Hang Lung Mathematics Awards
Vol. 8 (2018), pp. 29–46

ON THE DIVISIBILITY OF CATALAN NUMBERS

TEAM MEMBER
Tsz Chung LI

TEACHER
Dr. Kit Wing YU

SCHOOL
United Christian College

Abstract. In this paper, we propound an efficacious method to derive the p-adic valuation of the Catalan number by analyzing the properties of the coefficients in the base p-expansion of n. We unearth a new connection between those coefficients and the p-adic valuation of the Catalan number. In fact, we have discovered that the highest power of p dividing the Catalan number is relevant to the number of digits greater than or equal to half of $p+1$, the nature of distribution of digits equal to half of $p-1$ and the frequency of carries when 1 is added to n. Meanwhile, we remark that the method we apply is more natural than the current way used by Alter and Kubota which is quite artificial.

Applications, examples of our new formula and details about Catalan numbers are also included in this paper.

1. Introduction

Catalan numbers, denoted by C_n, are a sequence of natural numbers which appear in many combinatorial and geometric problems, see, for examples, [11] and [16, Ch. 7]. This number was discovered by the Belgian mathematician Eugene Charles Catalan (1814-1894) when he studied the well-formed sequences of parentheses, see [11] and [16, Ch. 7]. Also the numbers were discovered by Euler when he investigated the problem of the number of triangulations of convex polygons ([8], [11] and [16, Ch. 7]). A brief history about this number is given in Appendix A.

Here is the definition of Catalan numbers:

Definition 1. *Let $C_0 = 1$. For $n \geq 1$, we define the Catalan numbers C_n by the recursion relation*

$$C_n = \sum_{k=0}^{n-1} C_k C_{n-1-k} = C_0 C_{n-1} + C_1 C_{n-2} + \cdots + C_{n-1} C_0$$

By applying the theory of generating functions and Segner's recurrence formula [10], Cn can be represented by the following explicit formula:

$$C_n = \frac{(2n)!}{(n+1)!n!} = \frac{1}{n+1}\binom{2n}{n} \tag{1}$$

where $\binom{n}{r}$ denotes the usual binomial coefficient and $n!$ the usual factorial. It is easy to deduce from the formula 1 that

$$C_{n+1} = \frac{4n-6}{n}C_n$$

holds for every positive integer n. We note that more classical formulas for Catalan numbers involving binomial coefficients can be found, for example, in [15].

Besides the study of proving useful identities connecting C_n and the binomial coefficient or other important constants, people were also interested in the parity of Catalan numbers. Let $n \geq 2$ be an integer and $v_p(n)$, called the p-adic valuation of n, be the highest power of p dividing n. By this notation, it is well-known that $v_2(C_n) = 0$ if and only if $n = 2k - 1$ for some positive integer k. In other words, we have

$$2 \nmid C_{2^k-1} \tag{2}$$

Several proofs of this result can be found in [1], [4], [6], [11] and [17]. For odd primes p, Alter and Kubota [2, Theorem 1] generalized (1.2) to the following result:

Theorem 2. *For odd primes p, we have $p \nmid C_{p^k-1}$.*

They further showed that the set $\{B_k(p)\}$, where each $B_k(p)$ denotes the consecutive Catalan numbers which are multiples of p, forms into blocks. The lengths and positions of each $B_k(p)$ are fully determined, see [2, Theorems 2, 3 & 5]. The main tool they applied is to consider Lemma 1 in [2]. In fact, they considered an associated sequence $\{c_n\}$ of $\{C_n\}$. Then they obtained a recurrence relation of c_n which involves a fraction. Next, by analyzing the factors of the denominator and numerator of such fraction, they were able to determine the relationship between $v_p(C_{n-1})$ and $v_p(C_{n-sp^m-1})$, where $sp^m \leq n < (s+1)p^m$ for integers $m \geq 0$ and $0 < s < p$. Furthermore, they expressed Lemma 1 in the following useful form [2, Theorem 7] and applied it to prove other interesting results [2, Theorems 8 & 9].

Theorem 3. *Suppose that $n = r_m p^m + \cdots + r_1 p + r_0$ where $0 \leq r_0, r_1, \ldots, r_m < p$. For every $i = 0, 1, \ldots, m$, we set $n_i = r_i p^i + \cdots + r_1 p + r_0$ and*

$$\epsilon_i = \begin{cases} 1, & \text{if } n_i \geq \frac{p^{i+1}+1}{2}; \\ 0, & \text{otherwise} \end{cases}$$

Then we have $v_p(C_{n-1}) = \sum_{i=0}^{m} \epsilon_i$.

Now Theorem B gives a way to find $v_p(C_n)$ in terms of the ni defined in it. However, the weakness of Theorem B is that such n_i and ϵ_i are constructed quite artificial. Thus one may ask: Does there exist any "natural" formula to find $u_p(C_n)$? To overcome this problem, it seems to the authors that Deutsch and Sagan [5,Theorem 2.1] are the first to answer this problem partially in 2006. In fact, they applied the theory of group actions to show that the largest power of 2 dividing the Catalan number C_n, namely $v_2(C_n)$, is given by

$$v_2(C_n) = \delta_2(n+1) - 1, \tag{3}$$

where $\delta_2(n)$ is the sum of the digits in the base-2 expansion of n. By the formula (3), one may wonder the existence of a connection between $v_p(C_n)$ and the digits of n for odd primes p. This kind of wonders gives birth to the paper you are reading.

