# Congruence Properties of Combinatorial Sequences via Walnut and the Rowland-Yassawi-Zeilberger Automaton

Narad Rampersad\*

Department of Mathematics and Statistics University of Winnipeg Winnipeg, MB, Canada

n.rampersad@uwinnipeg.ca

Jeffrey Shallit<sup>†</sup>

School of Computer Science University of Waterloo Waterloo, ON, Canada

shallit@uwaterloo.ca

Submitted: Oct 18, 2021; Accepted: Jul 22, 2022; Published: Aug 12, 2022 © The authors. Released under the CC BY-ND license (International 4.0).

#### Abstract

Certain famous combinatorial sequences, such as the Catalan numbers and the Motzkin numbers, when taken modulo a prime power, can be computed by finite automata. Many theorems about such sequences can therefore be proved using Walnut, which is an implementation of a decision procedure for proving various properties of automatic sequences. In this paper we explore some results (old and new) that can be proved using this method.

Mathematics Subject Classifications: 68R15

### 1 Introduction

We study the properties of two famous combinatorial sequences. For  $n \geq 0$ , let

$$C_n = \frac{1}{n+1} \binom{2n}{n}$$

denote the n-th Catalan number and let

$$M_n = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} C_k$$

denote the n-th Motzkin number. For more about the Catalan numbers, see [17].

<sup>\*</sup>Supported by NSERC Discovery Grant 2019-04111.

<sup>&</sup>lt;sup>†</sup>Supported by NSERC Discovery Grant 2018-04118.

Many authors have studied congruence properties of these and other sequences modulo primes p or prime powers  $p^{\alpha}$ . Notably, Alter and Kubota [2] studied the Catalan numbers modulo p, and Deutsch and Sagan [9] studied many sequences, including the Catalan numbers, Motzkin numbers, Central Delannoy numbers, Apéry numbers, etc., modulo certain prime powers. Eu, Liu, and Yeh [10] studied the Catalan and Motzkin numbers modulo 4 and 8, and Krattenthaler and Müller [11] studied the Motzkin numbers and related sequences modulo powers of 2. Rowland and Yassawi [14] and Rowland and Zeilberger [15] gave different methods to compute finite automata that compute the sequences  $(C_n \mod p^{\alpha})_{n\geqslant 0}$  and  $(M_n \mod p^{\alpha})_{n\geqslant 0}$  (and many other similar sequences), where  $p^{\alpha}$  is a prime power. Rowland and Zeilberger provide a number of these automata for different  $p^{\alpha}$  at the website

https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/meta.html

along with the Maple code used to compute them. We use some of these automata, along with the program Walnut [12], available at the website

to study properties of these sequences.

We use the Rowland-Zeilberger algorithm (and Walnut) as a black-box, so we do not discuss the theory behind it. We just mention that this algorithm applies to any sequence of numbers that can be defined as the *constant term* of  $[P(x)]^nQ(x)$ , where P and Q are Laurent polynomials. In particular, the n-th Catalan number is the constant term of  $(1/x + 2 + x)^n(1 - x)$  and the n-th Motzkin number is the constant term of  $(1/x + 1 + x)^n(1 - x^2)$ .

Note that the automata produced by this method read their input in least-significant-digit-first format. All of the automata in this paper therefore also follow this convention. We use the notation  $(n)_k$  to denote the base-k representation of n in the lsd-first format.

Burns has posted several manuscripts to the arXiv [3, 4, 5, 6, 7, 8] in which he investigates various properties of the Catalan and Motzkin numbers modulo primes p by analyzing structural properties of automata computed using the Rowland–Yassawi algorithm. This paper takes a similar approach, but we use Walnut to simplify/automate much of the analysis.

