IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

On the p-adic valuation of Stirling numbers of the first kind

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/1622122 since 2017-01-16T17:48:33Z

Published version:

DOI:10.1007/s10474-016-0680-4

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

12 October 2024

ON THE p-ADIC VALUATION OF STIRLING NUMBERS OF THE FIRST KIND

PAOLO LEONETTI

Università "Luigi Bocconi", Department of Statistics, Milan, Italy
email: `leonetti.paolo@gmail.com`

CARLO SANNA

Università degli Studi di Torino, Department of Mathematics, Turin, Italy
email: `carlo.sanna.dev@gmail.com`

Abstract. For all integers $n \geq k \geq 1$, define $H(n,k) := \sum 1/(i_1 \cdots i_k)$, where the sum is extended over all positive integers $i_1 < \cdots < i_k \leq n$. These quantities are closely related to the Stirling numbers of the first kind by the identity $H(n,k) = s(n+1,k+1)/n!$. Motivated by the works of Erdős–Niven and Chen–Tang, we study the p-adic valuation of $H(n,k)$. Lengyel proved that $\nu_p(H(n,k)) > -k \log_p n + O_k(1)$ and we conjecture that there exists a positive constant $c = c(p,k)$ such that $\nu_P(H(n,k)) < -c \log n$ for all large n. In this respect, we prove the conjecture in the affirmative for all $n \leq x$ whose base p representations start with the base p representation of $k-1$, but at most $3x^{0.835}$ exceptions. We also generalize a result of Lengyel by giving a description of $\nu_2(H(n,2))$ in terms of an infinite binary sequence.

1. Introduction

It is well known that the n-th harmonic number $H_n := 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ is not an integer whenever $n \geq 2$. Indeed, this result has been generalized in several ways (see, e.g., [2, 7, 13]). In particular, given integers $n \geq k \geq 1$, Erdős and Niven [8] proved that

$$H(n,k) := \sum_{1 \leq i_1 < \cdots < i_k \leq n} \frac{1}{i_1 \cdots i_k}$$

is an integer only for finitely many n and k. Precisely, Chen and Tang [4] showed that $H(1,1)$ and $H(3,2)$ are the only integral values. (See also [11] for a generalization to arithmetic progressions.)

A crucial step in both the proofs of Erdős–Niven and Chen–Tang's results consists in showing that, when n and k are in an appropriate range, for some prime number p the p-adic valuation of $H(n,k)$ is negative, so that $H(n,k)$ cannot be an integer.

2010 *Mathematics Subject Classification.* Primary: 11B73; Secondary: 11B50, 11A51.
Key words and phrases. Stirling number of the first kind, harmonic numbers, p-adic valuation.

Moreover, a study of the p-adic valuation of the harmonic numbers was initiated by Eswarathasan and Levine [9]. They conjectured that for any prime number p the set $\mathcal{J}_p$ of all positive integers n such that $\nu_p(H_n) > 0$ is finite. Although Boyd [3] gave a probabilistic model predicting that $\#\mathcal{J}_p = O(p^2(\log\log p)^{2+\varepsilon})$, for any $\varepsilon > 0$, and Sanna [22] proved that $\mathcal{J}_p$ has asymptotic density zero, the conjecture is still open. Another result of Sanna [22] is that $\nu_p(H_n) = -\lfloor \log_p n \rfloor$ for any n in a subset $\mathcal{S}_p$ of the positive integers with logarithmic density greater than 0.273.

In this paper, we study the p-adic valuation of $H(n,k)$. Let $s(n,k)$ denotes an unsigned Stirling number of the first kind [10, §6.1], i.e., $s(n,k)$ is the number of permutations of $\{1,\dots,n\}$ with exactly k disjoint cycles. Then $H(n,k)$ and $s(n,k)$ are related by the following easy identity.

Lemma 1.1. *For all integers $n \geq k \geq 1$, we have $H(n,k) = s(n+1,k+1)/n!$.*

In light of Lemma 1.1, and since the p-adic valuation of the factorial is given by the formula [10, p. 517, 4.24]

$$\nu_p(n!) = \frac{n - s_p(n)}{p-1},$$

where $s_p(n)$ is the sum of digits of the base p representation of n, it follows that

$$\nu_p(H(n,k)) = \nu_p(s(n+1,k+1)) - \frac{n - s_p(n)}{p-1}, \tag{1}$$

hence the study of $\nu_p(H(n,k))$ is equivalent to the study of $\nu_p(s(n+1,k+1))$. That explains the title of this paper.

In this regard, p-adic valuations of sequences with combinatorial meanings have been studied by several authors (see, e.g., [5, 15, 17, 19, 20, 21, 23]). In particular, the p-adic valuation of Stirling numbers of the second kind have been extensively studied [1, 6, 12, 14, 16]. On the other hand, very few seems to be known about the p-adic valuation of Stirling numbers of the first kind. Indeed, up to our knowledge, the only systematic work on this topic is due to Lengyel [18]. Among several results, he showed (see the proof of [18, Theorem 1.2]) that, for all primes p and positive integers k, it holds

$$\nu_p(H(n,k)) > -k\log_p n + O_k(1). \tag{2}$$

The main aim of this article is to provide an upper bound for $\nu_p(H(n,k))$. In this respect, we believe that inequality (2) is *nearly* optimal, and our Theorem 2.3 confirms this in the special case when the base p representation of n starts with the base p representation of $k-1$.

We restrict ourselves to this special case since the proofs are already quite involved. However, we think that our method could be improved to remove this condition on the base p representation of n. Probably, the first step in that direction would be finding a general p-adic expansion of $H(n,k)$ with coefficients depending only of the digits of the base p representation of n, extending Lemma 3.2 in the following.

