On the p-adic valuation of harmonic numbers

ON THE p-ADIC VALUATION OF HARMONIC NUMBERS

CARLO SANNA

ABSTRACT. For any prime number p, let J_p be the set of positive integers n such that p divides the numerator of the n-th harmonic number H_n. An old conjecture of Eswarathasan and Levine states that J_p is finite. We prove that for $x \geq 1$ the number of integers in $J_p \cap [1, x]$ is less than $129p^{2/3}x^{0.765}$. In particular, J_p has asymptotic density zero. Furthermore, we show that there exists a subset S_p of the positive integers, with logarithmic density greater than 0.273, and such that for any $n \in S_p$ the p-adic valuation of H_n is equal to $-\lfloor \log_p n \rfloor$.

1. Introduction

For each positive integer n, let

$$H_n := 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$$

be the n-th harmonic number. The arithmetic properties of harmonic numbers have been studied since a long time. For example, Wolstenholme [7] proved in 1862 that for any prime number $p \geq 5$ the numerator of H_{p-1} is divisible by p^2; while in 1915, Taeisinger [6, p. 3115] showed that H_n is never an integer for $n > 1$.

For each prime number p, let J_p be the set of positive integers n such that the numerator of H_n is divisible by p. Eswarathasan and Levine [4] conjectured that J_p is finite for all primes p, and provided a method to compute the elements of J_p. If J_p is finite, then, after sufficient computation, their method gives a proof that it is finite. They computed $J_2 = \varnothing$, $J_3 = \{2, 7, 22\}$, $J_5 = \{4, 20, 24\}$, and

$$J_7 = \{6, 42, 48, 295, 299, 337, 341, 2096, 2390, 14675, 16731, 16735, 102728\}.$$

Boyd [2], using some p-adic expansions, improved the algorithm of Eswarathasan and Levine, and determined J_p for all primes $p \leq 547$, but 83, 127, and 397; confirming that J_p is finite for those prime numbers. Notably, he showed that J_{11} has 638 elements, the largest being an integer of 31 digits. Boyd gave also an heuristic model predicting that J_p is always finite and that its cardinality is $\#J_p = O\big(p^2(\log\log p)^{2+\varepsilon}\big)$. However, the conjecture of Eswarathasan and Levine is still open.

We write $J_p(x) := J_p \cap [1, x]$, for $x \geq 1$. Our first result is the following.

Theorem 1.1. *For any prime number p and any $x \geq 1$, it holds*

$$\#J_p(x) < 129p^{2/3}x^{0.765}.$$

In particular, J_p has asymptotic density zero.

For any prime number p, let $\nu_p(\cdot)$ be the usual p-adic valuation over the rational numbers. Boyd [2, Proposition 3.3] proved the following lemma.

Lemma 1.2. *For any prime p, the set J_p is finite if and only if $\nu_p(H_n) \to -\infty$, as $n \to +\infty$.*

Therefore, the study of J_p is strictly related to the negative growth of the p-adic valuation of H_n. It is well-known and easy to prove that $\nu_2(H_n) = -\lfloor \log_2 n \rfloor$. (Hereafter, $\lfloor x \rfloor$ denotes the greatest integer not exceeding the real number x.) Moreover, Kamano [5, Theorem 2] proved

2010 *Mathematics Subject Classification.* Primary: 11B99, 11B83. Secondary: 11B05.

Key words and phrases. harmonic numbers, p-adic valuation, asymptotic density, logarithmic density.

that $\nu_3(H_n)$ can be determined easily from the expansion of n in base 3. Note that, since obviously $\nu_p(k) \leq \lfloor \log_p n \rfloor$ for any $k \in \{1, \ldots, n\}$, we have the lower bound

$$\nu_p(H_n) \geq -\lfloor \log_p n \rfloor. \tag{1}$$

Our next result shows that in (1) the equality holds quite often. We recall that the logarithmic density of a set of positive integers S is defined as

$$\delta(S) := \lim_{x \to +\infty} \frac{1}{\log x} \sum_{n \,\in\, S \,\cap\, [1,x]} \frac{1}{n},$$

whenever this limit exists.

Theorem 1.3. *For any prime number p, there exists a set S_p of positive integers, with logarithmic density $\delta(S_p) > 0.273$, and such that $\nu_p(H_n) = -\lfloor \log_p n \rfloor$ for each $n \in S_p$.*

2. Proof of Theorem 1.1

For any prime p, define the sequence of sets $J_p^{(1)}, J_p^{(2)}, \ldots$ as follow:

$$J_p^{(1)} := \{n \in \{1, \ldots, p-1\} : p \mid H_n\},$$
$$J_p^{(k+1)} := \{pn + r : n \in J_p^{(k)},\ r \in \{0, \ldots, p-1\},\ p \mid H_{pn+r}\} \quad \forall k \geq 1.$$

First, we need the following lemma.