### 2 Motzkin numbers

Deutsch and Sagan [9] gave a characterization of  $\mathbf{m}_2 = (M_n \mod 2)_{n \ge 0}$  that involves the Thue–Morse sequence

$$\mathbf{t} = (t_n)_{n \ge 0} = (0, 1, 1, 0, 1, 0, 0, 1, \ldots).$$

Let

$$\mathbf{c} = (c_n)_{n \ge 0} = (1, 3, 4, 5, 7, \ldots)$$

denote the starting positions of the "runs" in  $\mathbf{t}$ , excluding the first run (which, of course, starts at position 0).

**Theorem 1** (Deutsch and Sagan). The Motzkin number  $M_n$  is even if and only if either  $n \in 4\mathbf{c} - 2$  or  $n \in 4\mathbf{c} - 1$ .

*Proof.* We can prove this result using Walnut. The Rowland–Zeilberger algorithm produces the automaton in Figure 1, which, when fed with  $(n)_2$ , outputs  $M_n$  mod 2.

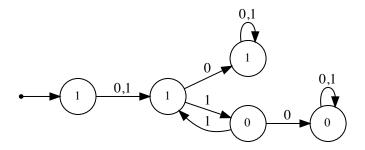


Figure 1: Automaton for  $M_n \mod 2$ .

Next we use Walnut to construct an automaton for the sequence  $\mathbf{c}$ . The commands

produce the automaton  $tm\_block\_start$  given in Figure 2, which computes c. We see that the elements of c are

 $\{m4^k : m \text{ is odd and } k \geqslant 0\}.$ 

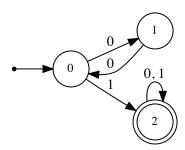


Figure 2: Automaton for starting positions of "runs" in t.

To complete the proof of the theorem, it suffices to execute the Walnut command eval even\_mot "?lsd\_2 An (MOT2[n]=@0 <=> Ei \$tm\_block\_start(i) & (n+2=4\*i | n+1=4\*i))":

which produces the output "TRUE".

Deutsch and Sagan also characterized  $\mathbf{m}_3 = (M_n \mod 3)_{n \ge 0}$ :

**Theorem 2.** (Deutsch and Sagan) The Motzkin number  $M_n$  satisfies

$$M_n \equiv_3 \begin{cases} 1, & \text{if either } (n)_3 = 0w, w \in \{0, 1\}^* \text{ or } (n+2)_3 = 0w, w \in \{0, 1\}^*, \\ 2, & \text{if } (n+1)_3 = 0w, w \in \{0, 1\}^*, \\ 0, & \text{otherwise.} \end{cases}$$

This can also be obtained directly from the automaton for  $\mathbf{m}_3$ . If we examine  $\mathbf{m}_5 = (M_n \mod 5)_{n \geqslant 0}$ , however, we discover that its behaviour is very different from that of  $\mathbf{m}_3$ . Deutsch and Sagan determined the positions of the 0's in  $\mathbf{m}_5$ .

**Theorem 3.** (Deutsch and Sagan) The Motzkin number  $M_n$  is divisible by 5 if and only if n is of the form

$$(5i+1)5^{2j}-2$$
,  $(5i+2)5^{2j-1}-1$   $(5i+3)5^{2j-1}-2$   $(5i+4)5^{2j}-1$ .

*Proof.* The Rowland–Zeilberger algorithm gives the automaton in Figure 3, which fully characterizes  $\mathbf{m}_5$ .

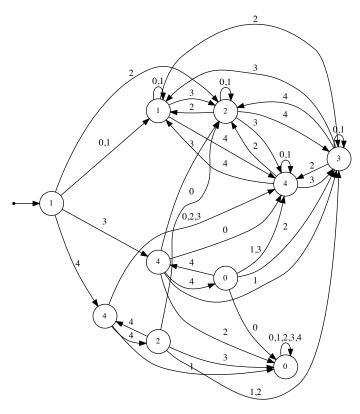


Figure 3: Automaton for  $M_n \mod 5$ .