Lastly, we also formulate the following:

Conjecture 1.1. *For any prime number p and any integer $k \geq 1$, there exists a constant $c = c(p,k) > 0$ such that $\nu_p(H(n,k)) < -c\log n$ for all sufficiently large integers n.*

2. Main results

Before stating our results, we need to introduce some notation and definition. For any prime number p, we write

$$\langle a_0, \dots, a_v \rangle_p := \sum_{i=0}^{v} a_i p^{v-i}, \text{ where } a_0, \dots, a_v \in \{0, \dots, p-1\}, \ a_0 \neq 0, \tag{3}$$

to denote a base p representation. In particular, hereafter, the restrictions of (3) on $a_0, \dots, a_v$ will be implicitly assumed any time we will write something like $\langle a_0, \dots, a_v \rangle_p$.

For any positive integer $a = \langle a_0, \dots, a_v \rangle_p$, let $\mathcal{S}_p(a)$ be the set of all positive integers whose base p representations start with the base p representation of a, that is,

$$\mathcal{S}_p(a) := \big\{ \langle b_0, \dots, b_u \rangle_p : u \geq v \text{ and } b_i = a_i \text{ for } i = 0, \dots, v \big\}.$$

We call *p-tree of root $a = \langle a_0, \dots, a_v \rangle_p$* a set of positive integers $\mathcal{T}$ such that:

(T1) $\langle a_0, \dots, a_v \rangle_p \in \mathcal{T}$;
(T2) If $\langle b_0, \dots, b_u \rangle_p \in \mathcal{T}$ then $u \geq v$ and $b_i = a_i$ for $i = 0, \dots, v$;
(T3) If $\langle b_0, \dots, b_u \rangle_p \in \mathcal{T}$ and $u > v$ then $\langle b_0, \dots, b_{u-1} \rangle_p \in \mathcal{T}$.

Hence, it is clear that $\mathcal{T} \subseteq \mathcal{S}_p(a)$. Moreover, for any $n = \langle d_0, \dots, d_s \rangle_p \in \mathcal{S}_p(a) \setminus \mathcal{T}$ we denote by $\mu_p(\mathcal{T}, n)$ the least positive integer r such that $\langle d_0, \dots, d_r \rangle_p \notin \mathcal{T}$. Note that $\mu_p(\mathcal{T}, n)$ is indeed well defined and that obviously $\mu_p(\mathcal{T}, n) \leq s$. Finally, the *girth* of $\mathcal{T}$ is the least integer g such that for all $\langle b_0, \dots, b_u \rangle_p \in \mathcal{T}$ we have $\langle b_0, \dots, b_u, c \rangle_p \in \mathcal{T}$ for at most g values of $c \in \{0, \dots, p-1\}$.

We are ready to state our results about the p-adic valuation of $H(n, k)$.

Theorem 2.1. *Let p be a prime number and let $k \geq 2$ be an integer. Then there exist a p-tree $\mathcal{T}_p(k)$ of root $k-1$ and a nonnegative integer $W_p(k)$ such that for all integers $n = \langle d_0, \dots, d_s \rangle_p \in \mathcal{S}_p(k-1)$ we have:*

(i) *If $n \notin \mathcal{T}_p(k)$ then $\nu_p(H(n,k)) = W_p(k) + \mu_p(\mathcal{T}_p(k), n) - ks$;*
(ii) *If $n \in \mathcal{T}_p(k)$ then $\nu_p(H(n,k)) > W_p(k) - (k-1)s$.*

Moreover, the girth of $\mathcal{T}_p(k)$ is less than $p^{0.835}$. In particular, $\mathcal{T}_2(k)$ is infinite and its girth is equal to 1.

Note that the case $k = 1$ has been excluded from the statement. (As mentioned in the introduction, see [3, 9, 22] for results on the p-adic valuation of $H(n, 1) = H_n$.)

For given p and k, the proof of Theorem 2.1 gives a full description of the p-tree $\mathcal{T}_p(k)$ and of the nonnegative integer $W_p(k)$. Moreover, in Section 5 we explain a method to effectively compute the elements of $\mathcal{T}_p(k)$. Therefore, Theorem 2.1(i) gives an effective formula for $\nu_p(H(n,k))$ for any $n \in \mathcal{S}_p(k-1) \setminus \mathcal{T}_p(k)$. Note also that the bound on the girth of $\mathcal{T}_p(k)$ implies that $\mathcal{S}_p(k-1) \setminus \mathcal{T}_p(k)$ has infinitely many elements. Furthermore, for some p and k we have that $\mathcal{T}_p(k)$ is finite (see Section 5), hence in such cases computing $\nu_p(H(n,k))$ for the finitely many $n \in \mathcal{T}_p(k)$ and using Theorem 2.1(i) for $n \in \mathcal{S}_p(k-1) \setminus \mathcal{T}_p(k)$, we obtain a complete description of $\nu_p(H(n,k))$ for all $n \in \mathcal{S}_p(k-1)$.

Since the statement of Theorem 2.1 is a bit complicated, for the sake of clarity we give a numerical example: Take $p = 3$ and $k = 2$. Then $\mathcal{T}_p(k)$ is the finite set of 8 integers drawn in Figure 1, while $W_p(k) = 0$. If we choose $n = 1257 = \langle 1, 2, 0, 1, 1, 2, 0 \rangle_3$, then it follows easily that $n \in \mathcal{S}_p(k-1) \setminus \mathcal{T}_p(k)$ and $\mu_p(\mathcal{T}_p(k), n) = 3$ thus Theorem 2.1 gives $\nu_p(H(n,k)) = 0 + 3 - 2 \cdot 6 = -9$.

Lengyel [18, Theorem 2.5] proved that for each integer $m \geq 2$ it holds

$$\nu_2(s(2^m, 3)) = 2^m - 3m + 3$$

which, in light of identity (1), is in turn equivalent to

$$\nu_2(H(2^m - 1, 2)) = 4 - 2m. \tag{4}$$

As an application of Theorem 2.1, we give a corollary that generalizes (4) and provides a quite precise description of $\nu_2(H(n, 2))$.

