THE $p$-ADIC GROWTH OF HARMONIC SUMS

KEITH CONRAD

The numbers

$$H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n} = \sum_{k=1}^{n} \frac{1}{k}$$

are called harmonic sums. Integrals can approximate these sums as real numbers, e.g., $H_n$ and $\int_1^n \frac{dt}{t} = \log n$ differ by a bounded amount ($H_n - \int_1^n \frac{dt}{t}$ is positive and monotonically decreasing), so the $H_n$’s slowly diverge in $\mathbb{R}$. We are interested here in the arithmetic behavior of harmonic sums, i.e., their $p$-adic growth. This is definitely not an important topic, but it is interesting to compare what happens with this rational sequence across the real and different $p$-adic landscapes.

Here are some initial values:

<table>
<thead>
<tr>
<th>$n$</th>
<th>$H_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3/2</td>
</tr>
<tr>
<td>3</td>
<td>11/6</td>
</tr>
<tr>
<td>4</td>
<td>25/12</td>
</tr>
<tr>
<td>5</td>
<td>137/60</td>
</tr>
<tr>
<td>6</td>
<td>49/20</td>
</tr>
<tr>
<td>7</td>
<td>363/140</td>
</tr>
<tr>
<td>8</td>
<td>761/280</td>
</tr>
<tr>
<td>9</td>
<td>7129/2520</td>
</tr>
<tr>
<td>10</td>
<td>7381/2520</td>
</tr>
<tr>
<td>11</td>
<td>83711/27720</td>
</tr>
</tbody>
</table>

Table 1. Harmonic sums

For $n > 1$, it appears that $H_n \notin \mathbb{Z}$. More precisely, the numerator of $H_n$ is odd and the denominator of $H_n$ is even. This suggests a strategy for proving $H_n$ is not an integer.

**Theorem 1** (Theisinger). For all $n \geq 2$, $H_n \notin \mathbb{Z}$.

**Proof.** Let $L$ be the least common multiple of $1, 2, \ldots, n$, so $H_n$ can be written as a fraction with denominator $L$. For $1 \leq k \leq n$, write $L = k a_k$ with $a_k \in \mathbb{Z}^+$, so $1/k = a_k/L$. Then

$$H_n = \sum_{k=1}^{n} \frac{1}{k} = \frac{\sum_{k=1}^{n} a_k}{L}.$$  

Since $n \geq 2$, $L$ is even. We will show $\sum_{k=1}^{n} a_k$ is odd, so the ratio is not integral.

Set $2^r$ to be the largest power of 2 up to $n$: $2^r \leq n < 2^{r+1}$. The only integer up to $n$ divisible by $2^r$ is $2^r$ itself, since $2 \cdot 2^r > n$. Therefore $L = 2^r b$ where $b$ is odd, so $2^r b = k a_k$ for $1 \leq k \leq n$. When $k = 2^r$ we see that $a_k = b$ is odd. When $k \neq 2^r$, $k$ is not divisible by $2^r$, so $a_k$ must be even. Therefore in the numerator $\sum_{k=1}^{n} a_k$, one term (at $k = 2^r$) is odd and the rest are even, so the total sum is odd.

$\square$
Theisinger [4] did not prove Theorem 1 using anything $p$-adic: he used determinants and Bertrand’s postulate, which says there is always a prime between $n$ and $2n$ when $n \geq 2$.\footnote{Theisinger referred to Bertrand’s postulate as Chebyshev’s theorem, since Chebyshev had proved it.}

The proof above for Theorem 1 presents it as a 2-adic result: $H_n \notin \mathbb{Z}$ because $H_n \notin \mathbb{Z}_2$. Here is a second proof of Theorem 1 brings out 2-adic features more clearly.

**Proof.** Let $2^r \leq n < 2^{r+1}$, so $r \geq 1$ and the highest power of 2 that appears in some reciprocal in the sum defining $H_n$ is $2^r$. The only reciprocal in $H_n$ with denominator divisible by $2^r$ is $1/2^r$. Indeed, every other such reciprocal would be $1/(2^r c)$ for odd $c > 1$, but if that term is in the sum then so is $1/(2^r \cdot 2) = 1/2^{r+1}$, which is false since $2^{r+1} > n$. Therefore $1/2^r$ has a more highly negative 2-adic valuation than every other term in $H_n$, so it is not cancelled out in the sum. This means $H_n \notin \mathbb{Z}_2$, so $H_n \notin \mathbb{Z}$. \hfill $\square$