In this paper, we investigate the divisibility of the Catalan numbers C_n by the odd prime p. In fact, we can find an explicit form of a greatest integer function by using Hermite's identity as the key. This gives a new method to establish a formula of $v_p(C_n)$ which is similar to Theorem B. Furthermore, by examining the formula (6) in details, we have discovered that $v_p(C_n)$ is closely connected to the nature of the digits of the base p-expansion of n (see formula (7) below). As a result, this gives us a new and efficient way to determine $v_p(C_n)$. The paper is organized as follows. In §2, two main results are stated. All the necessary lemmas for the proofs of the main results are given and shown in §3. The details of the proofs of Theorems 4 and 5 are given in §4 and §5 respectively. In §6, applications of Theorem 5 are given. In particular, we show that Theorem A is a special case of Theorem 5 and $p \nmid C_{\frac{p^k-1}{2^i}}$ for $1 \leq i \leq v_2(p-1)$. Several examples and remarks about the main results are discussed in §7. Appendix A is a brief history of Catalan numbers and Appendix B shows some applications of the numbers.

2. The main results

In the following discussion, we suppose that p is an odd prime and m, n are non-negative integers.

Theorem 4. *Suppose that*

$$n = r_m p^m + r_{m-1} p^{m-1} + \cdots + r_1 p + r_0 \tag{4}$$

is the base-p expansion of the integer n, where $0 \leq r_0, r_1, \ldots, r_m < p$. *Furthermore, for* $0 \leq i \leq m$, *suppose that* $n_i = r_i p^i + \cdots + r_1 p + r_0$ *and*

$$\epsilon_i = \begin{cases} 1, & \textit{if } n_i \geq \frac{p^{i+1}+1}{2}; \\ 0, & \textit{otherwise} \end{cases} \tag{5}$$

Then we have

$$v_p(C_n) = \sum_{i=0}^{m} \epsilon_i - N_p, \tag{6}$$

where N_p is the number of carries when 1 is added to n in base p.

Our next result tells us how to evaluate the summation in the formula (6). Before stating the result, we have to classify the digits of n into three classes. Let $S = \{r_0, r_1, \ldots, r_m\}$. Define subsets $S_1(p), S_2(p)$ and $S_3(p)$ of S by

$$S_1(p) = \left\{ r_i \in S \mid r_i \geq \frac{p+1}{2} \right\}, \quad S_2(p) = \left\{ r_i \in S \mid r_i = \frac{p-1}{2} \right\},$$
$$S_3(p) = \left\{ r_i \in S \mid r_i \leq \frac{p-3}{2} \right\},$$

Since every digit r_i must belong to one and only one $S_1(p), S_2(p)$ or $S_3(p)$, we have

$$S = S_1(p) \cup S_2(p) \cup S_3(p)$$

and $S_1(p), S_2(p)$ and $S_3(p)$ are disjoint.

Now we further decompose $S_2(p)$ into disjoint subsets. In fact, we arrange the digits of $S_2(p)$ in ascending order according to the subscripts and let i_1 be the *smallest* subscript of the digits in $S_2(p)$. Denote $T_1(p)$ to be the set of consecutive digits of $S_2(p)$ starting with the subscript i_1. Since $S_2(p)$ is finite, the smallest subscript of the digits exists in $S_2(p) \backslash T_1(p)$ and we call it r_{i_2}. Next, denote $T_2(p)$ to be the set of consecutive subscripts of $S_2(p) \backslash T_1(p)$ starting with i_2. Continuing this process $(k-2)$-times, we can express $S_2(p)$ in the following form:

$$S_2(p) = T_1(p) \cup T_2(p) \cup \cdots \cup T_k(p),$$

where $T_j(p)$ is the set of consecutive digits of $S_2(p) \backslash [T_1(p) \cup T_2(p) \cup \cdots \cup T_{j-1}(p)]$ starting with the subscript i_j for $j = 2, 3, \ldots, k$. By this construction, it is clear that $T_1(p), T_2(p), \ldots, T_k(p)$ are disjoint.

To get a better understanding of the structures of $S_2(p)$ and its subsets $T_1(p), T_2(p), \ldots, T_k(p)$ are shown as follows:

$$\overbrace{r_0, r_1, \ldots, r_{i_1-1}}^{S_1(p) \text{ or } S_3(p)}, \underbrace{r_{i_1}, r_{i_1+1}, \ldots, r_{i_1+|T_1(p)|-1}}_{T_1(p)}, \overbrace{r_{i_1+|T_1(p)|}, \ldots, r_{i_2-1}}^{S_1(p) \text{ or } S_3(p)},$$

$$\underbrace{r_{i_2}, r_{i_2+1}, \ldots, r_{i_2+|T_2(p)|-1}}_{T_2(p)}, \ldots\ldots\ldots\ldots, \overbrace{r_{i_{k-1}+|T_{k-1}(p)|}, \ldots, r_{i_k-1}}^{S_1(p) \text{ or } S_3(p)},$$

$$\underbrace{r_{i_k}, r_{i_k+1}, \ldots, r_{i_k+|T_k(p)|-1}}_{T_k(p)}, \overbrace{r_{i_k+|T_k(p)|}, \ldots, r_m}^{S_1(p) \text{ or } S_3(p)},$$

Now we are ready to state our next main theorem:

Theorem 5. *Suppose that $B(p) = \{r_{i_1-1}, r_{i_2-1}, \ldots, r_{i_k-1}\} \cap S_1(p)$. Then we have*

$$v_p(C_n) = |S_1(p)| - N_p + \sum_{r_{i_q}-1 \in B(p)} |T_q(p)|, \tag{7}$$

where N_p is the number of carries when 1 is added to n in base p and the absolute value refers to the cardinality of the set.