Corollary 2.2. *There exists a sequence $f_0, f_1, \ldots \in \{0, 1\}$ such that for any integer $n = \langle d_0, \ldots, d_s \rangle_2 \geq 2$ we have:*

(i) *If $d_0 = f_0, \ldots, d_{r-1} = f_{r-1}$, and $d_r \neq f_r$, for some positive integer $r \leq s$, then $\nu_2(H(n, 2)) = r - 2s$;*

(ii) *If $d_0 = f_0, \ldots, d_s = f_s$, then $\nu_2(H(n, 2)) > -s$.*

Precisely, the sequence $f_0, f_1, \ldots$ can be computed recursively by $f_0 = 1$ and

$$f_s = \begin{cases} 1 & \text{if } \nu_2(H(\langle f_0, \ldots, f_{s-1}, 1 \rangle_2, 2)) > -s, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

for any positive integer s. In particular, $f_0 = 1$, $f_1 = 1$, $f_2 = 0$.

Note that (4) is indeed a consequence of Corollary 2.2. In fact, on the one hand, for $m = 2$ the identity (4) can be checked quickly. On the other hand, for any integer $m \geq 3$ we have $2^m - 1 = \langle d_0, \ldots, d_{m-1} \rangle_2$ with $d_0 = \cdots = d_{m-1} = 1$, so that $d_0 = f_0$, $d_1 = f_1$, and $d_2 \neq f_2$, hence (4) follows from Corollary 2.2(i), with $s = m - 1$ and $r = 2$.

Finally, we obtain the following upper bound for $\nu_p(H(n, k))$.

Theorem 2.3. *Fix a prime number p, and integer $k \geq 2$, and $x \geq (k-1)p$. Then*

$$\nu_p(H(n, k)) < -(k-1)(\log_p n - \log_p(k-1) - 1)$$

holds for all $n \in \mathcal{S}_p(k-1) \cap [(k-1)p, x]$, but at most $3x^{0.835}$ exceptions.

Note that $\#(\mathcal{S}_p(k-1) \cap [(k-1)p, x]) \gg_{p,k} x$. Hence Theorem 2.3 gives an upper bound for $\nu_p(H(n, k))$ for almost all $n \in \mathcal{S}_p(k-1)$, with respect to the its asymptotic relative density. In particular, there exists a positive constant $c = c(p, k)$ such that

$$\nu_p(H(n, k)) < -c \log(n)$$

for almost all $n \in \mathcal{S}_p(k-1)$, which provides, in turn, a sort of evidence in support of Conjecture 1.1.

3. Preliminaries

Let us start by proving the identity claimed in Lemma 1.1.

Proof of Lemma 1.1. By [10, Eq. 6.11] and $s(n+1, 0) = 0$, we have the polynomial identity

$$\prod_{i=1}^{n} (X + i) = \sum_{k=0}^{n} s(n+1, k+1) X^k,$$

hence

$$1+\sum_{k=1}^{n} H(n,k)X^k = \prod_{i=1}^{n}\left(\frac{X}{i}+1\right) = \frac{1}{n!}\prod_{i=1}^{n}(X+i) = \sum_{k=0}^{n}\frac{s(n+1,k+1)}{n!}X^k$$

and the claim follows. □

From here later, let us fix a prime number p and let $k = \langle e_0,\dots,e_t\rangle_p + 1 \geq 2$ and $n = \langle d_0,\dots,d_s\rangle_p$ be positive integers with $s \geq t+1$ and $d_i = e_i$ for $i = 0,\dots,t$. For any $a_0,\dots,a_v \in \{0,\dots,p-1\}$, define

$$B_p(a_0,\dots,a_v) := \langle a_0,\dots,a_v\rangle_p - \langle a_0,\dots,a_{v-1}\rangle_p,$$

where by convention $\langle a_0,\dots,a_{v-1}\rangle_p = 0$ if $v = 0$, and also

$$\mathcal{B}_p(a_0,\dots,a_v) := \big\{c_p(i) : i = 1,\dots,B_p(a_0,\dots,a_v)\big\}$$

where $c_p(1) < c_p(2) < \cdots$ denotes the sequence of all positive integers not divisible by p. Lastly, put

$$\mathcal{A}_p(n,v) := \big\{m \in \{1,\dots,n\} : \nu_p(m) = s-v\big\},$$

for each integer $v \geq 0$. The next lemma relates $\mathcal{A}_p(n,v)$ and $\mathcal{B}_p(d_0,\dots,d_v)$.

Lemma 3.1. *For each nonnegative integer $v \leq s$, we have*

$$\mathcal{A}_p(n,v) = \big\{jp^{s-v} : j \in \mathcal{B}_p(d_0,\dots,d_v)\big\}.$$

In particular, $\#\mathcal{A}_p(n,v) = B_p(d_0,\dots,d_v)$ and $\mathcal{A}_p(n,v)$ depends only on $p,s,d_0,\dots,d_v$.

Proof. For $m \in \{1,\dots,n\}$, we have $m \in \mathcal{A}_p(n,v)$ if and only if $p^{s-v} \mid n$ but $p^{s-v+1} \nmid n$. Therefore,

$$\begin{aligned}\#\mathcal{A}_p(n,v) &= \left\lfloor\frac{n}{p^{s-v}}\right\rfloor - \left\lfloor\frac{n}{p^{s-v+1}}\right\rfloor = \left\lfloor\sum_{i=0}^{s} d_i p^{v-i}\right\rfloor - \left\lfloor\sum_{i=0}^{s} d_i p^{v-i-1}\right\rfloor \\ &= \sum_{i=0}^{v} d_i p^{v-i} - \sum_{i=0}^{v-1} d_i p^{v-i-1} = \langle d_0,\dots,d_v\rangle_p - \langle d_0,\dots,d_{v-1}\rangle_p \\ &= B_p(d_0,\dots,d_v),\end{aligned}$$

and

$$\begin{aligned}\mathcal{A}_p(n,v) &= \big\{c_p(i)p^{s-v} : i = 1,\dots,\#\mathcal{A}_p(n,v)\big\} \\ &= \big\{c_p(i)p^{s-v} : i = 1,\dots,B_p(d_0,\dots,d_v)\big\} \\ &= \big\{jp^{s-v} : j \in \mathcal{B}_p(d_0,\dots,d_v)\big\},\end{aligned}$$