The Walnut command

eval mot5mod0 "?lsd\_5 MOT5[n]=@0":

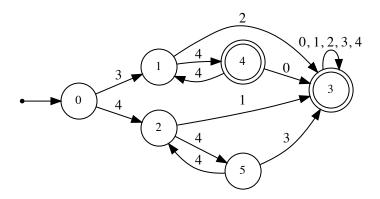


Figure 4: Automaton for the positions of the 0's in  $\mathbf{m}_5$ .

produces the automaton in Figure 4, from which one easily derives the result.

One notices that  $\mathbf{m}_3$  contains arbitrarily large runs of 0's, whereas  $\mathbf{m}_5$  does not have this property. We can use Walnut to determine the types of repetitions that are present in  $\mathbf{m}_5$ , but we first need to introduce some definitions.

Let  $w = w_1 w_2 \cdots w_n$  be a word of length n and period p; i.e.,  $w_i = w_{i+p}$  for i = 1 $0, \ldots, n-p$ . If p is the smallest period of w, we say that the exponent of w is n/p. We also say that w is an (n/p)-power of order p. Words of exponent 2 (resp. 3) are called squares (resp. cubes). If  $\mathbf{x}$  is an infinite sequence, we define the critical exponent of  $\mathbf{x}$  as

 $\sup\{e \in \mathbb{Q} : \text{ there is a factor of } \mathbf{x} \text{ with exponent } e\}.$ 

**Theorem 4.** The sequence  $\mathbf{m}_5$  has critical exponent 3. Furthermore, the only cubes in **m**<sub>5</sub> are 111, 222, 333, and 444.

*Proof.* We execute the Walnut commands

```
eval tmp "?lsd_5 Ei,n (n>=1) & At (t<=2*n) => MOT5[i+t]=MOT5[i+t+n]":
eval tmp "?lsd_5 Ei (n>=1) & At (t<2*n) => MOT5[i+t]=MOT5[i+t+n]":
```

and note that the first outputs "FALSE", indicating that  $\mathbf{m}_5$  has no factors of exponent larger than 3, and the second produces an automaton that only accepts n=1, indicating that the only cubes in  $\mathbf{m}_5$  have order 1. By inspecting a prefix of  $\mathbf{m}_5$ , one sees that 111, 222, 333, and 444 all occur.

We can also prove that every pattern of residues that appears in  $\mathbf{m}_5$  appears infinitely often, and furthermore, we can give a bound on when the next occurrence of a pattern will appear in  $\mathbf{m}_5$ . We say that a sequence  $\mathbf{x}$  is uniformly recurrent if for every factor w of  $\mathbf{x}$ , there is a constant c such that every occurrence of w in  $\mathbf{x}$  is followed by another occurrence of w at distance at most c.

Note that  $\mathbf{m}_3$  is *not* uniformly recurrent. This is due to the presence of arbitrarily large runs of 0's in  $\mathbf{m}_3$ . On the other hand, the sequence  $\mathbf{m}_5$  exhibits rather different behaviour.

**Theorem 5.** The sequence  $\mathbf{m}_5$  is uniformly recurrent. Furthermore, if w has length n and occurs at position i in  $\mathbf{m}_5$ , then there is another occurrence of w at some position j, where  $i < j \le i + 200n$ . The bound 200n cannot be replaced by 200n - 1.

*Proof.* This is proved with the Walnut commands

noting that the first eval command returns "TRUE" and the second returns "FALSE".  $\Box$ 

Burns [7] studied  $\mathbf{m}_p$  for p between 7 and 29 using automata computed using the Rowland–Yassawi algorithm. Among other things, his work suggests that depending on the value of p, the sequence  $\mathbf{m}_p$  either behaves like  $\mathbf{m}_3$ , where 0 has density 1 (i.e., p = 7, 17, 19), or  $\mathbf{m}_p$  behaves like  $\mathbf{m}_5$ , where 0 has density < 1 (i.e., p = 11, 13, 23, 29). Many of Burns' results could also be obtained using Walnut.