3. Preliminary lemmas

The proof of Theorem 4 depends solely on the famous Legendre's formula and a special of the classical Hermite's identity. We state them in Lemmas 6 and 7 respectively.

Lemma 6. *(Legendre's formula, 1808) Suppose that p is prime and n is a positive integer. Then we have*

$$v_p(n!) = \sum_{k=1}^{\infty} \left[\frac{n}{p^k}\right] = \frac{n - s_p(n)}{p-1},$$

where $[x]$ denotes the greatest integer less than or equal to x and $s_p(n)$ is the sum of the digits of n in the base-p expansion.

Lemma 7. *For every real number x, we have*

$$[2x] = [x] + \left[x + \frac{1}{2}\right]$$

Proof. We observe that both sides of the equation increase by 2 if x increases by 1. Thus it suffices to prove the equation when $0 \le x < 1$. If $0 \le x < \frac{1}{2}$, then we have $0 \le 2x < 1$ and $\frac{1}{2} \le x + \frac{1}{2} < 1$ which mean that

$$[2x] = 0 = 0 + 0 = [x] + \left[x + \frac{1}{2}\right]$$

If $\frac{1}{2} \le x < 1$, then we have $1 \le 2x < 2$ and $1 \le x + \frac{1}{2} < \frac{3}{2} < 2$. Therefore we get

$$[2x] = 1 = 0 + 1 = [x] + \left[x + \frac{1}{2}\right]$$

This finishes the proof of the lemma. □

Next we evaluate the value of $u_p(n+1)$ which will be used in the proof of Lemma 8.

Lemma 8. *Suppose that n is the integer in the form (2.1). Then $v_p(n+1)$ equals to the number of carries when 1 is added to n in base p.*

Proof. It is clear that a carry exists when 1 is added to n if and only if $r_0 = p-1$. Suppose that such a carry exists and N_p is the *largest* positive integer such that

$$r_0 = r_1 = \cdots = r_{N_p-1} = p-1$$

but $0 \le r_{N_p} < p-1$. Thus we have

$$\begin{aligned} n+1 &= r_m p^m + \cdots + r_{N_p} p^{N_p} + (p-1)p^{N_p-1} + \cdots + (p-1)p + (p-1) + 1 \\ &= r_m p^m + \cdots + (r_{N_p}+1)p^{N_p} \\ &= [r_m p^{m-N_p} + \cdots + (r_{N_p}+1)]p^{N_p} \end{aligned} \tag{8}$$

which implies that

$$v_p(n+1) = N_p$$

Since N_p carries happen in the sum (8), the desired result follows and the proof is complete. □

4. Proof of Theorem 4

By the formula (1) and Lemma 6, we have

$$\begin{aligned} v_p(C_n) &= v_p((2n)!) - 2v_p(n!) - v_p(n+1) \\ &= \sum_{k=1}^{\infty} \left[\frac{2n}{p^k}\right] - 2v_p(n!) - v_p(n+1) \end{aligned} \tag{9}$$

By Lemma 7, we can rewrite the expression (9) as

$$v_p(C_n) = \sum_{k=1}^{\infty} \left[\frac{n}{p^k} + \frac{1}{2}\right] - v_p(n!) - v_p(n+1) \tag{10}$$

Now we are going to find an explicit formula of the sum on the right-hand side in the expression (10). Note that we can split it into two sums:

$$\sum_{k=1}^{\infty} \left[\frac{n}{p^k} + \frac{1}{2}\right] = \sum_{k=1}^{m+1} \left[\frac{n}{p^k} + \frac{1}{2}\right] + \sum_{k=m+2}^{\infty} \left[\frac{n}{p^k} + \frac{1}{2}\right]$$

Since $0 \le r_0, r_1, \ldots, r_m \le p-1$, it is obvious that we have

$$n = r_m p^m + \cdots + r_1 p + r_0 \le (p-1)(p^m + \cdots + p + 1) = p^{m+1} - 1$$

By this result and the fact that $p \ge 3$, if $k \ge m+2$, then we know that

$$0 \le \frac{n}{p^k} + \frac{1}{2} \le \frac{n}{p^{m+2}} + \frac{1}{2} \le \frac{p^{m+1}-1}{p^{m+2}} + \frac{1}{2} = \frac{1}{p} + \frac{1}{2} - \frac{1}{p^{m+2}} < 1$$