This proof gives a formula for the 2-adic valuation of the harmonic sums: $\text{ord}_2(H_n) = -r$ where $2^r \leq n < 2^{r+1}$. To see this formula in action, we rewrite the initial harmonic sums in Table 2 with the power of 2 in the denominators made explicit. The exponent jumps when $n$ is a power of 2.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$H_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3/2</td>
</tr>
<tr>
<td>3</td>
<td>11/2 \cdot 3</td>
</tr>
<tr>
<td>4</td>
<td>25/2^2 \cdot 3</td>
</tr>
<tr>
<td>5</td>
<td>137/2^2 \cdot 15</td>
</tr>
<tr>
<td>6</td>
<td>49/2^2 \cdot 5</td>
</tr>
<tr>
<td>7</td>
<td>363/2^2 \cdot 35</td>
</tr>
<tr>
<td>8</td>
<td>761/2^3 \cdot 35</td>
</tr>
<tr>
<td>9</td>
<td>7129/2^3 \cdot 315</td>
</tr>
<tr>
<td>10</td>
<td>7381/2^3 \cdot 2520</td>
</tr>
<tr>
<td>11</td>
<td>83711/2^3 \cdot 3465</td>
</tr>
</tbody>
</table>

**Table 2.** Harmonic sums viewed 2-adically

Since $|H_n|_2 = 2^r$ if $2^r \leq n < 2^{r+1}$, $|H_n|_2 \rightarrow \infty$ as $n \rightarrow \infty$ (explicitly, $n/2 < |H_n|_2 \leq n$). Does $|H_n|_p \rightarrow \infty$ as $n \rightarrow \infty$ for odd primes $p$? We will address this later.

**Theorem 2** (Kürschák). *For $m \leq n - 2$, $H_n - H_m \notin \mathbb{Z}_2$. In particular, $H_n - H_m \notin \mathbb{Z}$.*

Taking $m = 1$ and $n \geq 3$ recovers Theorem 1 for $n \geq 3$ since $H_1 \in \mathbb{Z}$. Theorem 2 is false if $m = n - 1$ and $n$ is odd, since then $H_n - H_m = 1/n \in \mathbb{Z}_2$.

**Proof.** Writing

$$H_n - H_m = \sum_{k=m+1}^{n} \frac{1}{k},$$

we will show there is a unique term in the sum with the most negative 2-adic valuation. Let $r = \max_{m < k \leq n} \text{ord}_2(k)$. Since $n \geq m + 2$, the sum $H_n - H_m$ has at least two terms in it, so some $k$ is even and therefore $r \geq 1$.

We will show there is only one integer from $m$ up to $n-1$ with 2-adic valuation $r$. Suppose there are two such numbers. Write them as $2^c d$ and $2^d c$ with odd $c < d$. Then $c+1$ is even and $2^c c < 2^d(c+1) < 2^d d$, so $1/(2^d(c+1))$ appears in $H_n - H_m$. But $\text{ord}_2(2^d(c+1)) \geq r + 1$ since $c$ is odd. This contradicts the definition of $r$. Therefore there is only one term in $H_n - H_m$ with 2-adic valuation $-r$, so $\text{ord}_2(H_n - H_m) = -r$. \hfill $\square$
This type of 2-adic argument (in the setting of rational numbers, not 2-adic numbers) is due to Kirschčak [3], so he should be credited with the 2-adic proof of Theorem 1.

We now turn to the $p$-adic behavior of $H_n$ for $p \neq 2$. The material below is based on the notes “Harmonics and Primes” by Nicholas Rogers.

The harmonic sums $H_1, H_2, \ldots, H_{p-1}$ are all $p$-adically integral, since the denominators appearing in them are prime to $p$ and thus are $p$-adic integers (even $p$-adic units). Since $1/p$ is outside the ring $\mathbb{Z}_p$, $H_p = H_{p-1} + 1/p$ is outside $\mathbb{Z}_p$: $\operatorname{ord}_p(H_p) = -1$. It need not be true (as it is for $p = 2$) that $H_n \not\in \mathbb{Z}_p$ for all $n \geq p$. For example, $H_3, H_4$ and $H_5$ are outside $\mathbb{Z}_3$ (all with 3-adic valuation $-1$), but $H_6, H_7$, and $H_8$ are in $\mathbb{Z}_3$. Then $H_9, H_{10}, \ldots, H_{20}$ are outside of $\mathbb{Z}_3$. The harmonic sums return to $\mathbb{Z}_3$ for $H_{21}, H_{22}$, and $H_{23}$, then leave $\mathbb{Z}_3$ and come back in again for $H_{66}, H_{67}$, and $H_{68}$. Then the harmonic sums leave $\mathbb{Z}_3$ and never return. In fact, $|H_n|_3 \rightarrow \infty$ as $n \rightarrow \infty$. We will not discuss the proof of this, but the next theorem gives a simple connection between $p$-adic integrality and $p$-adic divergence of the harmonic sums.

**Theorem 3.** For a prime $p$, the following conditions are equivalent:

1. $\{n \geq 1 : H_n \in \mathbb{Z}_p\}$ is finite,
2. $|H_n|_p \rightarrow \infty$ as $n \rightarrow \infty$.