**Problem 6.** Characterize the primes p for which  $\mathbf{m}_p$  is uniformly recurrent.

Indeed, based on Burns' results and the discussion in the next section, we guess that the answer to this problem is given by the sequence

$$2, 5, 11, 13, 23, 29, 31, 37, 53, \dots$$

of primes that do not divide any central trinomial number. This is sequence  $\underline{A113305}$  of [16].

### 3 Central trinomial coefficients

The Motzkin numbers are closely related to the central trinomial coefficients  $T_n$ . The usual definition of  $T_n$  is as the coefficient of  $x^n$  in  $(1 + x + x^2)^n$ , but the definition

$$T_n = \sum_{k \geqslant 0} \binom{n}{2k} \binom{2k}{k}$$

better illustrates the connection between these numbers and the Motzkin numbers. The number  $T_n$  is also the constant term of  $(1/x + 1 + x)^n$ , which is the form needed for the Rowland–Zeilberger algorithm. Deutsch and Sagan studied the divisibility of  $T_n$  modulo primes and Noe [13] did the same for generalized central trinomial numbers.

**Theorem 7** (Deutsch and Sagan). The central trinomial coefficient  $T_n$  satisfies

$$T_n \equiv_3 \begin{cases} 1, & \text{if } (n)_3 \text{ does not contain a 2;} \\ 0, & \text{otherwise.} \end{cases}$$

Deutsch and Sagan proved this by an application of Lucas' Theorem; it is also immediate from the automaton produced by the Rowland–Zeilberger algorithm. As with the Motzkin numbers, the behaviour of  $T_n$  modulo 5 is rather different from that modulo 3. We collect some properties below (compare with those of  $\mathbf{m}_5$  from the previous section).

**Theorem 8.** Let  $\mathbf{t}_5 = (T_n \mod 5)_{n \ge 0}$ . Then

- 1.  $\mathbf{t}_5$  does not contain 0 (i.e.,  $T_n$  is never divisible by 5);
- 2.  $\mathbf{t}_5$  has critical exponent 3; furthermore, the only cubes in  $\mathbf{t}_5$  are 111, 222, 333, and 444:
- 3.  $\mathbf{t}_5$  is uniformly recurrent; Furthermore, if w has length n and occurs at position i in  $\mathbf{t}_5$ , then there is another occurrence of w at some position j, where  $i < j \leq i + 200n 192$ . The constant 192 cannot be replaced with 193.
- 4. If w has length n and appears in  $\mathbf{t}_5$ , then w appears in the prefix of  $\mathbf{t}_5$  of length 121n. The quantity 121n cannot be replaced with 121n 1.

*Proof.* Properties 1)–3) can all be obtained by similar Walnut commands to those used in the previous section for the Motzkin numbers. We just need the automaton for  $\mathbf{t}_5$ . The Rowland–Zeilberger algorithm gives the pleasantly symmetric automaton in Figure 5.

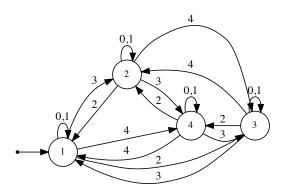


Figure 5: Automaton for  $T_n \mod 5$ .

For Property 4), we use the Walnut commands

```
def pr_tri5 "?lsd_5 Aj Ei i+n<=s & At t<n => TRI5[i+t]=TRI5[j+t]": eval tmp "?lsd_5 An pr_ti5(n,121*n)": eval tmp "?lsd_5 An pr_ti5(n,121*n-1)":
```

and note that the last two commands return "TRUE" and "FALSE", respectively.

We should note that in the special case of the central trinomial coefficients, it is not necessary to resort to either the Rowland–Zeilberger or Rowland–Yassawi algorithms to compute the automaton for  $T_n \mod p$ . Using the following result of Deutsch and Sagan, one can directly define the automaton for  $T_n \mod p$ .

