3F'' + 5tF^3F''' + t^2F^3F^{(4)} - 16F^2(F')^2 - 15tF^2F'F'' - 39t^2F^2(F'')^2$$
$$+ 20t^2F^2F'F''' + 10tF(F')^3 + 12t^2F(F')^2F'' + 6t^2(F')^4 = 0.$$

General philosophy:

Inclusions: (1) $\subset$ (2) $\subset$ (3) $\subset$ (4).

Note: P-recursive sequences $\{a_n\}$ can be computed efficiently by definition.

Proposition: ADE sequences $\{a_n\}$ can be computed efficiently.

- Most combinatorial sequences have *nice* g.f. (D-finite, ADE, etc.)
- Proving that $\mathcal{A}(t) = \sum_n a_n t^n$ is *not* D-finite or ADE is difficult.
- Thus, proving $\{a_n\}$ non-D-finite and non-ADE are important first steps.

How hard can that be? Non-combinatorial examples:

Theorem (Jacobi, 1848): $\sum_n t^{n^2}$ is ADE.

Theorem (Lipshitz, Rubel, 1986): $\sum_n t^{2^n}$ is *not* ADE.

Conjecture: $\sum_n t^{n^3}$ is *not* ADE.

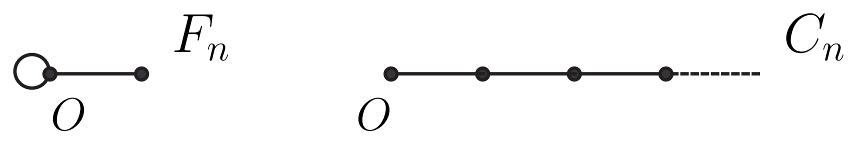
Theorem (Flajolet, Gerhold and Salvy, 2005): $\sum_n p_n t^n$ is *not* D-finite, where p_n is n-th prime.

Conjecture: $\sum_n p_n t^n$ is *not* ADE.

Walks on graphs:

Definition: Let $\Gamma = (V, E)$ be a graph, $O \in V$ fixed.
Let a_n be the number of $x \to y$ walks in Γ of length n.

Examples:



Further examples: $\Gamma \subset \mathbb{Z}^d$, i.e. a region in the grid
$\Gamma = \text{Cayley}(G, S)$, where G infinite group, $G = \langle S \rangle$, $S = S^{-1}$ finite (*cogrowth sequence*).

Theorem (folklore): $G = \mathbb{Z}^d$, any finite S, then $\{a_n\}$ is P-recursive.

Proposition (Furstenberg, 1967): $\Gamma = \mathbb{Z}^2$, then $\left\{a_n = \binom{2n}{n}^2\right\}$ is not algebraic.

Theorem (Haiman, 1993): $\Gamma = \mathbb{F}_k$, S standard, then $\{a_n\}$ is algebraic.

Walks on Cayley graphs:

Theorem (Elder, Rechnitzer, Janse, van Rensburg, Wong, 2014)
Cogrowth sequence $\{a_n\}$ is P-recursive for $G = \mathrm{BS}(N, N)$, $S = \{x, x^{-1}, y, y^{-1}\}$, where $\mathrm{BS}(k, \ell) = \langle x, y \,|\, x^k y = y x^\ell \rangle$.

Theorem (Garrabrant, P., 2017) Cogrowth sequence $\{a_n\}$ is *not* P-recursive for
(1) virtually solvable groups of exponential growth with finite Prüfer rank;
(2) amenable linear groups of superpolynomial growth;
(3) groups of weakly exponential growth
(4) *Baumslag–Solitar groups* $\mathrm{BS}(k, 1)$, where $k \geq 2$;
(5) *lamplighter groups* $L(d, H) = H \wr \mathbb{Z}^d$, where H is finite abelian, $d \geq 1$.

Theorem (Bell, Mishna, 2018+) Cogrowth sequence $\{a_n\}$ is *not* P-recursive for amenable groups of superpolynomial growth.

Walks on Cayley graphs:

Main lemma: (Birkhoff, Trjitzinsky, Katz, …)
Let $\{a_n\}$ be a P-recursive, $a_n < C^n$ for some $C > 0$ and all $n \geq 1$. Then

$$a_n \sim \sum_{i=1}^{m} K_i \lambda_i^n n^{\alpha_i} (\log n)^{\beta_i},$$

where $K_i \in \mathbb{R}_+$, $\lambda_i \in \overline{\mathbb{Q}}$, $\alpha_i \in \mathbb{Q}$, and $\beta_i \in \mathbb{N}$.

