The number of Catalan words avoiding a finite set of patterns has a rational generating function

Justin Troyka

California State University, Los Angeles

Joint work in progress with [Jay Pantone and] Vincent Vatter

Permutation Patterns
June 7, 2025

I. Patterns in Catalan words
II. Infinite antichain of Catalan words
III. Rationality of finitely based classes

I. **Patterns in Catalan words**
II. Infinite antichain of Catalan words
III. Rationality of finitely based classes

I. Patterns in Catalan words

A Catalan word or Catalan sequence of size n is a word $w = w_1 w_2 \dots w_n$ with entries from $\mathbb{N}$ such that

i. $w_1 = 0$

ii. and $w_i \leqslant w_{i-1} + 1$ for $i \geqslant 2$.

I. Patterns in Catalan words

A Catalan word or Catalan sequence of size n is a word $w = w_1 w_2 \ldots w_n$ with entries from $\mathbb{N}$ such that

i. $w_1 = 0$
ii. and $w_i \leqslant w_{i-1} + 1$ for $i \geqslant 2$.

- Yes: 001120123342331
- No: 00112341331

I. Patterns in Catalan words

A Catalan word or Catalan sequence of size n is a word $w = w_1 w_2 \dots w_n$ with entries from $\mathbb{N}$ such that

i. $w_1 = 0$

ii. and $w_i \leqslant w_{i-1} + 1$ for $i \geqslant 2$.

- Yes: 001120123342331
- No: 0011234$1$$3$31

I. Patterns in Catalan words

A Catalan word or Catalan sequence of size n is a word $w = w_1 w_2 \dots w_n$ with entries from $\mathbb{N}$ such that

i. $w_1 = 0$
ii. and $w_i \leqslant w_{i-1} + 1$ for $i \geqslant 2$.

- Yes: 001120123342331
- No: 0011234133 1

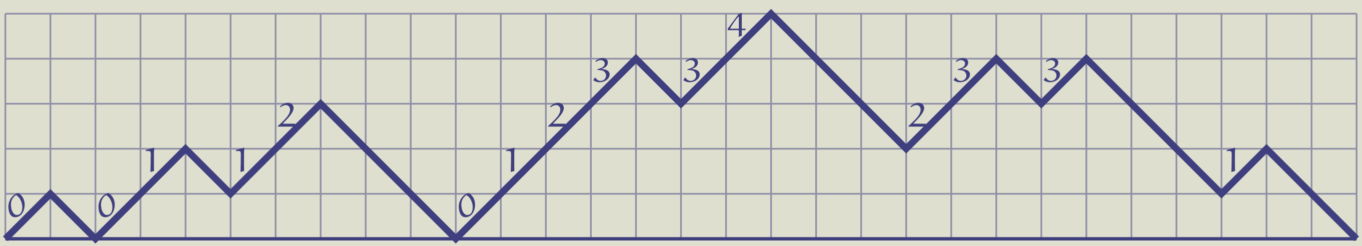
Catalan words are a subset of restricted growth functions, which are a subset of both ascent sequences and inversion sequences.

I. Patterns in Catalan words

- $\mathcal{C}$ = set of Catalan words
- $\mathcal{C}_n$ = set of Catalan words of size n

Fact: $|\mathcal{C}_n| = c_n$ (the nth Catalan number).
Bijection with Dyck paths: label the up steps by height

$001120123342331 \in \mathcal{C}_{15}$



I. Patterns in Catalan words

Given Catalan words $v, w \in \boldsymbol{C}$, we say $v \leqslant w$, or v is contained in w (as a pattern), if there exist $i_1 < i_2 < \cdots < i_k$ such that $\mathrm{st}(w_{i_1} w_{i_2} \ldots w_{i_k}) = v$.

- "st" denotes the standardization of a word.

$v \leqslant w$ means w has a subsequence that is order isomorphic to v.

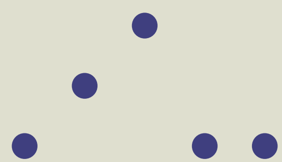
I. Patterns in Catalan words

Given Catalan words $v, w \in \mathcal{C}$, we say $v \leqslant w$, or v is contained in w (as a pattern), if there exist $i_1 < i_2 < \cdots < i_k$ such that $\mathrm{st}(w_{i_1} w_{i_2} \ldots w_{i_k}) = v$.