**Theorem 9.** (Deutsch and Sagan) Let  $(n)_p = n_0 n_1 \cdots n_r$ . Then

$$T_n \equiv_p \prod_{i=0}^r T_{n_i}.$$

An immediate consequence is that  $T_n$  is divisible by p if and only if one of the  $T_{n_i}$  is divisible by p. This criterion allows one to determine the primes that do not divide any central trinomial coefficient; i.e., those in  $\underline{A113305}$  of [16], which we conjectured in the previous section to be the ones that answer the question of Problem 6.

We can also give the following sufficient condition for  $\mathbf{t}_p = (T_n \mod p)_{n \ge 0}$  to be uniformly recurrent. For  $i = 0, \dots, p-1$ , let  $\tau_i = T_i \mod p$ .

**Theorem 10.** Let p be prime and let  $\Sigma = \{\tau_i : i = 0, ..., p-1\}$ . If  $\Sigma$  does not contain 0 but does contain a primitive root modulo p, then  $\mathbf{t}_p$  is uniformly recurrent.

*Proof.* Clearly the order of the product in Theorem 9 does not matter; it follows then that Theorem 9 holds for the most-significant-digit-first representation of n, as well as the least-significant-digit-first representation. If we consider Theorem 9 with n written in msd-first notation, we see that  $\mathbf{t}_p$  is generated by iterating the morphism  $f: \Sigma^* \to \Sigma^*$  defined by

$$f(\tau_i) = (\tau_i \tau_0 \bmod p)(\tau_i \tau_1 \bmod p) \cdots (\tau_i \tau_{p-1} \bmod p)$$

for 
$$i = 0, ..., p - 1$$
; i.e.,  $\mathbf{t}_p = f^{\omega}(1)$ .

Recall that if there exists t such that for every  $a, b \in \Sigma$  the word  $f^t(a)$  contains b, we say that f is a primitive morphism. Now  $\tau_0 = 1$ , so for  $i = 0, \ldots, p-1$ , we can write  $f(\tau_i) = \tau_i x_i$  for some word  $x_i$ . It follows that  $f^p(\tau_i) = \tau_i f(x_i) f^2(x_i) \cdots f^{p-1}(x_i)$ . Furthermore, if  $0 \notin \Sigma$  and  $x_0$  contains a primitive root modulo p, then for every i, each non-zero residue modulo p appears in one of  $\tau_i, f(x_i), f^2(x_i), \ldots, f^{p-1}(x_i)$ . This proves that the morphism f is primitive. A standard result from the theory of morphic sequences states that any fixed point of a primitive morphism is uniformly recurrent [1, Theorem 10.9.5].

**Example 11.** For p = 5, we have  $(T_0, T_1, T_2, T_3, T_4) = (1, 1, 3, 7, 19)$ , so  $(\tau_0, \tau_1, \tau_2, \tau_3, \tau_4) = (1, 1, 3, 2, 4)$  contains the primitive root 2. The word

$$\mathbf{t}_5 = 113241132433412221434423111324 \cdots$$

is uniformly recurrent and is equal to  $f^{\omega}(1)$ , where f is the morphism defined by

 $1 \to 11324$ 

 $2 \rightarrow 22143$ 

 $3 \rightarrow 33412$ 

 $4 \to 44231.$ 

A computer calculation shows that for each prime p appearing in the list of initial values  $2, 5, 11, 13, \ldots, 479$  of  $\underline{A113305}$ , the first p terms of  $\mathbf{t}_p$  always contain a primitive root modulo p. Hence, each of these  $\mathbf{t}_p$ 's are uniformly recurrent.

## 4 Catalan numbers

Alter and Kubota [2] studied the sequences  $\mathbf{c}_p = (C_n \mod p)_{n \ge 0}$ , where p is prime. They proved that the runs of 0's in  $\mathbf{c}_p$  have lengths

$$\frac{p^{m+1+\delta_{3p}}-3}{2},\tag{1}$$

where  $\delta_{3p}$  is 1 when p=3 and 0 otherwise. This implies, of course, that for every prime p, the sequence  $\mathbf{c}_p$  is not uniformly recurrent. Alter and Kubota also proved that the blocks of non-zero values in  $\mathbf{c}_p$  have length

$$\frac{p+3(1+2\delta_{3p})}{2}.$$

For p = 3, Deutsch and Sagan [9, Theorem 5.2] gave a complete characterization of  $\mathbf{c}_3$ . We can obtain a similar characterization using Walnut.