Lemma 2.1. *For all prime numbers p, it holds $J_p^{(k)} = J_p \cap [p^{k-1}, p^k[$, for each integer $k \geq 1$. In particular, $J_p = \bigcup_{k=1}^\infty J_p^{(k)}$.*

Proof. From [4, Eq. 2.5] we know that if n is a positive integer and $r \in \{0, \ldots, p-1\}$, then $pn + r \in J_p$ implies that $n \in J_p$. Therefore, the claim follows quickly by induction on k. □

Now we prove a result regarding the number of elements of J_p in a short interval.

Lemma 2.2. *For any prime p, and any real numbers x and y, with $1 \leq y < p$, we have*

$$\#\big(J_p \cap [x, x+y]\big) < \frac{3y^{2/3}}{2} + 1.$$

Proof. Set $c := \#\big(J_p \cap [x, x+y]\big)$. If $c \leq 1$, then there is nothing to prove. Hence, suppose $c \geq 2$ and let $n_1 < \cdots < n_c$ be the elements of $J_p \cap [x, x+y]$. Moreover, define $d_i := n_{i+1} - n_i$, for any $i = 1, \ldots, c-1$. Given a positive integer d, consider the polynomial

$$f_d(X) := (X+1)(X+2)\cdots(X+d). \tag{2}$$

Taking the logarithms of both sides of (2) and deriving, we obtain the identity

$$\frac{f_d'(X)}{f_d(X)} = \frac{1}{X+1} + \frac{1}{X+2} + \cdots + \frac{1}{X+d}.$$

Thus for any $i = 1, \ldots, c-1$ we have

$$\frac{f_{d_i}'(n_i)}{f_{d_i}(n_i)} = \frac{1}{n_i+1} + \frac{1}{n_i+2} + \cdots + \frac{1}{n_{i+1}} = H_{n_{i+1}} - H_{n_i} \equiv 0 \bmod p,$$

so that $f_{d_i}'(n_i) \equiv 0 \bmod p$. Since $f_d'(X)$ is a non-zero polynomial of degree $d-1$, there are at most $d-1$ solutions modulo p of the equation $f_d'(X) \equiv 0 \bmod p$. Therefore, for any $z \geq 1$, on the one hand we have

$$\#\{i : d_i \leq z\} = \sum_{1 \,\leq\, d \,\leq\, z} \#\{i : d_i = d\} \leq \sum_{1 \,\leq\, d \,\leq\, z} (d-1) < \frac{z^2}{2}. \tag{3}$$

On the other hand,

$$\#\{i : d_i > z\} < \frac{1}{z} \sum_{i=1}^{c-1} d_i = \frac{n_c - n_1}{z} \leq \frac{y}{z} \tag{4}$$

In conclusion, by summing (3) and (4), we get

$$c - 1 = \#\{i : d_i \le z\} + \#\{i : d_i > z\} < \frac{z^2}{2} + \frac{y}{z}$$

and the claim follows taking $z = y^{1/3}$. □

We are ready to prove Theorem 1.1. If $p < 83$ then, from the values of $\#J_p$ computed by Boyd [2, Table 2], one can check that $\#J_p/p^{2/3} < 129$, so the claim is obvious. Hence, suppose $p \ge 83$, and put $A := \frac{3}{2}(p-1)^{2/3} + 1$. By the definition of the sets $J_p^{(k)}$, and since Lemma 2.2, we get that

$$\#J_p^{(1)} = \#\big(J_p \cap [1, p-1]\big) < A,$$

while

$$\#J_p^{(k+1)} = \sum_{n \in J_p^{(k)}} \#\big(J_p \cap [pn, pn + p - 1]\big) < \#J_p^{(k)} \cdot A,$$

hence it follows by induction that $\#J_p^{(k)} < A^k$.

Now let s be the positive integer determined by $p^{s-1} \le x < p^s$. Note that $p^s \notin J_p$, indeed $\nu_p(H_{p^s}) = -s$ (this is a particular case of Lemma 3.2 in the next section). Thanks to Lemma 2.1 and the previous considerations, we have

$$\begin{aligned}\#J_p(x) \le \#J_p(p^s) = \#J_p(p^s - 1) &= \sum_{k=1}^{s} \#\big(J_p \cap [p^{k-1}, p^k[\big) = \sum_{k=1}^{s} \#J_p^{(k)} \\ &< \sum_{k=1}^{s} A^k < \frac{A^2}{A-1} \cdot A^{s-1} = \frac{A^2}{A-1} \cdot \big(p^{s-1}\big)^{\log_p A} < \frac{A^2}{A-1} \cdot x^{0.765} < 129 p^{2/3} x^{0.765},\end{aligned}$$

since $p^{s-1} \le x$, while it can be checked quickly that $\log_p A < 0.765$. The proof is complete.