In other words, we have

$$\sum_{k=m+2}^{\infty} \left[\frac{n}{p^k} + \frac{1}{2}\right] = 0$$

Next, it is clear from the base-p expansion (4) that

$$\begin{aligned}\sum_{k=1}^{m+1}\left[\frac{n}{p^k}+\frac{1}{2}\right] &= \left[r_m p^{m-1}+\dots+r_1+\left(\frac{r_0}{p}+\frac{1}{2}\right)\right]\\ &\quad+\left[r_m p^{m-2}+\dots+r_2+\left(\frac{r_1}{p}+\frac{r_0}{p^2}+\frac{1}{2}\right)\right]\\ &\quad+\dots+\left[\frac{r_m p^m+\dots+r_1 p+r_0}{p^{m+1}}+\frac{1}{2}\right]\\ &= (r_m p^{m-1}+\dots+r_1)+(r_m p^{m-2}+\dots+r_2)+\dots+r_m\\ &\quad+\left[\frac{n_0}{p}+\frac{1}{2}\right]+\left[\frac{n_1}{p^2}+\frac{1}{2}\right]+\dots+\left[\frac{n_m}{p^{m+1}}+\frac{1}{2}\right]\\ &= r_m(p^{m-1}+\dots+p+1)+r_{m-1}(p^{m-2}+\dots+p+1)+\dots\\ &\qquad+r_2(p+1)+r_1+\sum_{i=0}^{m}\epsilon_i\\ &= r_m\cdot\frac{p^m-1}{p-1}+r_{m-1}\cdot\frac{p^{m-1}-1}{p-1}+\dots+r_2\cdot\frac{p^2-1}{p-1}+r_1+\sum_{i=0}^{m}\epsilon_i\\ &= \frac{(r_m p^m+\dots+r_2p^2+r_1p)-(r_m+\dots+r_2+r_1)}{p-1}+\sum_{i=0}^{m}\epsilon_i\\ &= \frac{n-s_p(n)}{p-1}+\sum_{i=0}^{m}\epsilon_i \qquad (11)\end{aligned}$$

Thus we derive from the expressions (10) and (11) with the help of Lemma 6 that

$$v_p(C_n)=\sum_{i=0}^{m}\epsilon_i - v_p(n+1) \qquad (12)$$

By Lemma 8, we know that $v_p(n+1)=N_p$, the number of carries when 1 is added to n in base p. Hence we have established the formula (6) from the expression (12) and this completes the proof of Theorem 4.

5. Proof of Theorem 5

Suppose that n takes the form (4) and each ϵ_i is defined by (5). We have to evaluate the sum

$$\sum_{i=0}^{m}\epsilon_i$$

and this depends on how the digits in $S_1(p), S_2(p)$ and $S_3(p)$ contribute to the p-adic $v_p(C_n)$ respectively.

Here we analyze the effects of $S_1(p)$ and $S_3(p)$ on $v_p(C_n)$ first. If $r_i \in S_1(p)$, then since $0 \le r_0, r_1, \ldots, r_{i-1} \le p-1$, we have

$$n_i = r_i p^i + \cdots + r_1 p + r_0 \ge \frac{p+1}{2} p^i > \frac{p^{i+1}}{2} \tag{13}$$

Next, suppose that $r_i \in S_3(p)$. Then we have

$$\begin{aligned} n_i &\le \left(\frac{p-3}{2}\right) p^i + (p-1)(p^{i-1} + \cdots + p + 1) \\ &= \frac{p^{i+1}}{2} - \frac{3p^i}{2} + p^i - 1 \\ n_i &< \frac{p^{i+1}}{2} \end{aligned} \tag{14}$$

Thus we see immediately from the inequalities (13) and (14) that

$$\epsilon_i = \begin{cases} 1, & \text{if } r_i \in S_1(p); \\ 0, & \text{if } r_i \in S_3(p) \end{cases} \tag{15}$$

In other words, we can conclude from the result (15) that the following fact holds:

Fact 1. Every digit in $S_1(p)$ contributes 1 to $v_p(C_n)$, but digits of $S_3(p)$ do nothing to $v_p(C_n)$.

Now our proof will be complete if we can show that the contribution of $S_2(p)$ to $v_p(C_n)$ is exactly

$$\sum_{r_{i_q}-1 \in B(p)} |T_q(p)|$$

To this end, we suppose that $q \in \{1, 2, \ldots, k\}$. If $r_{i_q-1} \in S_1(p)$, then we know from the inequality (13) that

$$n_{i_q-1} > \frac{p^{i_q}}{2} \quad \text{or equivalently} \quad \epsilon_{i_q-1} = 1 \tag{16}$$

Similarly, if $r_{i_q-1} \in S_3(p)$, then the inequality (14) implies that

$$n_{i_q-1} < \frac{p^{i_q}}{2} \quad \text{or equivalently} \quad \epsilon_{i_q-1} = 0 \tag{17}$$

Based on the inequality (16), we can derive easily that

$$\begin{aligned} n_j &= r_j p^j + \cdots + r_{i_q} p^{i_q} + n_{i_q-1} \\ &= \left(\frac{p-1}{2}\right)(p^j + \cdots + p^{i_q}) + n_{i_q-1} \\ &= \frac{p^{j+1} - p^{i_q}}{2} + n_{i_q-1} \\ n_j &> \frac{p^{j+1}}{2} \end{aligned} \tag{18}$$

where $j = i_q, i_{q+1}, \ldots, i_q + |T_q(p)| - 1$. In other words, these mean that $\epsilon_j = 1$ for $j = i_q, i_{q+1}, \ldots, i_q + |T_q(p)| - 1$. Similarly, it follows from the inequality (17) that

$$n_j = r_j p^j + \cdots + r_{i_q} p^{i_q} + n_{i_q - 1} = \frac{p^{j+1} - p^{i_q}}{2} + n_{i_q-1} < \frac{p^{j+1}}{2} \tag{19}$$

so that $\epsilon_j = 0$, where $j = i_q, i_q + 1, \ldots, i_q + |T_q(p)| - 1$. Thus we obtain from the inequalities (18) and (19) the following fact:

Fact 2. If $r_{i_q-1} \in B(p) = \{r_{i_1-1}, r_{i_2-1}, \ldots, r_{i_k-1}\} \cap S_1(p)$, then it contributes totally $|T_q(p)|$ to $v_p(C_n)$; otherwise, it plays no role in $v_p(C_n)$.