as claimed. □

Before stating the next lemma, we need to introduce some additional notation. First, we define

$$\mathcal{C}_p(n,k) := \bigcup_{v=0}^{t}\mathcal{A}_p(n,v) \quad\text{and}\quad \Pi_p(k) := \prod_{j\in\mathcal{C}_p(n,k)}\frac{1}{\mathrm{free}_p(j)},$$

where $\mathrm{free}_p(m) := m/p^{\nu_p(m)}$ for any positive integer m. Note that, since $d_i = e_i$ for $i = 0, \dots, t$, from Lemma 3.1 it follows easily that indeed $\Pi_p(k)$ depends only on p and k, and not on n. Then we put

$$U_p(k) := \sum_{v=0}^{t} B_p(e_0, \dots, e_v)v + t + 1,$$

while, for $a_0, \dots, a_{t+v+1} \in \{0, \dots, p-1\}$, with $v \geq 0$ and $a_i = e_i$ for $i = 0, \dots, t$, we set

$$H'_p(a_0, \dots, a_{t+v}) := \sum_{\substack{0 \leq v_1, \dots, v_k \leq t+v \\ v_1 + \cdots + v_k = U_p(k) + v}} \sum_{\substack{j_1/p^{v_1} < \cdots < j_k/p^{v_k} \\ j_1 \in \mathcal{B}_p(a_0, \dots, a_{v_1}), \dots, j_k \in \mathcal{B}_p(a_0, \dots, a_{v_k})}} \frac{1}{j_1 \cdots j_k}$$

and

$$H_p(a_0, \dots, a_{t+v+1}) := H'_p(a_0, \dots, a_{t+v}) + \Pi_p(k) \sum_{j \in \mathcal{B}_p(a_0, \dots, a_{t+v+1})} \frac{1}{j}.$$

Note that $\nu_p(H_p(a_0, \dots, a_{t+v+1})) \geq 0$, this fact will be fundamental later.

The following lemma gives a kind of p-adic expansion for $H(n,k)$. We use $O(p^v)$ to denote a rational number with p-adic valuation greater than or equal to v.

Lemma 3.2. *We have*

$$H(n,k) = \sum_{v=0}^{s-t-1} H_p(d_0, \dots, d_{t+v+1}) \cdot p^{v-ks+U_p(k)} + O\big(p^{s-t-ks+U_p(k)}\big).$$

Proof. Clearly, we can write

$$H(n,k) = \sum_{v=0}^{V_p(n,k)} J_p(n,k,v) \cdot p^{v-V_p(n,k)},$$

where

$$V_p(n,k) := \max\big\{\nu_p(i_1 \cdots i_k) : 1 \leq i_1 < \cdots < i_k \leq n\big\},$$

and

$$J_p(n,k,v) := \sum_{\substack{1 \leq i_1 < \cdots < i_k \leq n \\ \nu_p(i_1 \cdots i_k) = V_p(n,k) - v}} \frac{1}{\mathrm{free}_p(i_1 \cdots i_k)},$$

for each nonnegative integer $v \leq V_p(n,k)$.

We shall prove that $V_p(n,k) = ks - U_p(k)$. On the one hand, we have

$$\begin{aligned} \sum_{v=0}^{t} B_p(e_0, \dots, e_v) &= \sum_{v=0}^{t} \big(\langle e_0, \dots, e_v \rangle_p - \langle e_0, \dots, e_{v-1} \rangle_p\big) \\ &= \langle e_0, \dots, e_t \rangle_p = k - 1. \end{aligned} \tag{6}$$

On the other hand, by (6) and thanks to Lemma 3.1, we obtain

$$\#\mathcal{C}_p(n,k) = \sum_{v=0}^{t} \#\mathcal{A}_p(n,v) = \sum_{v=0}^{t} B_p(e_0, \dots, e_v) = k - 1. \tag{7}$$

Hence, in order to maximize $\nu_p(i_1 \cdots i_k)$ for positive integers $i_1 < \cdots < i_k \le n$, we have to choose $i_1, \dots, i_k$ by picking all the $k-1$ elements of $\mathcal{C}_p(n,k)$ and exactly one element from $\mathcal{A}_p(n, t+1)$. Therefore, using again (6) and Lemma 3.1, we get

$$\begin{aligned} V_p(n,k) &= \sum_{v=0}^{t} \#\mathcal{A}_p(n,v)(s-v) + (s-t-1) \\ &= \sum_{v=0}^{t} B_p(e_0, \dots, e_v)(s-v) + (s-t-1) \\ &= \left(\sum_{v=0}^{t} B_p(e_0, \dots, e_v) + 1 \right) s - U_p(k) \\ &= ks - U_p(k), \end{aligned} \tag{8}$$

as desired.

Similarly, if $\nu_p(i_1 \cdots i_k) = V_p(n,k) - v$, for some positive integers $i_1 < \cdots < i_k \le n$ and some nonnegative integer $v \le s-t-1$, then only two cases are possible: $\nu_p(i_1), \dots, \nu_p(i_k) \ge s-t-v$; or $i_1, \dots, i_k$ consist of all the $k-1$ elements of $\mathcal{C}_p(n,k)$ and one element of $\mathcal{A}_p(n, t+v+1)$. As a consequence,

$$J_p(n,k,v) = \sum_{\substack{1 \le i_1 < \cdots < i_k \le n \\ \nu_p(i_1 \cdots i_k) = V_p(n,k) - v \\ \nu_p(i_1), \dots, \nu_p(i_k) \ge s-t-v}} \frac{1}{\mathrm{free}_p(i_1 \cdots i_k)} + \Pi_p(k) \sum_{i \in \mathcal{A}_p(n, t+v+1)} \frac{1}{\mathrm{free}_p(i)}, \tag{9}$$

for all nonnegative integers $v \le s-t-1$.