3. Proof of Theorem 1.3

For any integer $b \ge 2$ and any $d \in \{1, \dots, b-1\}$, let $F_b(d)$ be the set of positive integers that have the most significant digit of their base b expansion equal to d. The set $F_b(d)$ has not an asymptotic density, however $F_b(d)$ has a logarithmic density. In fact, $F_b(d)$ satisfies a kind of Benford's law [1], as shown by the following lemma.

Lemma 3.1. *For all integers $b \ge 2$ and $d \in \{1, \dots, b-1\}$, it holds $\delta(F_b(d)) = \log_b\big(1 + 1/d\big)$.*

Proof. See [3]. □

Write $J_p^* := \{1, \dots, p-1\} \setminus J_p^{(1)}$.

Lemma 3.2. *For p prime, $d \in J_p^*$, and $n \in F_p(d)$, it holds $\nu_p(H_n) = -\lfloor \log_p n \rfloor$.*

Proof. Since $n \in F_p(d)$, we can write $n = p^k d + r$, where $k := \lfloor \log_p n \rfloor$ and $r < p^k$ is a non-negative integer. Hence,

$$H_n = \sum_{\substack{m=1 \\ p^k \nmid m}}^{n} \frac{1}{m} + \sum_{j=1}^{d} \frac{1}{p^k j} = \sum_{\substack{m=1 \\ p^k \nmid m}}^{n} \frac{1}{m} + \frac{H_d}{p^k}. \tag{5}$$

On the one hand, it is clear that the last sum in (5) has p-adic valuation greater than $-k$. On the other hand, we have $\nu_p(H_d/p^k) = -k$, since $d \in J_p^*$ and so $p \nmid H_d$.

In conclusion, $\nu_p(H_n) = -k$ as desired. □

Now we can prove Theorem 1.3. Define the set S_p as

$$S_p := \bigcup_{d \in J_p^*} F_p(d).$$

It follows immediately from Lemma 3.2 that $\nu_p(H_n) = -\lfloor \log_p n \rfloor$, for each $n \in S_p$. Moreover, since the sets $F_p(d)$ are disjoint, and thanks to Lemma 3.1, we have

$$\delta(S_p) := \sum_{d \in J_p^*} \delta(F_p(d)) = \sum_{d \in J_p^*} \log_p\left(1 + \frac{1}{d}\right) \geq \sum_{d = \#J_p^{(1)}+1}^{p-1} \log_p\left(1 + \frac{1}{d}\right) \tag{6}$$

$$= \log_p\left(\frac{p}{\#J_p^{(1)} + 1}\right) = 1 - \frac{\log\left(\#J_p^{(1)} + 1\right)}{\log p}.$$

Suppose $p \geq 1013$. By Lemma 2.2 we have

$$\#J_p^{(1)} = \#(J_p \cap [1, p-1]) < \tfrac{3}{2}(p-2)^{2/3} + 1,$$

hence from (6) we get

$$\delta(S_p) > 1 - \frac{\log\left(\frac{3}{2}(p-2)^{2/3} + 2\right)}{\log p} > 0.273.$$

At this point, the proof is only a matter of computation. The author used the Python programming language (since it has native support for arbitrary-sized integers) to computed the numerators of the harmonic numbers H_n, up to $n = 1012$. Then he determined $\#J_p^{(1)}$ for each prime number $p < 1013$, and using (6) he checked that the inequality $\delta(S_p) > 0.273$ holds. This required only a few seconds on a personal computer.

4. Concluding remarks

From the proof of Theorem 1.1, it is clear that with our methods one cannot obtain an upper bound better than $\#J_p(x) < Cp^{2/3}x^{2/3+\varepsilon}$, for some $C, \varepsilon > 0$. Similarly, in the statement of Theorem 1.3 a logarithmic density greater than $1/3 - \varepsilon$ cannot be achieved.

One way to obtain better results could be an improvement of Lemma 2.2, we leave this as an open question for the readers.

References

1. A. Berger and T. P. Hill, *An introduction to Benford's law*, Princeton University Press, 2015.
2. D. W. Boyd, *A p-adic study of the partial sums of the harmonic series*, Experiment. Math. **3** (1994), no. 4, 287–302.
3. R. L. Duncan, *Note on the initial digit problem*, Fibonacci Quart. **7** (1969), no. 5, 474–475.
4. A. Eswarathasan and E. Levine, *p-integral harmonic sums*, Discrete Math. **91** (1991), no. 3, 249–257.
5. K. Kamano, *On 3-adic valuations of generalized harmonic numbers*, Integers **11** (2011).
6. E. W. Weisstein, *Concise encyclopedia of mathematics*, 2 ed., CRC Press, 2002.
7. J. Wolstenholme, *On certain properties of prime numbers*, Quart. J. Pure Appl. Math. **5** (1862), 35–39.

Università degli Studi di Torino, Department of Mathematics, Turin, Italy
E-mail address: carlo.sanna.dev@gmail.com