Hence, combining **Fact 1** and **Fact 2**, we are able to show that the formula (7) holds and this completes the proof of the theorem.

Remark 9. *It is not hard to see that our Theorem 5 is more convenient and useful than Theorem 3 for computing $v_p(C_n)$ because we only check the $(m+1)$ digits $r_0, r_1, \ldots, r_m$ rather than the $(m+1)$ sets of inequalities. Besides, it reveals the much more "natural"' and important connection between $v_p(C_n)$ and the digits of n in base-p expansion.*

6. Applications of Theorem 5

In this section, we give two applications of our Theorem 5. The first corollary shows that Theorem 2 is a special of Theorem 5. The second corollary is a new result which is similar to Theorem 2 in a certain sense.

Corollary 10. *For odd primes p and $k \in \mathbb{N}$, we have $p \nmid C_{p^k-1}$.*

Proof. Since

$$\begin{aligned} p^k - 1 &= (p-1)(p^{k-1} + \cdots + p + 1) \\ &= (p-1)p^{k-1} + \cdots + (p-1)p + (p-1) \end{aligned} \tag{20}$$

we have

$$r_0 = r_1 = \cdots = r_{k-1} = p - 1 \geq \frac{p+1}{2}$$

In this case, we have $|S_1(p)| = k$ and $B(p) = \emptyset$. In addition, we know from the expansion (20) that there are exactly k carries when 1 is added to $p^k - 1$. By Theorem 5, we certainly have

$$v_p(C_{p^k-1}) = 0$$

□

Corollary 11. *For odd primes p and $k \in \mathbb{N}$, we have*

$$p \nmid C_{\frac{p^k-1}{2^i}},$$

where $i = 1, 2, \ldots, v_2(p-1)$.

Proof. Since p is odd, $p-1$ is even and thus $v_2(p-1) \geq 1$. Since $1 \leq i \leq v_2(p-1)$ and

$$\frac{p^k - 1}{2^i} = \frac{p-1}{2^i}(p^{k-1} + \cdots + p + 1), \tag{21}$$

$\dfrac{p^k-1}{2^i}$ are integers for $1 \leq i \leq v_2(p-1)$. If $j = p - \frac{p-1}{2^{i-1}}$ for $1 \leq i \leq v_2(p-1)$, then we have

$$\frac{p-j}{2} = \frac{p-1}{2^i} \tag{22}$$

Combining the expressions (21) and (22), we derive that

$$\frac{p^k - 1}{2^i} = \frac{p-j}{2}(p^{k-1} + \cdots + p + 1), \tag{23}$$

where $j = p - \frac{p-1}{2^{i-1}}$ for $1 \leq i \leq v_2(p-1)$. There are two cases:

- **Case (1):** $p = 3$. In this case, we have $i = j = 1$. Thus we know from the expression (21) that

$$\frac{3^k - 1}{2} = \left(\frac{3-1}{2}\right)(3^{k-1} + \cdots + 3 + 1)$$

which imply that

$$r_0 = r_1 = \cdots = r_{k-1} = \frac{3-1}{2} \tag{24}$$

Therefore, $S_1(3) = \emptyset$ and then $B(3) = \emptyset$. By the digits (24), it is obvious that no carry exists when 1 is added to $\frac{3^k-1}{2}$ so that $N_3 = 0$. Hence it derives from Theorem 5 that $v_3(C_{\frac{3^k-1}{2}}) = 0$, i.e.,

$$3 \nmid C_{\frac{3^k-1}{2}}$$

- **Case(2):** $p \geq 5$. If $i = 1$, then we have $j = p - (p-1) = 1$ and the expression (23) gives

$$\frac{p^k - 1}{2} = \left(\frac{p-1}{2}\right)p^{k-1} + \cdots + \left(\frac{p-1}{2}\right)p + \frac{p-1}{2}$$

Similar to **Case (1)**, we know that

$$r_0 = r_1 = \cdots = r_{k-1} = \frac{p-1}{2} \tag{25}$$

Thus $S_1(p) = \emptyset$ and $B(p) = \emptyset$. Now the digits (25) show that no carry exists when 1 is added to $\frac{p^k-1}{2}$ so that $N_p = 0$. By Theorem 5, we have $v_p(C_{p^k-12})$, i.e.,

$$p \nmid C_{p^k-12}$$

Next, we suppose that $2 \le i \le v_2(p-1)$. Recall the definition of j and the fact that $p \ge 5$, we have

$$j = p - \frac{p-1}{2^{i-1}} \ge p - \frac{p-1}{2} = \frac{p+1}{2} \ge 3$$

Thus we follow from this and the expression (23) that $S_1(p) = \emptyset$. By similar argument as above, we see that $N_p = 0$. By Theorem 5, we get

$$p \nmid C_{\frac{p^k-1}{2^i}}$$

where $2 \le i \le v_2(p-1)$.

This completes the proof of the corollary. □

For instance, let $p = 5$ and $k = 2$. Then $\frac{5^2-1}{2} = 12$ and $\frac{5^2-1}{2^2} = 6$, so we have

$$C_{12} = 208012 = 2^2 \times 7 \times 17 \times 19 \times 23 \quad \text{and} \quad C_6 = 132 = 2^2 \times 3 \times 11$$

Hence they imply that $5 \nmid C_{12}$ and $5 \nmid C_6$. However, since $C_3 = 5$, Corollary 11 is not true for $i = v_2(p-1) + 1$.