By putting $v_\ell := s - \nu_p(i_\ell)$ and $j_\ell := \mathrm{free}_p(i_\ell)$ for $\ell = 1, \dots, k$, the first sum of (9) can be rewritten as

$$\begin{aligned} &\sum_{\substack{0 \le v_1, \dots, v_k \le t+v \\ (s-v_1) + \cdots + (s-v_k) = V_p(n,k) - v}} \sum_{\substack{i_1 < \cdots < i_k \\ i_1 \in \mathcal{A}_p(n, v_1), \dots, i_k \in \mathcal{A}_p(n, v_k)}} \frac{1}{\mathrm{free}_p(i_1 \cdots i_k)} \\ &= \sum_{\substack{0 \le v_1, \dots, v_k \le t+v \\ v_1 + \cdots + v_k = U_p(k) + v}} \sum_{\substack{j_1/p^{v_1} < \cdots < j_k/p^{v_k} \\ j_1 \in \mathcal{B}_p(d_0, \dots, d_{v_1}), \dots, j_k \in \mathcal{B}_p(d_0, \dots, d_{v_k})}} \frac{1}{j_1 \cdots j_k} = H'_p(d_0, \dots, d_{t+v}), \end{aligned}$$

where we have also made use of (8) and Lemma 3.1, hence

$$J_p(n,k,v) = H_p(d_0, \dots, d_{t+v+1}), \tag{10}$$

for any nonnegative integer $v \le s-t-1$.

At this point, being $s > t$, by (8) it follows that $V_p(n,k) > s-t-1$, hence

$$H(n,k) = \sum_{v=0}^{s-t-1} J_p(n,k,v) \cdot p^{v - ks + U_p(k)} + O\big(p^{s-t-ks+U_p(k)}\big), \tag{11}$$

since clearly $\nu_p(J_p(n,k,v)) \ge 0$ for any nonnegative integer $v \le V_p(n,k)$.

In conclusion, the claim follows from (10) and (11). □

Finally, we need two lemmas about the number of solutions of some congruences. For rational numbers a and b, we write $a \equiv b \bmod p$ to mean that $\nu_p(a-b) > 0$.

Lemma 3.3. *Let r be a rational number and let x, y be positive integers with $y < p$. Then the number of integers $v \in [x, x+y]$ such that $H_v \equiv r \bmod p$ is less than $\frac{3}{2}y^{2/3}+1$.*

Proof. The case $r = 0$ is proved in [22, Lemma 2.2] and the proof works exactly in the same way even for $r \neq 0$. □

Lemma 3.4. *Let q be a rational number and let a be a positive integer. Then the number of $d \in \{0, \dots, p-1\}$ such that*

$$\sum_{i=a}^{a+d} \frac{1}{c_p(i)} \equiv q \bmod p \tag{12}$$

is less than $p^{0.835}$.

Proof. It is easy to see that there exists some $h \in \{0, \dots, p-2\}$ such that

$$c_p(i) = \begin{cases} c_p(a) + i - a & \text{for } i = a, \dots, a+h, \\ c_p(a) + i - a + 1 & \text{for } i = a+h+1, \dots, a+p-1. \end{cases}$$

Therefore, by putting $x := c_p(a)$, $y := h$, and $r := q + H_{x-1}$ in Lemma 3.3, we get that the number of $d \leq h$ satisfying (12) is less than $\frac{3}{2}h^{2/3} + 1$. Similarly, by putting $x := c_p(a) + h + 2$, $y := p - h - 2$, and

$$r := q + H_{x-1} - \sum_{i=a}^{a+h} \frac{1}{c_p(i)}$$

in Lemma 3.3, we get that the number of $d \in [h+1, p-1]$ satisfying (12) is less than $\frac{3}{2}(p-h-2)^{2/3} + 1$. Thus, letting N be the number of $d \in \{0, \dots, p-1\}$ that satisfy (12), we have

$$N \leq \frac{3}{2}h^{2/3} + 1 + \frac{3}{2}(p-h-2)^{2/3} + 1 \leq 3\left(\frac{p-2}{2}\right)^{2/3} + 2.$$

Furthermore, it is clear the d and $d+1$ cannot both satisfy (12), hence $N \leq \lceil p/2 \rceil$. Finally, a little computation shows that the maximum of

$$\log_p\left(\min\left(3\left(\frac{p-2}{2}\right)^{2/3} + 2, \left\lceil \frac{p}{2} \right\rceil\right)\right)$$

is obtained for $p = 59$ and is less than 0.835, hence the claim follows. □

4. Proof of Theorem 2.1

Now we are ready to prove Theorem 2.1. For any $a_0, \dots, a_{t+u+1} \in \{0, \dots, p-1\}$, with $u \geq 0$ and $a_i = e_i$ for $i = 0, \dots, t$, let

$$\Sigma_p(a_0, \dots, a_{t+u+1}) := \sum_{v=0}^{u} H_p(a_0, \dots, a_{t+v+1}) \cdot p^v.$$

Furthermore, define the sequence of sets $\mathcal{T}_p^{(0)}(k), \mathcal{T}_p^{(1)}(k), \dots$ as follows: $\mathcal{T}_p^{(0)}(k) := \{\langle e_0, \dots, e_t \rangle_p\}$, and for any integer $u \geq 0$ put $\langle a_0, \dots, a_{t+u+1} \rangle_p \in \mathcal{T}_p^{(u+1)}(k)$ if and only

if $\langle a_0,\dots,a_{t+u}\rangle_p \in \mathcal{T}_p^{(u)}(k)$ and $\nu_p(\Sigma_p(a_0,\dots,a_{t+u+1})) \geq u+1$. At this point, setting

$$\mathcal{T}_p(k) := \bigcup_{u=0}^{\infty} \mathcal{T}_p^{(u)}(k),$$

it is straightforward to see that $\mathcal{T}_p(k)$ is a p-tree of root $\langle e_0,\dots,e_t\rangle_p$. Put $W_p(k) := U_p(k)-t-1$.