- "st" denotes the standardization of a word.

$v \leqslant w$ means w has a subsequence that is order isomorphic to v.

- 00112012334233**1** contains 01120 since $\mathrm{st}(12241) = 01120$.
- 00112012334233**1** avoids 01200, since no subsequence goes



I. Patterns in Catalan words

If R is a set of words, then

- $\mathcal{C}(R) =$ set of Catalan words avoiding every element of R.
- $\mathcal{C}_n(R) =$ set of Catalan words of size n avoiding every element of R.

I. Patterns in Catalan words

If R is a set of words, then

- $\mathcal{C}(R) =$ set of Catalan words avoiding every element of R.
- $\mathcal{C}_n(R) =$ set of Catalan words of size n avoiding every element of R.

Examples:

- $\mathcal{C}(012)$ is the set of Catalan words with only $0'$s and $1'$s.
  - $|\mathcal{C}_n(012)| = 2^{n-1}$.

I. Patterns in Catalan words

If R is a set of words, then

- $\mathcal{C}(R) =$ set of Catalan words avoiding every element of R.
- $\mathcal{C}_n(R) =$ set of Catalan words of size n avoiding every element of R.

Examples:

- $\mathcal{C}(012)$ is the set of Catalan words with only $0'$s and $1'$s.
  - $|\mathcal{C}_n(012)| = 2^{n-1}$.
- $\mathcal{C}(10)$ is the set of Catalan words that are weakly increasing.
  - $\mathcal{C}(10) = \mathcal{C}(010)$.
  - $|\mathcal{C}_n(010)| = 2^{n-1}$.

I. Patterns in Catalan words

If R is a set of words, then

- $\mathcal{C}(R) =$ set of Catalan words avoiding every element of R.
- $\mathcal{C}_n(R) =$ set of Catalan words of size n avoiding every element of R.

Examples:

- $\mathcal{C}(012)$ is the set of Catalan words with only $0'$s and $1'$s.
  - $|\mathcal{C}_n(012)| = 2^{n-1}$.
- $\mathcal{C}(10)$ is the set of Catalan words that are weakly increasing.
  - $\mathcal{C}(10) = \mathcal{C}(010)$.
  - $|\mathcal{C}_n(010)| = 2^{n-1}$.
- $\mathcal{C}(000)$ is the set of Catalan words with $\leqslant 2$ occurrences of each value.
  - $\{|\mathcal{C}_n(000)|\}_{n\geqslant 0} = \{1, 1, 2, 4, 9, 19, 42, 90, \dots\}$.

I. Patterns in Catalan words

Baril, Kirgizov, & Vajnovszki (2018) enumerate $\mathcal{C}_n(v)$ for every size-3 word v (refined by descent number).

Catalan words avoiding a pattern of length three.

Pattern p	Sequence $\lvert\mathcal{C}_n(p)\rvert$	Generating function	OEIS [12]
012, 001, 010	2^{n-1}	$\frac{1-x}{1-2x}$	A011782
021	$(n-1)\cdot 2^{n-2}+1$	$\frac{1-4x+5x^2-x^3}{(1-x)(1-2x)^2}$	A005183
102, 201	$\frac{3^{n-1}+1}{2}$	$\frac{1-3x+x^2}{(1-x)(1-3x)}$	A007051
120, 101	F_{2n-1}	$\frac{1-2x}{1-3x+x^2}$	A001519
011	$\frac{n(n-1)}{2}+1$	$\frac{1-2x+2x^2}{(1-x)^3}$	A000124
000		$\frac{1-2x^2}{1-x-3x^2+x^3}$	
100	$\lceil\frac{(1+\sqrt{3})^{n+1}}{12}\rceil$	$\frac{1-2x-x^2+x^3}{1-3x+2x^3}$	A057960
110	$\frac{1}{2}\sum_{k=0}^{\lfloor\frac{n}{2}\rfloor}\binom{n+1}{2k+1}2^k-\frac{n-1}{2}$	$\frac{1-3x+2x^2+x^3}{(1-x)^2(1-2x-x^2)}$	
210		$\frac{1-5x+7x^2-x^3-x^4}{(1-2x)(1-4x+3x^2+x^3)}$	

I. Patterns in Catalan words

Baril, Khalil, & Vajnovszki (2021) enumerate $\mathcal{C}_n(u, v)$ for every pair of size-3 words u and v.