7. Examples and concluding remarks

In this section, several examples are considered and other studies related to the divisibility of C_n are also discussed. The following two examples explain how efficiency our main formula (7) is.

Example 12. *Consider the number $n = 9936$. We want to evaluate $v_3(C_{9936})$ and $v_5(C_{9936})$ by using the formula (7) in Theorem 5. We first factorize C_{9936}. In fact, we know from the website [20] that*

Then we paste this number to another website [21] to get its factorization:

$$C_{9936} = 2^6 \times 3^6 \times 5^2 \times 7^3 \times \cdots \times 19867 \tag{26}$$

(a) *To find $v_3(C_{9936})$, we note that*

$$\begin{aligned} 9936 = {} & 0 \times 3^0 + 0 \times 3^1 + 0 \times 3^2 + 2 \times 3^3 + 2 \times 3^4 + 1 \times 3^5 + 1 \times 3^6 \\ & + 1 \times 3^7 + 1 \times 3^8 \end{aligned} \tag{27}$$

By the expansion (27), we list the sequence of digits of 9936 as follows:

Therefore, we have $|S_1(3)| = 2$ and $B(3) = \{2\}$. It is easy to see from the expansion (27) that no carry exists when 1 is added to 9936, so $N_3 = 0$. Hence they imply that

$$v_3(C_{9936}) = |S_1(3)| - N_3 + |T_1(3)| = 2 - 0 + 4 = 6$$

which is consistent with the factorization (26).

(b) *Similarly, we see that*

$$9936 = 1 \times 5^0 + 2 \times 5 + 2 \times 5^2 + 4 \times 5^3 + 0 \times 5^4 + 3 \times 5^5 \tag{28}$$

By the expansion (28), we list the sequence of digits of 9936 as follows:

Notice that $|S_1(5)| = 2, B(5) = \emptyset$ *and* $N_5 = 0$, *so we follow from the formula (7) that*

$$v_5(C_{9936}) = 2$$

which is also consistent with the factorization (26).

Example 13. *Now we consider another number* $n = 11186$. *Similar as Example 12, we want to evaluate* $v_3(C_{11186})$ *and* $v_5(C_{11186})$. *By the website [20] again, we have*

Next, the website [21] gives its factorization:

$$C_{11186} = 2^8 \times 3^2 \times 5^3 \times 7^3 \times \cdots \times 22369 \tag{29}$$

(a) *Note that*

$$\begin{aligned} 11186 = {} & 2 \times 3^0 + 2 \times 3^1 + 0 \times 3^2 + 0 \times 3^3 + 0 \times 3^4 + 1 \times 3^5 \\ & + 0 \times 3^6 + 2 \times 3^7 + 1 \times 3^8 \end{aligned} \tag{30}$$

It is easy to see from the expansion (30) that the sequence of its digits is given by

Thus we have $|S_1(3)| = 3$ *and* $B(3) = \{2\}$. *To compute* N_3, *we see that the first two digits are 2 so that two carries occur when 1 is added to 11186 in the 3-expansion (30). In this case,* $N_3 = 2$ *and the formula (7) shows that*

$$v_3(C_{11186}) = |S_1(3)| - N_3 + |T_2(3)| = 3 - 2 + 1 = 2 \tag{31}$$

Hence the result (31) is compatible with the factorization (29).

(b) *Similarly, we have*

$$11186 = 1 \times 5^0 + 2 \times 5 + 2 \times 5^2 + 4 \times 5^3 + 2 \times 5^4 + 3 \times 5^5$$

and its corresponding sequence of digits is given by

Since $|S_1(5)| = 2, B(5) = \{4\}$ *and* $N_5 = 0$, *these and the formula (7) imply that*

$$v_5(C_{11186}) = |S_1(5)| - N_5 + |T_2(5)| = 2 - 0 + 1 = 3$$

Remark 14. *Besides the formula (3), Deutsch and Sagan [5, Theorem 5.1] gave a complete characterization of the residue of* C_n *modulo 3 for a particular class of positive integers* n. *Thereafter people started to do research on the congruence class of* C_n *modulo prime powers. For instances, Eu, Liu and Yeh [7] determined the congruence of* C_n *modulo 8 in 2008, then in 2010, Liu and Yeh [13] classified* *the congruence of C_n modulo 64. Other recent developments or results on the congruence of C_n modulo 2^k or p^k can be found in [3], [12] and [19].*

Remark 15. *There are several monographs about Catalan numbers which were published in the last ten years. They are Koshy [9] in 2009, Roman [14] and Stanley [18] both in 2015.*

Remark 16. *Finally, we remark that Motzkin numbers $\{M_n\}$ form another important sequence of numbers which are closely related to Catalan numbers. In fact, it is well-known that*

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

for every positive integer n. It is natural to ask the following question: Does there exist a formula connecting $v_p(M_n)$ and the digits of n for odd primes p?

8. Conclusion

By and large, we propound a new method to establish a formula of the p-adic valuation of Catalan numbers and discover the connection which is previously unknown between this p-adic valuation and the coefficients of n in the base p-expansion. As we can derive the p-adic valuation of Catalan numbers merely by determining a few digits where Theorem 3 requires calculation of various sets of inequalities, this method is conclusively more natural and convenient than Theorem 3.