If $n \notin \mathcal{T}_p(k)$ then, for the sake of convenience, set $r := \mu_p(\mathcal{T}_p(k),n)$. Thus $r > t$, $\langle d_0,\dots,d_{r-1}\rangle_p \in \mathcal{T}_p(k)$ but $\langle d_0,\dots,d_r\rangle \notin \mathcal{T}_p(k)$, so that

$$\nu_p(\Sigma_p(d_0,\dots,d_r)) \leq r-t-1. \tag{13}$$

Now we distinguish between two cases. If $r = t+1$, then $\nu_p(\Sigma_p(d_0,\dots,d_{t+1})) = 0$ and by Lemma 3.2 we obtain $\nu_p(H(n,k)) = W_p(k)+r-ks$. If $r > t+1$ then by $\langle d_0,\dots,d_{r-1}\rangle \in \mathcal{T}_p(k)$ we get that $\nu_p(\Sigma_p(d_0,\dots,d_{r-1})) \geq r-t-1$, which together with (13) and

$$\Sigma_p(d_0,\dots,d_r) = \Sigma_p(d_0,\dots,d_{r-1}) + H_p(d_0,\dots,d_r)\cdot p^{r-t-1}$$

implies that $\nu_p(\Sigma_p(d_0,\dots,d_r)) = r-t-1$, hence by Lemma 3.2 we get $\nu_p(H(n,k)) = W_p(k)+r-ks$, and (i) is proved.

If $n \in \mathcal{T}_p(k)$ then, by the definition of $\mathcal{T}_p(k)$, we have $\nu_p(\Sigma_p(d_0,\dots,d_s)) \geq s-t$. Therefore, by Lemma 3.2 it follows that $\nu_p(H(n,k)) > W_p(k)-(k-1)s$, and this proves (ii).

It remains only to bound the girth of $\mathcal{T}_p(k)$. Let u be a nonnegative integer and pick $\langle a_0,\dots,a_{t+u}\rangle_p \in \mathcal{T}_p^{(u)}(k)$. By the definition of $\mathcal{T}_p^{(u+1)}(k)$, we have $\langle a_0,\dots,a_{t+u+1}\rangle_p \in \mathcal{T}_p^{(u+1)}(k)$ if and only if $\nu_p(\Sigma_p(a_0,\dots,a_{t+u+1})) \geq u+1$, which in turn is equivalent to

$$\begin{aligned}&\sum_{v=0}^{u-1} H_p(a_0,\dots,a_{t+v+1})\cdot p^{v-u} + H_p'(a_0,\dots,a_{t+u}) + \Pi_p(k)\sum_{j\in\mathcal{B}_p(a_0,\dots,a_{t+u+1})}\frac{1}{j} \\ &\equiv \sum_{v=0}^{u} H_p(a_0,\dots,a_{t+v+1})\cdot p^{v-u} \equiv 0 \bmod p.\end{aligned} \tag{14}$$

Using the definition of $\mathcal{B}_p(a_0,\dots,a_{t+u+1})$ and the facts that

$$B_p(a_0,\dots,a_{t+u+1}) = a_{t+u+1} + (p-1)\sum_{v=0}^{u} a_v p^{u-v},$$

and $\nu_p(\Pi_p(k)) = 0$, we get that (14) is equivalent to

$$\begin{aligned}\sum_{i=a}^{a+a_{t+u+1}} \frac{1}{c_p(i)} &\equiv -\sum_{i=1}^{a-1}\frac{1}{c_p(i)} \\ &- \frac{1}{\Pi_p(k)}\left(\sum_{v=0}^{u-1} H_p(a_0,\dots,a_{t+v+1})\cdot p^{v-u} + H_p'(a_0,\dots,a_{t+u})\right) \bmod p,\end{aligned} \tag{15}$$

where

$$a := (p-1)\sum_{v=0}^{u} a_v p^{u-v}.$$

Note that both a and the right-hand side of (15) do not depend on a_{t+u+1}. As a consequence, by Lemma 3.4 we get that $\langle a_0, \dots, a_{t+u+1}\rangle_p \in \mathcal{T}_p^{(u+1)}(k)$ for less than $p^{0.835}$ values of $a_{t+u+1} \in \{0, \dots, p-1\}$. Thus the girth of $\mathcal{T}_p(k)$ is less than $p^{0.835}$.

Finally, consider the case $p = 2$. Obviously, $1/c_2(i) \equiv 1 \bmod 2$ for any positive integer i, while the right-hand side of (15) is equal to 0 or 1 (mod 2). Therefore, there exists one and only one choice of $a_{t+u+1} \in \{0, 1\}$ such that (15) is satisfied. This means that $\mathcal{T}_2(k)$ is infinite and its girth is equal to 1.

The proof is complete.

5. The computation of $\mathcal{T}_p(k)$

Given p and k, it might be interesting to effectively compute the elements of $\mathcal{T}_p(k)$. Clearly, $\mathcal{T}_p(k)$ could be infinite — by Theorem 2.1 this is indeed the case when $p = 2$ — hence the computation should proceed by first enumerating all the elements of $\mathcal{T}_p^{(0)}(k)$, then all the elements of $\mathcal{T}_p^{(1)}(k)$, and so on. An obvious way to do this is using the recursive definition of the $\mathcal{T}_p^{(u)}(k)$'s. However, it is easy to see how this method is quite complicated and impractical. A better idea is noting that from Theorem 2.1 we have

$$\begin{aligned}\mathcal{T}_p^{(u+1)}(k) = \big\{\langle a_0, \dots, a_{t+u}, b\rangle_p : \langle a_0, \dots, a_{t+u}\rangle_p \in \mathcal{T}_p^{(u)}(k), \\ \nu_p(H(\langle a_0, \dots, a_{t+u}, b\rangle_p, k)) > W_p(k) - (k-1)s\big\},\end{aligned} \tag{16}$$

for all integers $u \geq 0$. Therefore, starting from $\mathcal{T}_p^{(0)}(k) = \{\langle e_0, \dots, e_t\rangle_p\}$, formula (16) gives a way to compute recursively all the elements of $\mathcal{T}_p(k)$. In particular, if $\mathcal{T}_p(k)$ is finite, then after sufficient computation one will get $\mathcal{T}_p^{(u)}(k) = \varnothing$ for some positive integer u, so the method actually proves that $\mathcal{T}_p(k)$ is finite.