Theorem: (Garrabrant, P., 2017) There is a $\langle S \rangle = \mathbb{F}_k$, s.t.
the cogrowth sequence $\{a_n\}$ is *not* P-recursive.

Theorem: (Garrabrant, P., 2018+) There is a $\langle S \rangle = \mathbb{F}_k$, s.t.
the cogrowth sequence $\{a_n\}$ is *not* ADE.

Permutation classes

Permutation $\sigma \in S_n$ contains $\pi \in S_k$ if M_π is a submatrix of M_σ.
Otherwise, σ *avoids* π. Such π are called *patterns*.
For example, (4564123) contains (321) but avoids (4321).

$$\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \qquad \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

Fix a set of patterns $\mathcal{F} \subset S_k$. Denote by $C_n(\mathcal{F})$ the number of $\sigma \in S_n$ which avoids all $\pi \in \mathcal{F}$.

Question 1: Is $\mathcal{A}(t) = \sum_n C_n(\mathcal{F}) t^n$ always D-finite or ADE?

Question 2: Can $C_n(\mathcal{F})$ always be computed in poly(n) time?

Notable results and examples:

(0) $C_n(12\cdots k, \ell\cdots 21) = 0$, $\forall n > (k-1)(\ell-1)$ [Erdős, Szekeres, 1935]

(1) $C_n(123) = C_n(213) = \mathrm{Cat}(n)$ [MacMahon, 1915], [Knuth, 1973]

(2) $C_n(123, 132, 213) = \mathrm{Fib}(n+1)$ [Simion, Shmidt, 1985]

(3) $C_n(2413, 3142) = \mathrm{Shröder}(n)$ [Shapiro, Stephens, 1991]

(4) $C_n(1234) = C_n(2143)$ has D-finite g.f. [Gessel, 1990]

(5) $C_n(1342) = C_n(2416385)$ has algebraic g.f. [Bona, 1997]

(6) $C_n(\mathcal{F}) < K(\mathcal{F})^n$ [Marcus, Tardos, 2004], improving [Alon, Friedgut, 2000]

(7) $K(\pi) = e^{k^{\Omega(1)}}$ w.h.p., for $\pi \in S_k$ random [Fox, 2013]

(8) σ contains π is NP-complete [Bose, Buss, Lubiw, 1998]

(9) can be decided in $O(n \log n)$ for π fixed [Guillemot, Marx, 2014]

Our main results

Noonan–Zeilberger Conjecture (1996):
The g.f. for $\{C_n(\mathcal{F})\}$ is D-finite, for all fixed $\mathcal{F} \subset S_k$.

Theorem 1. [Garrabrant, P., 2015]
The NZ Conjecture is false. To be precise, there is a set $\mathcal{F} \subset S_{80}$, $|\mathcal{F}| < 31000$, such that $\sum_n C_n(\mathcal{F}) t^n$ is not *D-finite.*

Theorem 2. [Garrabrant, P., 2018+]
There is a set $\mathcal{F} \subset S_{80}$, such that $\sum_n C_n(\mathcal{F}) t^n$ is not *ADE.*

Historical notes: NZ Conjecture was first stated by Gessel in 1990. In 2005, Zeilberger changes his mind, conjectures that $\{C_n(1324)\}$ is a counterexample.
In 2014, Zeilberger changes his mind half-way back, promises \$100 bounty, pays up in 2015.

Computability implications

Theorem 3. [Garrabrant, P., 2015]
The problem whether $C_n(\mathcal{F}) = C_n(\mathcal{F}') \mod 2\ \forall n$, *is undecidable.*

Corollary 1. For all k large enough, there exists $\mathcal{F}, \mathcal{F}' \subset S_k$, s.t. the first time $C_n(\mathcal{F}) \neq C_n(\mathcal{F}') \mod 2$ is for

$$n > 2^{2^{2^{2^{2^k}}}}.$$

Corollary 2. There exist two finite sets of patterns $\mathcal{F}$ and $\mathcal{F}'$ in S_k, s.t. the problem of whether $C_n(\mathcal{F}) = C_n(\mathcal{F}')$ mod 2, for all $n \in \mathbb{N}$, is independent of ZFC.