9. A brief history of Catalan numbers

Catalan numbers appear in many combinatorial and geometric problems. They have over 200 years of history. It was first discovered in 1730 by a Chinese scientist. During this period, many mathematicians derived Catalan number by different ways like polygons cutting and bracket sequences. In the early years, only Catalan numbers with small value can be computed but more numbers have been computed when more formulas of Catalan numbers were discovered by mathematicians.

9.1. The development of Catalan numbers

In 1730s, the Mongolian mathematician Ming Antu mentioned the expansion of $\sin(2\alpha)$, it was the first time to involve Catalan numbers:

$$\sin(2\alpha) = 2\sin\alpha - \sum_{n=1}^{\infty} \frac{C_{n-1}}{4^{n-1}} \sin^{2n+1}\alpha = 2\sin\alpha - \sin^3\alpha - \frac{1}{4}\sin^5\alpha - \dots$$

Figure 1 is a piece of the manuscript of Ming Antu and the Catalan numbers C_n for $n = 0, 1, 2, 3, 4, 5$ and 6 are boxed.

Later in 1751, Euler defined Catalan numbers as the number of triangulations of $(n+2)$-gon and he applied a binomial formula to derive the following product formula:

$$C_{n-2} = \frac{2 \cdot 6 \cdots (4n-10)}{2 \cdot 3 \cdots (n-1)} \tag{32}$$

In the 1750s, they can only compute the values of Catalan numbers when $n \leq 8$. However, in 1758, Segner worked on the number of triangulations of a n-gon and ultimately discovered the formula in Definition **??**. Meanwhile, the values of Catalan numbers for $n \leq 23$ were instantly computed by using that formula.

In 1795, Fuss introduced Fuss-Catalan numbers, it was the number of subdivisions of an n-gon into k-gons. Such numbers form a family of generalized Catalan numbers and was finished before Catalan. Furthermore, the sequence of Catalan numbers is a special case of the family of Fuss-Catalan numbers.

In 1838, Lame discovered another way to obtain formula (32). Firstly, he counted the number A_n of triangulations of a $(n+2)$-gon with one of its $(n-1)$ diagonals oriented and derived

$$A_n = 2(n-1)C_n$$

Afterwards, he summed over all possible directed diagonals and the result showed the following formula:

$$A_n = n(C_1 C_{n-1} + C_2 C_{n-2} + \cdots + C_n C_1)$$

Eventually, he obtained formula (32) by combining the two formulas of A_n with Segner's formula. It is crucial that his work inspired Catalan who obtained the current standard formula (32).

9.2. Problems involving Catalan numbers

In 1859, Cayley, who was interested in counting plane trees, proved that the number of plane trees is the C_n and then introduced another formula for C_n:

$$C_{m-1} = \frac{1 \cdot 3 \cdot 5 \cdots (2m-3)}{1 \cdot 2 \cdot 3 \cdots m} 2^{m-1}$$

In 1838, Rodrigues counted in two ways the number B_n of triangulations of $(n+2)$-gon where either an edge or a diagonal was oriented. Afterwards, he noticed that

$$B_n = 2(2n+1)C_n \quad \text{and} \quad B_n = (n+2)C_{n+1}$$

which demonstrated that they involved Catalan numbers.

In the same year, he investigated the bracket sequences and noticed that there was a relation between bracket sequences and C_n. Firstly, he denoted P_n as the number of bracket sequences of $x_1, x_2, \ldots, x_n$. After that, he discovered that

$$P_n = n!C_n \quad \text{and} \quad P(n+1) = (4n-2)P_n$$

as variable $x(n+1)$ can be inserted to the left of any of the $(2n-1)$ left brackets or to the right of any of the $(2n-1)$ right brackets, e.g.

$$((x_2)((x_1)(x_3))) \to ((x_2)(((x_1)(x_4))(x_3))$$

In 1878, Whitworth introduced ballot sequences. He resolved the problem which ballot numbers have not been computed yet and eventually realized that the numbers he computed were C_n.

9.3. The works of Catalan and naming

It has known that Catalan obtained the current standard formula (1.1). Moreover, in 1838, he defined ballot numbers, disguised as the number of certain triangulations and gave a formula for the ballot numbers in terms of C_n. In 1878, Catalan published a book on the divisibility of the Catalan numbers.

Despite he did many on investigating Catalan numbers, they were not called as Catalan numbers entirely as sometimes they were called the Segner numbers or the "Euler-Segner numbers". Nevertheless, in 1968, an American combinatorialist Riordan used the name Catalan numbers in his monograph and it struck a chord. It was enormously crucial since his book was very influential at that period. His monograph was the first book containing the name and it consequently spread after 1968. Not least, in 1976, Martin Gardner's Scientific American column popularized this name greatly, the name thus has become popular since that time.

10. Several applications of Catalan numbers

Apart from evaluating Catalan numbers by using some algebraic formulas, we can also discover them from geometric problems. In fact, C_n appears in many real-life problems. In this appendix, some applications of C_n will be presented. For further information, the reader is suggested to read the monographs [9] and [14].

10.1. $(n+2)$-gon triangulation

If we triangulate a $(n+2)$-gon and count the possible ways of it, the result demonstrates Catalan numbers. Figure **??** illustrates that the ways of triangulating are 1, 2, 5, 14 for the first four polygons respectively. It satisfies the result of using formula (1) to compute the Catalan numbers.

10.2. Balanced parentheses

Denote n as the number of pairs of parentheses and count the ways of valid combination where an open parenthesis must match a closed parenthesis. Table 1 shows the numbers of ways for the first 4 terms of n are 1, 2, 5, 14 which, again, are the sequence of the Catalan numbers.