**Theorem 12** (Deutsch and Sagan). The runs of 0's in  $c_3$  begin at positions n, where either

$$(n)_3 \in 211^* \text{ or } (n)_3 \in 211^*0\{0,1\}^*,$$

and have length  $(3^{i+2}-3)/2$ , where i is the length of the leftmost block of 1's in  $(n)_3$ . The blocks of non-zero values in  $\mathbf{c}_3$  are given by the following:

- The block 11222 occurs at position 0.
- The block 111222 occurs at all positions n where  $(n)_3 \in 222*0w$  for some  $w \in \{0, 1\}^*$  that contains an odd number of 1's.
- The block 222111 occurs at all positions n where  $(n)_3 \in 222^*0w$  for some  $w \in \{0, 1\}^*$  that contains an even number of 1's.

*Proof.* We use the automaton for  $\mathbf{c}_3$  given in Figure 6.

The Walnut command

produces the automaton in Figure 7. Examining the transition labels of the first component of the input gives the claimed representation for the starting positions of the runs of 0's and examining the transition labels of the second component gives the claimed length.

For the blocks of non-zero values, we execute the Walnut commands

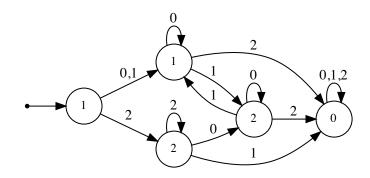


Figure 6: Automaton for  $C_n \mod 3$ .

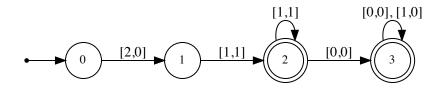


Figure 7: Automaton for runs of 0's in  $c_3$ .

Note that the length of the runs given in Theorem 12 is exactly what is given by the result of Alter and Kubota stated above in Eq. (1).

We can also perform the same calculation for p = 5 to obtain

**Theorem 13.** The runs of 0's in  $c_5$  begin at positions n, where either

$$(n)_5 \in 32^* \text{ or } (n)_5 \in 32^* \{0, 1\} \{0, 1, 2\}^*,$$

and have length  $(5^{i+2}-3)/2$ , where i is the length of the leftmost block of 2's in  $(n)_5$ .

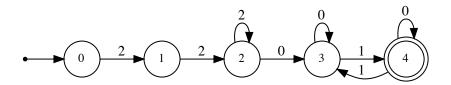


Figure 8: Automaton for blocks 111222 in  $c_3$ .

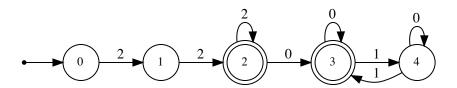


Figure 9: Automaton for blocks 222111 in  $c_3$ .

*Proof.* We use the automaton for  $c_5$  given in Figure 10. The Walnut command

produces the automaton in Figure 11. Examining the transition labels, as in the proof of Theorem 12, gives the result.  $\Box$ 

Again, note that the lengths of the runs match what is given by Eq. (1).

**Theorem 14.** The sequence  $c_5$  begins with the non-zero block 112. The other non-zero blocks in  $c_5$  are 1331, 2112, 3443, and 4224.

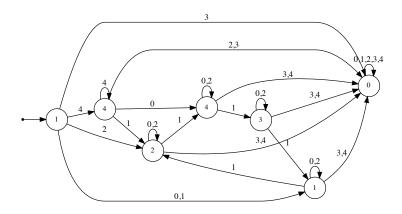


Figure 10: Automaton for  $C_n \mod 5$ .