Complexity result and Wilf's question

Theorem 4. [Garrabrant, P., 2015]
If $\mathsf{EXP} \neq \oplus\mathsf{EXP}$*, then there exists a finite set of patterns* $\mathcal{F}$*, such that the sequence* $\{C_n(\mathcal{F})\}$ *cannot be computed in time polynomial in* n*.*

Reminder: EXP = exponential time,
$\oplus\mathsf{P}$ = parity version of the class of counting problem $\#\mathsf{P}$,
$\oplus\mathsf{EXP}$ = parity version of the class of counting problem $\#\mathsf{EXP}$.
$\mathsf{EXP} \neq \oplus\mathsf{EXP}$ assumption is similar to $\mathsf{P} \neq \oplus\mathsf{P}$.

Remark: This answers Wilf's question (1982)
"Can one describe a reasonable and natural family of combinatorial enumeration problems for which there is provably no polynomial-in-n time formula or algorithm to compute $f(n)$*?"*

Simulating Turing Machines

Let $\mathbb{X}$ denote the set of sequences $\{\xi_\Gamma(n)\}$, where
Γ is a two-stack automaton with source S and sink T, and
$\xi_\Gamma(n)$ is the number of balanced $S-T$ paths of length n.
(Here *balanced* means that both stacks are empty at the end).

Main Lemma

Let $\xi : \mathbb{N} \to \mathbb{N}$ *be a function in* $\mathbb{X}$*. Then there exist* $k, a, b \in \mathbb{N}$ *and sets of patterns* $\mathcal{F}, \mathcal{F}' \in S_k$*, such that*
$\xi(n) = C_{an+b}(\mathcal{F}) - C_{an+b}(\mathcal{F}') \bmod 2$*, for all* $n \geq 1$*.*

Main Lemma can be used to derive both Theorem 3 and Theorem 4.

Note: Here mod 2 can be changed to any mod p, but cannot be completely removed.

Proof of Theorem 1.

Lemma 1. Let $\{a_n\}$ be a P-recursive sequence (i.e. with D-finite g.f.)
Let $\overline{\alpha} = (\alpha_1, \alpha_2, \ldots)$, $\overline{\alpha} \in \{0,1\}^\infty$ defined by $\alpha_n = a_n \bmod 2$.
Then there is a finite binary word w which is NOT a subword of $\overline{\alpha}$.

Lemma 2. There is a two-stack automaton Γ s.t. the number of balanced paths $\xi_\Gamma(n)$ is given by the sequence
$0,\ 1,\ 0, 0,\ 0, 1,\ 1, 0,\ 1, 1,\ 0, 0, 0,\ 0, 0, 1,\ 0, 1, 0, \ldots$

Lemma 1, Lemma 2 and the Main Lemma imply Theorem 1.

Proof of Theorem 2.

Lemma 1′. Let $\{a_n\}$ be a sequence, and let $\{n_i\}$ be the sequence of indices with odd a_n. Suppose
1) for all $b, c \in \mathbb{N}$, there exists i such that $n_i = b \bmod 2c$,
2) $n_i/n_{i+1} \to 0$ as $i \to \infty$.
Then the g.f. for $\{a_n\}$ is not ADE.

Observe: $\{a_n = n! + n\}$ satisfies conditions of Lemma 1′.

Lemma 2′. There is a two-stack automaton Γ s.t. the number of balanced paths $\xi_\Gamma(n) = n! + n$.

Lemma 1′, Lemma 2′ and the Main Lemma imply Theorem 2.

Notes on the proofs:

(i) We use exactly 6854 partial patterns.

(i) Automaton Γ in Lemma 2 uses 31 vertices, which is why the alphabet has size 10×10 only.

(iii) The largest matrix in $\mathcal{F}$ has 8×8 blocks, which is why Theorem 1 has permutations in S_{80}.

(iv) Proof of Lemma 1 has only 2 paragraphs, but it took over a year to find a statement. Lemma 1′ took another year.

(v) Condition n_i/n_{i+1} in Lemma 1′ cannot be weakened, e.g. Cat(n) is odd if and only if $n = 2^m - 1$.

Open problems:

Conjecture 1. The *Wilf-equivalence problem* of whether $C_n(\mathcal{F}_1) = C_n(\mathcal{F}_2)$ for all $n \in \mathbb{N}$, is undecidable.

Conjecture 2. The *Wilf-equivalence problem* for single permutations: $C_n(\sigma) = C_n(\omega)$ for all $n \in \mathbb{N}$, is decidable.

Conjecture 3. Sequence $\{C_n(1324)\}$ is not P-recursive.

Conjecture 4. There exists a fixed set of patterns $\mathcal{F}$, s.t. computing $\{C_n(\mathcal{F})\}$is $\#\mathsf{EXP}$-complete.

Thank you!