The authors implemented this algorithm in SageMath, since it allows computations with arbitrary-precision p-adic numbers. In particular, they found that $\mathcal{T}_3(2), \dots, \mathcal{T}_3(6)$ are all finite sets, with respectively $8, 24, 16, 7, 23$ elements, while the cardinality of $\mathcal{T}_3(7)$ is at least 43. Through these numerical experiments, it seems that, in general, $\mathcal{T}_p(k)$ does not exhibit any trivial structure (see Figures 1, 2, 3), hence the question of the finiteness of $\mathcal{T}_p(k)$ is probably a difficult one.

6. Proof of Corollary 2.2

Only for this section, let us focus on the case $p = 2$ and $k = 2$, so that $t = 0$, $e_0 = 1$, and $W_2(2) = 0$. Thanks to Theorem 2.1 we know that $\mathcal{T}_2(2)$ is infinite and its girth is equal to 1. Hence, it follows easily that there exists a sequence $f_0, f_1, \dots \in \{0, 1\}$ such that $\mathcal{T}_2^{(u)}(2) = \{\langle f_0, \dots, f_u\rangle_2\}$ for all integers $u \geq 0$. In particular, $f_0 = e_0 = 1$. At this point, (i) and (ii) are direct consequences of Theorem 2.1, while the recursive formula (5) is just a special case of (16).

7. Proof of Theorem 2.3

On the one hand, if $n = \langle d_0, \ldots, d_s \rangle_p \in (\mathcal{S}_p(k-1) \setminus \mathcal{T}_p(k)) \cap [(k-1)p, x]$ then by Theorem 2.1 we get that

$$\begin{aligned}
\nu_p(H(n,k)) &= W_p(k) + \mu_p(\mathcal{T}_p(k), n) - ks \\
&\leq W_p(k) - (k-1)s \\
&= \sum_{v=0}^{t} B_p(e_0, \ldots, e_v)v - (k-1)s \\
&\leq \sum_{v=0}^{t} B_p(e_0, \ldots, e_v)t - (k-1)s < -(k-1)(\log_p n - \log_p(k-1) - 1),
\end{aligned}$$

where we have made use of (6) and the inequalities $\mu_p(\mathcal{T}_p(k), n) \leq s$, $s > \log_p n - 1$, and $t \leq \log_p(k-1)$.

On the other hand, by Theorem 2.1, the girth of $\mathcal{T}_p(k)$ is less than $p^{0.835}$, hence it follows easily that $\#\mathcal{T}_p^{(u)}(k) < p^{0.835u}$, for any positive integer u. As a consequence,

$$\#(\mathcal{T}_p(k) \cap [(k-1)p, x]) \leq \sum_{u=1}^{\lfloor \log_p x \rfloor - t} \#\mathcal{T}_p^{(u)}(k) < \sum_{u=1}^{\lfloor \log_p x \rfloor} p^{0.835u} < 3x^{0.835},$$

and the claim follows.

8. Figures

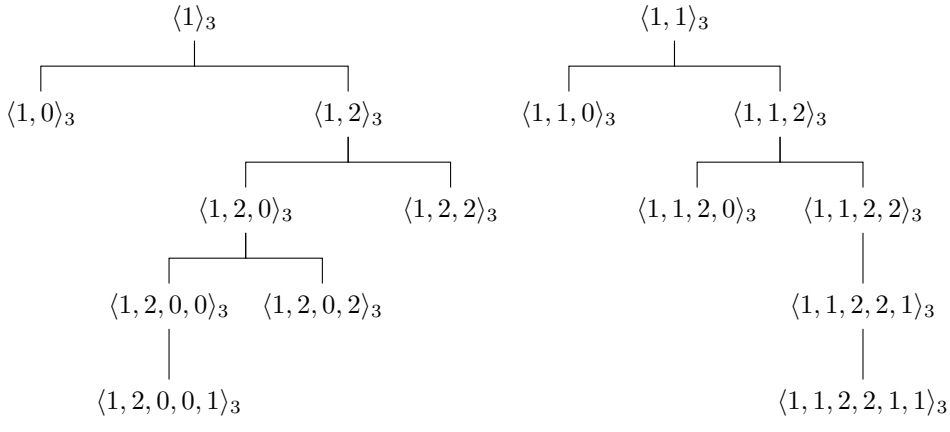

Figure 1. The 8 elements of $\mathcal{T}_3(2)$ (left tree), and the 7 elements of $\mathcal{T}_3(5)$ (right tree).