(they don't have a nice table)

I. Patterns in Catalan words

Baril, Mansour, Ramírez, & Shattuck (preprint 2024) enumerate $C_n(v)$ for every size-4 word v (refined by descent number).

Class	p	$C_p(x)$	Reference/OEIS #
1	1234	$\frac{1-2x}{1-3x+x^2}$	Cor. 2.2/A001519
2	1243	$\frac{(1-x)(1-5x+7x^2-x^3)}{(1-2x)^2(1-3x+x^2)}$	Cor. 2.5/A244885
3	1324 1423	$\frac{1-7x+16x^2-12x^3+x^4}{(1-2x)(1-3x)(1-3x+x^2)}$	Cor. 2.8, Thm. 2.10
4	1342	$\frac{1-6x+12x^2-9x^3+3x^4}{(1-x)(1-3x+x^2)^2}$	Cor. 2.12/A116845
5	1432	$\frac{1-11x+49x^2-112x^3+136x^4-78x^5+9x^6+6x^7-x^8}{(1-x)(1-2x)^2(1-3x+x^2)(1-4x+3x^2+x^3)}$	Cor. 2.15
6	2134 3412	$\frac{(1-x)(1-3x)}{1-5x+6x^2-x^3}$	Thm. 3.1/A080937
7	2143 4312	$\frac{(1-3x+x^2)(1-7x+17x^2-17x^3+6x^4-x^5)}{(1-x)(1-2x)^2(1-6x+10x^2-4x^3+x^4)}$	Thms. 4.2 and 3.7
8	2314 2413 3124 4123	$\frac{1-5x+6x^2-x^3}{(1-2x)(1-4x+2x^2)}$	Thm. 3.2
9	2341	$\frac{1-3x+x^2}{(1-x)(1-3x)}$	Thm. 3.3/A024175
10	2431	$\frac{1-8x+23x^2-27x^3+8x^4+5x^5-x^6}{(1-2x)^2(1-5x+6x^2-x^4)}$	Thm. 3.5
11	3142	$\frac{1-13x+69x^2-192x^3+297x^4-244x^5+82x^6+9x^7-11x^8+x^9}{(1-x)(1-3x+x^2)(1-10x+37x^2-61x^3+39x^4+x^5-5x^6)}$	Thm. 4.4
12	3214 4132 4213	$\frac{(1-2x)(1-4x+2x^2)}{(1-x)(1-6x+9x^2-x^3)}$	Thm. 3.8/A080938
13	3241 4231	$\frac{1-8x+21x^2-18x^3+x^5}{(1-3x)(1-6x+10x^2-3x^3-x^4)}$	Thm. 3.9
14	3421	$\frac{(1-x)(1-6x+10x^2-2x^3-2x^4)}{(1-3x+x^2)(1-5x+6x^2-x^4)}$	Thm. 3.10
15	4321	Too lengthy to state here	Thm. 4.6

I. Patterns in Catalan words

BMRS conjecture that $\mathcal{C}(w)$ has a rational generating function for every word w.

Which of these properties does $\mathcal{C}$ have:

1. Every proper subclass is rational:
2. Every proper finitely based subclass is rational:
3. Every proper well-quasi-ordered subclass is rational:

I. Patterns in Catalan words

BMRS conjecture that $\mathcal{C}(w)$ has a rational generating function for every word w.

Which of these properties does $\mathcal{C}$ have:

1. Every proper subclass is rational: No!
2. Every proper finitely based subclass is rational: Yes!
3. Every proper well-quasi-ordered subclass is rational: *(open)*

I. Patterns in Catalan words

BMRS conjecture that $\mathcal{C}(w)$ has a rational generating function for every word w.

Which of these properties does $\mathcal{C}$ have:

1. Every proper subclass is rational: No!
2. Every proper finitely based subclass is rational: Yes!
3. Every proper well-quasi-ordered subclass is rational: *(open)*

- $\mathcal{S}(321)$ has properties 2 and 3 (Albert, Brignall, Ruškuc, & Vatter 2019).
- $\mathcal{S}(312)$ has all three properties (Albert & Atkinson 2005).