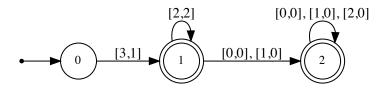


Figure 11: Automaton for runs of 0's in  $c_5$ .

*Proof.* The Walnut command

```
eval cat5max1234 "?lsd_5 n>=1 & (At t<n => CAT5[i+t]!=@0) & CAT5[i+n]=@0 & (i=0|CAT5[i-1]=@0)":
```

produces the automaton in Figure 12. We see that the initial non-zero block has length 3

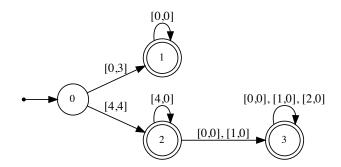


Figure 12: Automaton for non-zero blocks in  $c_5$ .

and all others have length 4. We omit the Walnut command to verify the values of these length 4 blocks, but it is easy to formulate.  $\Box$ 

## 5 Conclusion

We have shown how to use Walnut to obtain automated proofs of certain results in the literature concerning the Catalan and Motzkin numbers modulo p, as well as the central trinomial coefficients modulo p. We were also able to use Walnut to examine other properties of these sequences that have not previously been explored, such as the presence (or absence) of certain repetitive patterns and the property of being uniformly recurrent. We hope these results encourage other researchers to continue to further explore these properties for other sequences.

## References

- [1] J.-P. Allouche, J. Shallit, Automatic Sequences, Cambridge, 2003.
- [2] R. Alter, K. K. Kubota, Prime and prime power divisibility of Catalan numbers, J. Combin. Theory Ser. A 15 (1973), 243–256.
- [3] R. Burns, The Catalan numbers have no forbidden residue modulo primes. Preprint available at arXiv:1703.02705.
- [4] R. Burns, Structure and asymptotics for Catalan numbers modulo primes using automata. Preprint available at arXiv:1701.02975.
- [5] R. Burns, Asymptotic density of Catalan numbers modulo 3 and powers of 2. Preprint available at arXiv:1611.03705.
- [6] R. Burns, Structure and asymptotics for Motzkin numbers modulo primes using automata. Preprint available at arXiv:1703.00826.
- [7] R. Burns, Structure and asymptotics for Motzkin numbers modulo small primes using automata. Preprint available at arXiv:1612.08146.
- [8] R. Burns, Asymptotic density of Motzkin numbers modulo small primes. Preprint available at arXiv:1611.04910.
- [9] E. Deutsch, B. Sagan, Congruences for Catalan and Motzkin numbers and related sequences, J. Number Theory 117 (2006), 191–215.
- [10] S.-P. Eu, S.-C. Liu, Y.-N. Yeh, Catalan and Motzkin numbers modulo 4 and 8, European J. Combin. 29 (2008), 1449–1466.
- [11] C. Krattenthaler, T.W. Müller, Motzkin numbers and related sequences modulo powers of 2, European J. Combin. 73 (2018), 114–137.
- [12] H. Mousavi, Automatic theorem proving in Walnut. Documentation (2016–2021) available at arXiv:1603.06017.
- [13] T. Noe, On the divisibility of generalized central trinomial coefficients, J. Integer Seq. 9 (2006), Article 06.2.7.
- [14] E. Rowland, R. Yassawi, Automatic congruences for diagonals of rational functions, J. Théorie Nombres Bordeaux 27 (2015), 245–288.
- [15] E. Rowland, D. Zeilberger, A Case Study in Meta-AUTOMATION: AUTOMATIC Generation of Congruence AUTOMATA For Combinatorial Sequences, J. Difference Equations and Applications 20 (2014), 973–988.
- [16] N. J. A. Sloane et al., The On-Line Encyclopedia of Integer Sequences, 2021, available at https://oeis.org.
- [17] R. Stanley, Catalan Numbers, Cambridge University Press, 2015.