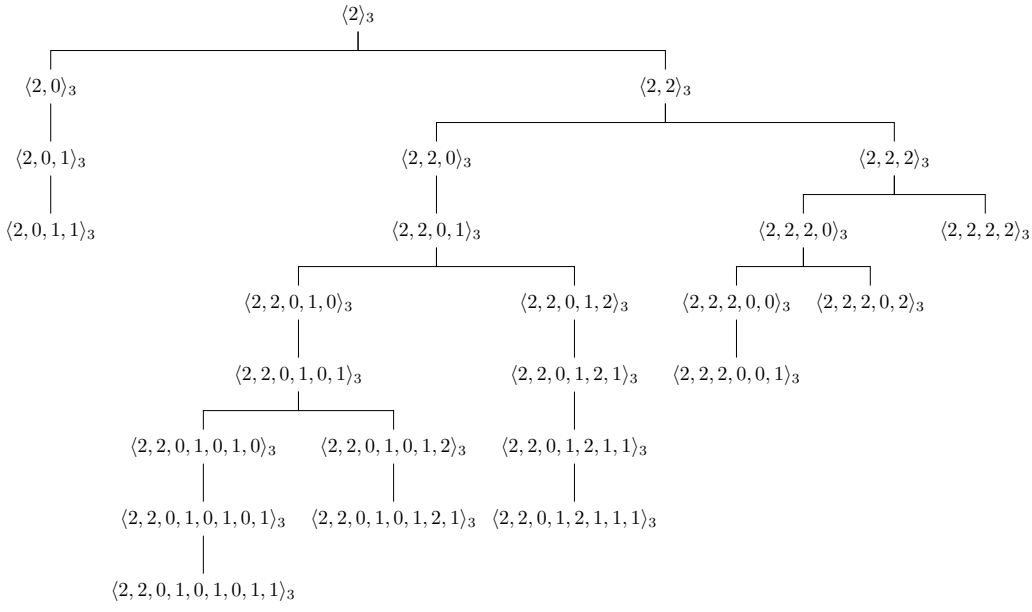

FIGURE 2. The 24 elements of $\mathcal{T}_3(3)$.

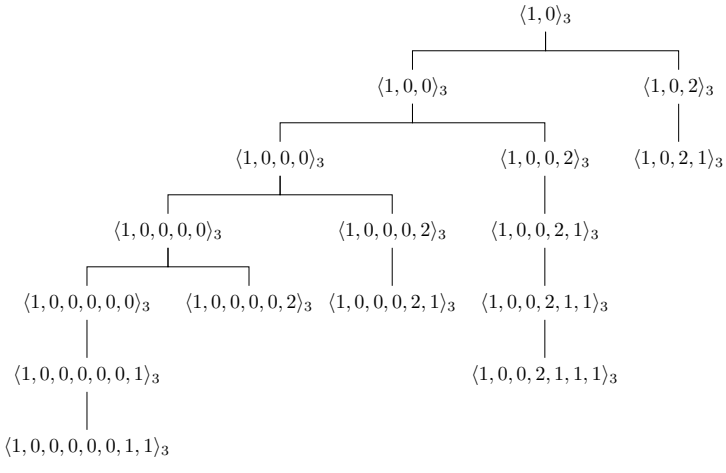

FIGURE 3. The 16 elements of $\mathcal{T}_3(4)$.

REFERENCES

[1] T. Amdeberhan, D. Manna, and V. H. Moll, *The 2-adic valuation of Stirling numbers*, Experiment. Math. **17** (2008), no. 1, 69–82.

[2] H. Belbachir and A. Khelladi, *On a sum involving powers of reciprocals of an arithmetical progression*, Ann. Math. Inform. **34** (2007), 29–31.

[3] D. W. Boyd, *A p-adic study of the partial sums of the harmonic series*, Experiment. Math. **3** (1994), no. 4, 287–302.

[4] Y.-G. Chen and M. Tang, *On the elementary symmetric functions of* $1, 1/2, \ldots, 1/n$, Amer. Math. Monthly **119** (2012), no. 10, 862–867.

[5] H. Cohn, 2-*adic behavior of numbers of domino tilings*, Electron. J. Combin. **6** (1999), Research Paper 14, 7 pp. (electronic).

[6] S. De Wannemacker, *On 2-adic orders of Stirling numbers of the second kind*, Integers **5** (2005), no. 1, A21, 7.

[7] P. Erdős, *Verallgemeinerung eines elementar-zahlentheoretischen Satzes von Kürschák.*, Mat. Fiz. Lapok **39** (1932), 17–24 (Hungarian).

[8] P. Erdős and I. Niven, *Some properties of partial sums of the harmonic series*, Bull. Amer. Math. Soc. **52** (1946), 248–251.

[9] A. Eswarathasan and E. Levine, *p-integral harmonic sums*, Discrete Math. **91** (1991), no. 3, 249–257.

[10] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete mathematics*, second ed., Addison-Wesley Publishing Company, Reading, MA, 1994, A foundation for computer science.

[11] S. Hong and C. Wang, *The elementary symmetric functions of reciprocals of elements of arithmetic progressions*, Acta Math. Hungar. **144** (2014), no. 1, 196–211.

[12] S. Hong, J. Zhao, and W. Zhao, *The 2-adic valuations of Stirling numbers of the second kind*, Int. J. Number Theory **8** (2012), no. 4, 1057–1066.

[13] J. Kürschák, *Über die harmonische Reihe.*, Mat. Fiz. Lapok **27** (1918), 299–300 (Hungarian).

[14] T. Lengyel, *On the divisibility by* 2 *of the Stirling numbers of the second kind*, Fibonacci Quart. **32** (1994), no. 3, 194–201.

[15] T. Lengyel, *The order of the Fibonacci and Lucas numbers*, Fibonacci Quart. **33** (1995), no. 3, 234–239.

[16] T. Lengyel, *On the 2-adic order of Stirling numbers of the second kind and their differences*, 21st International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2009), Discrete Math. Theor. Comput. Sci. Proc., AK, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2009, pp. 561–572.

[17] T. Lengyel, *Exact p-adic orders for differences of Motzkin numbers*, Int. J. Number Theory **10** (2014), no. 3, 653–667.

[18] T. Lengyel, *On p-adic properties of the Stirling numbers of the first kind*, J. Number Theory **148** (2015), 73–94.

[19] D. Marques and T. Lengyel, *The 2-adic order of the Tribonacci numbers and the equation* $T_n = m!$, J. Integer Seq. **17** (2014), no. 10, Article 14.10.1, 8.

[20] N. Murru and C. Sanna, *On the k-regularity of the k-adic valuation of Lucas sequences*, J. Théor. Nombres Bordeaux (to appear).

[21] A. Postnikov and B. E. Sagan, *What power of two divides a weighted Catalan number?*, J. Combin. Theory Ser. A **114** (2007), no. 5, 970–977.

[22] C. Sanna, *On the p-adic valuation of harmonic numbers*, J. Number Theory **166** (2016), 41–46.

[23] C. Sanna, *The p-adic valuation of Lucas sequences*, Fibonacci Quart. **54** (2016), 118–224.