I. Patterns in Catalan words

II. **Infinite antichain of Catalan words**

III. Rationality of finitely based classes

II. Infinite antichain of Catalan words

$$\begin{aligned} \text{Define } A = \{ & 012\,012\,123\,123, \\ & 012\,012\,123\,234\,234, \\ & 012\,012\,123\,234\,345\,345, \\ & 012\,012\,123\,234\,345\,456\,456, \\ & \ldots \}. \end{aligned}$$

A is an infinite set of Catalan words such that no word contains another word as a pattern — an infinite antichain.

II. Infinite antichain of Catalan words

$$\begin{aligned}\text{Define } A = \{&012\,012\,123\,123,\\ &012\,012\,123\,234\,234,\\ &012\,012\,123\,234\,345\,345,\\ &012\,012\,123\,234\,345\,456\,456,\\ &\ldots\}.\end{aligned}$$

A is an infinite set of Catalan words such that no word contains another word as a pattern — an infinite antichain.

Theorem (T. & Vatter 2025+): $\mathcal{C}$ is not a well-quasi-order (wqo): there exists an infinite antichain in $\mathcal{C}$.

Justin Troyka <jmtr... Jul 1, 2025, 3:07 AM (6 days ago) Reply
to Vince

Sounds like a great plan. I'll see you next week! Maybe an infinite antichain of Catalan words will come to me in a dream while I'm flying over the Atlantic.

Justin Troyka

...

Vince Vatter Jul 1, 2025, 3:16 AM (6 days ago) Reply
to me

A thought that just occurred to me... it might be enlightening to take an Av(321) antichain, convert to panel encoding, and convert back to Catalan word... although I guess it might *not* convert back. I'm flying over tomorrow, so maybe I'll have an antichain thought before you.

...

Vince Vatter Jul 2, 2025, 1:48 PM (5 days ago) Reply
to me

Hi Justin,

On the plane, I think I came up with an antichain, or at least, something that makes me suspect we can build an infinite antichain of Catalan words.

II. Infinite antichain of Catalan words

Why is an infinite antichain significant?

- It means that not all pattern-avoiding classes can be defined by a finite set of patterns, since $\mathcal{C}(A)$ has an infinite basis.
- It allows us to define an uncountable set of pattern-avoiding classes, all with different enumerations: $\{\mathcal{C}(\mathrm{R})\colon \mathrm{R} \subseteq A\}$.
- In a sense, $\mathcal{C}$ has greater complexity than quasi-orders that do not have infinite antichains. (Contrast with the Robertson–Seymour Theorem about graph minors; contrast with Albert & Atkinson 2005 on $\mathcal{S}(312)$.)

II. Infinite antichain of Catalan words

Why is an infinite antichain significant?

- It means that not all pattern-avoiding classes can be defined by a finite set of patterns, since $\mathcal{C}(A)$ has an infinite basis.
- It allows us to define an uncountable set of pattern-avoiding classes, all with different enumerations: $\{\mathcal{C}(\mathrm{R})\colon \mathrm{R} \subseteq A\}$.
- In a sense, $\mathcal{C}$ has greater complexity than quasi-orders that do not have infinite antichains. (Contrast with the Robertson–Seymour Theorem about graph minors; contrast with Albert & Atkinson 2005 on $\mathcal{S}(312)$.)

Corollary (T. & Vatter 2025+): There exist uncountably many sets $\mathrm{R} \subseteq \mathcal{C}$ such that the generating function for $\mathcal{C}(\mathrm{R})$ is not rational (or D-algebraic) (or even computable!).

I. Patterns in Catalan words
II. Infinite antichain of Catalan words
III. **Rationality of finitely based classes**

III. Rationality of finitely based classes

Theorem (T. & Vatter 2025+): If R is a non-empty finite subset of $\mathcal{C}$, then $\mathcal{C}(R)$ has a rational generating function.

Our proof is similar to the proof of the same property of $\mathcal{S}(321)$ by Albert, Brignall, Ruškuc, & Vatter 2019. We rely on the fact that every regular language has a rational generating function.

III. Rationality of finitely based classes

Theorem (T. & Vatter 2025+): If R is a non-empty finite subset of $\mathcal{C}$, then $\mathcal{C}(R)$ has a rational generating function.