10.3. Mountain ranges

Let n be the number of upstrokes and the number of downstrokes. We use those upstrokes and downstrokes to form mountains where all strokes must not stay below the original line. From Figure 3, we notice that the number of ways for n is exactly Catalan number. It helps us to derive the further answer when n becomes greater since we cannot sketch a hundred of strokes.

10.4. Counting Diagonal-avoiding Paths

Let's count the number of paths from $(0,0)$ to $(n-1,n-1)$ on a $(n-1)\times(n-1)$ plane without falling below the line: $y=x$. Figure 4 below shows 1, 2, 5, 14 which are Catalan numbers. Nonetheless, we can also apply combination to count the possible paths and the result will be the formula (1.1).

10.5. Staircase tessellations

If you consider the number of ways to form staircase shapes with n rectangle, you can also obtain Catalan numbers. Figure 5 illustrates that the they are the first four terms of the sequence of Catalan numbers.

REFERENCES

[1] R. Alter and T.B. Curtz, *On binary nonassociative products and their relation to a classical problem of Euler*, Roczniki Polskiego Towarzystwa Mathematics, 17: 1-8, 1972.
[2] R. Alter and K.K. Kubota, *Prime and Prime Power Divisibility of Catalan Numbers*, J. Combin. Theory Ser. A, 15: 243-256, 1973.
[3] Y.G. Chen and W. Jiang, *Catalan numbers modulo a prime power*, integers, 13: Article 36, 2013.
[4] E. Deutsch, *An involution on Dyck paths and its consequences*, Discrete Math., 204: 163-166, 1999.
[5] E. Deutsch and B.E. Sagan, *Congruences for Catalan and Motzkin numbers and related sequences*, J. Number Theory, 117: 191-215, 2006.
[6] Ö. Eğecioğlu, *The parity of the Catalan numbers via lattice paths*, Fibonacci Quart., 21: 65-66, 1983.
[7] S.P. Eu, S.C. Liu and Y.N. Yeh, *Catalan and Motzkin numbers modulo 4 and 8*, European J. Combin., 29: 1449-1466, 2008.
[8] R.P. Grimaldi, *Discrete and Combinatorial Mathematics*, an Applied Introduction, 3rd ed., Addison-Wesley, 1994.
[9] T. Koshy, *Catalan Numbers with Applications*, Oxford University Press, 2009.
[10] T. Koshy, *Discrete Mathematics with Applications*, Elsevier Academic Press, 2004.
[11] T. Koshy and M. Salmassi, *Parity and primality of Catalan numbers*, College Math. J., 37: 52-53, 2006.
[12] H.Y. Lin, *Odd Catalan numbers modulo* 2^k, Integers, 12: 161-165, 2012.

[13] S.C. Liu and J.C.C. Yeh, *Catalan numbers module* 2^k, J. Integer Seqeuences, 13 Article 10.5.4, 2010.
[14] S. Roman, *An Introduction to Catalan Numbers*, Birkhäuser, 2015.
[15] J. Riordan, *Combinatorial Identities*, Krieger, Huntington, 1979.
[16] F. S. Roberts, *Applied Combinatorics*, Prentice-Hall, Englewood Cliffs, N. J., 1984.
[17] R. Simion and D. Ullman, *On the structure of the lattice of noncrossing partitions*, Discrete Math., 98: 193-206, 1991.
[18] R.P. Stanley, *Catalan Numbers*, Cambridge University Press, 2015.
[19] G. Xin and J.F. Xu, *A short approach to Catalan numbers modulo* 2^r, Electron. J. Combin., 18: Paper 177, 2011.
[20] http://www.javascripter.net/math/calculators/catalannumberscalculator.htm
[21] https://www.alpertron.com.ar/ECM.HTM

Reviewer's Comments

This article studied the p-adic valuation of the Catalan number. Here are the reviewer's comments on the paper:

1. Novelty and methodology: The paper improved a result given by Alter and Kubota [1]. It also mentioned some applications of the claimed results with a brief explanation of the history of Catalan number. Here are some of the reviewer's suggestions:
   - On page 30, it would be nice if the author can explain more why the claimed results are "more natural" than those found in the literature. And why are the old methods more "artificial"?
   - In the reviewer's opinion, Theorem 4 is quite similar to Theorem B as proved by Alter and Kubota in [1], while Theorem 5 seems to be a more important result given in the presented work (at least the reviewer can see more applications from Theorem 5). It would be more reasonable if the authors can pinpoint the difference between Theorem 4 and Theorem 5, and show the significance of Theorem 4 in a more explicit way.
2. Organisation: This paper is well-written, and the ideas and proofs are clear. It appears that the paper is nearly a publishable journal article! Here are some of the reviewer's suggestions:
   - In the proof of Theorem 5, the authors used two "facts", yet they are not really "facts" (the authors gave the proofs on them) and it may cause confusions to the readers.
   - On the final line of the proof of Theorem 5, rather than simply saying "Hence, combining Fact 1 and Fact 2, we are able to show that the formula (7) holds and this completes the proof of the theorem.", it would be better if the authors can explain more on the applications of Fact 1 and Fact 2 (in proving Theorem 5).

REFERENCES

[1] R. Alter and K.K. Kubota, Prime and Prime Power Divisibility of Catalan Numbers, J. Combin. Theory Ser. A, Vol. 15, 1973, pp. 243–256.