Our proof is similar to the proof of the same property of $\mathcal{S}(321)$ by Albert, Brignall, Ruškuc, & Vatter 2019. We rely on the fact that every regular language has a rational generating function.

1. There exist a, b such that $\mathcal{C}(R) \subseteq \mathcal{C}((012\ldots a)^b)$.
2. The panel encoding of $\mathcal{C}((012\ldots a)^b)$ maps each Catalan word to a word over a finite alphabet, and the image of this encoding is a regular language.
3. Within this regular language, the set of words corresponding to $\mathcal{C}(R)$ is a regular language.

III. Rationality of finitely based classes

1. There exist a, b such that $\mathcal{C}(R) \subseteq \mathcal{C}((012\ldots a)^{b})$.

Example: $R = \{01120\}$.
01120 is contained in $0123\,0123\,0123$ (using $a = 3$ and $b = 3$).
Then $\mathcal{C}(01120) \subseteq \mathcal{C}(0123\,0123\,0123)$.

III. Rationality of finitely based classes

2. The panel encoding of $\mathcal{C}((012\ldots a)^b)$ maps each Catalan word to a word over a finite alphabet, and the image of this encoding is a regular language.

This is almost exactly the same as in ABRV 2019.

Example: We will encode $00112012334 2331 \in \mathcal{C}(0123\,0123\,0123)$ as a word over the alphabet $\{0, \overleftarrow{0}, \overrightarrow{0}, \overleftrightarrow{0}, 1, 2, \#\}$.

$$\begin{array}{l|l|l} 00112\,\overrightarrow{0}\,\underline{\overleftarrow{1}\,23342331} & \overleftrightarrow{0}\,\underline{\overleftarrow{1}\,223122}0 & \overleftarrow{0}\,112011 \\ \hline 00112\,\overrightarrow{0} & \overleftrightarrow{0}\,0 & \overleftarrow{0}\,112011 \end{array}$$

The panel encoding is read from the second row:

$$00112\,\overrightarrow{0}\,\#\,\overleftrightarrow{0}\,0\,\#\,\overleftarrow{0}\,112011$$

III. Rationality of finitely based classes

2. The panel encoding of $\mathcal{C}((012\ldots a)^b)$ maps each Catalan word to a word over a finite alphabet, and the image of this encoding is a regular language.

- Avoiding $(012\ldots a)^b$ means there are $< b$ underlined subwords in each step, and so each block of the panel encoding has $< b$ left arrows and $< b$ right arrows. This is what makes the language regular!
- The language of panel encodings is the same as in ABRV, but with one extra condition: each block of the panel encoding starts with $\overleftarrow{0}$ or $\overleftrightarrow{0}$ except the first block.

III. Rationality of finitely based classes

3. Within this regular language [of panel encodings], the set of words corresponding to $\mathcal{C}(R)$ is a regular language.

As in ABRV, we can use a transducer on panel encodings that non-deterministically marks an occurrence of some pattern and checks whether it is σ.

Example: $\sigma = 01120$

$$
\begin{aligned}
0\,0\,1\,1\,2\,\overrightarrow{0}\,\#\,\overleftrightarrow{0}\,0\,\#\,\overleftarrow{0}\,1\,1\,2\,0\,1\,1 &\to 0\,0\,1\,{\color{red}1\,2}\,\overrightarrow{0}\,\#\,\overleftrightarrow{0}\,{\color{red}0}\,\#\,\overleftarrow{{\color{red}0}}\,1\,1\,{\color{red}2}\,0\,1\,1 \\
&\to {\color{red}1\,2}\,\overrightarrow{}\,\#\,\overleftrightarrow{}\,{\color{red}0}\,\#\,\overleftarrow{{\color{red}0}}\,{\color{red}2} \\
&\to {\color{red}0\,1\,1\,2\,0}
\end{aligned}
$$

Further questions

- Is there some kind of symmetry on Catalan words to explain why so many classes have the same enumeration? (“Recursive reversal”)
- Can we bound the degree of the rational generating function in order to automatically enumerate classes?
- Can our proof of rationality be adapted to prove the rationality of the bivariate generating function tracking descent number?
- Is every wqo class of Catalan words rational?
- Conjecture: for $w \in \mathcal{C}_k$, $|\mathcal{C}_n(w)|$ is maximized by $w = 0^k$.