**Proof.** Trivially the second condition implies the first. To prove the (more interesting) reverse implication, pick $n \geq 1$ and write $n = pq + r$ with $0 \leq r \leq p - 1$. Since $1/k \in \mathbb{Z}_p$ when $(p,k) = 1$, in the quotient group $\mathbb{Q}_p/\mathbb{Z}_p$ a harmonic sum is equal to the sum of its terms whose denominators are divisible by $p$, so

$$H_n \equiv \frac{1}{p} + \frac{1}{2p} + \cdots + \frac{1}{pq} = \frac{1}{p}H_q,$$

so $H_n - (1/p)H_q \in \mathbb{Z}_p$.

We are assuming that $|H_n|_p \leq 1$ only finitely many times, so there is some $N_0 \geq 1$ such that $|H_n|_p > 1$ for $n \geq N_0$. We will show this $N_0$ controls the $p$-adic divergence of the harmonic sums: for all $k \geq 0$,

$$n \geq p^kN_0 \implies |H_n|_p > p^k.$$

For $k = 0$ this is true by the definition of $N_0$. Assuming it is true for $k$, suppose $n \geq p^{k+1}N_0$. Writing $n = pq + r$ with $0 \leq r \leq p - 1$, from $pq + r \geq p^{k+1}N_0$ we must have $q \geq p^kN_0$, so $|H_q|_p > p^k$. Therefore $|(1/p)H_q|_p > p^{k+1}$. Since $H_n - (1/p)H_q \in \mathbb{Z}_p$,

$$\left|H_n - \frac{1}{p}H_q\right|_p \leq 1 < p^{k+1} = \left|\frac{1}{p}H_q\right|_p,$$

so by the nonarchimedean property $|H_n|_p = |(1/p)H_q|_p > p^{k+1}$. \hfill \Box

The same reasoning shows $|H_n|_p \rightarrow \infty$ as $n \rightarrow \infty$ if and only if the set $\{n \geq 1 : H_n \in p\mathbb{Z}_p\}$ is finite. (Replace the strict inequalities $|H_n|_p > 1$ and $|H_n|_p > p^k$ in the proof with $|H_n|_p \geq 1$ and $|H_n|_p \geq p^k$.) This set, introduced in [2], is denoted

$$J(p) = \{n \geq 1 : H_n \in p\mathbb{Z}_p\}.$$

To say $n \not\in J(p)$ means $H_n$ has no $p$ in its numerator. For instance, Theorem 1 says $J(2) = \emptyset$. In [2] it is proved that $J(3) = \{2, 7, 22\}$, $J(5) = \{4, 20, 24\}$, and

$$J(7) = \{6, 42, 48, 295, 299, 337, 341, 2096, 2390, 14675, 16731, 16735, 102728\}.$$

The set $J(11)$ has 638 elements [1]. It is conjectured that $J(p)$ is finite for all $p$, which by Theorem 3 is equivalent to saying $|H_n|_p \rightarrow \infty$ as $n \rightarrow \infty$ for all $p$.

The rest of our discussion will focus on the particular sum $H_{p-1}$.
**Theorem 4.** For each odd prime \( p \), \( H_{p-1} \in p\mathbb{Z}_p \). Therefore \( J(p) \neq \emptyset \) for \( p > 2 \).

**Proof.** In the sum

\[
H_{p-1} = 1 + \frac{1}{2} + \cdots + \frac{1}{p-1},
\]

each term is a unit in \( \mathbb{Z}_p \), so we can reduce modulo \( p \) and check that \( H_{p-1} \equiv 0 \mod p\mathbb{Z}_p \).

The integers \( \{1, 2, \ldots, p-1\} \) represent the units modulo \( p \), so their inverses do as well. Thus, after replacing each \( 1/k \) with the integer from 1 to \( p-1 \) that is equal to it in \( \mathbb{Z}_p/(p) \),

\[
H_{p-1} \equiv 1 + 2 + \cdots + p-1 \mod p
\]

\[
\equiv \frac{p(p-1)}{2} \mod p
\]

\[
\equiv 0 \mod p,
\]

where in the last step we needed \( p \neq 2 \).

\[\square\]

In Table 3, we factor the initial harmonic sums into primes, and can see in this data (for \( p = 5, 7, \text{ and } 11 \)) that not only is \( H_{p-1} \) divisible by \( p \), but by \( p^2 \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( H_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3/2</td>
</tr>
<tr>
<td>3</td>
<td>11/2 \cdot 3</td>
</tr>
<tr>
<td>4</td>
<td>5^2/2^2 \cdot 3</td>
</tr>
<tr>
<td>5</td>
<td>137/2^2 \cdot 3 \cdot 5</td>
</tr>
<tr>
<td>6</td>
<td>7^2/2^2 \cdot 5</td>
</tr>
<tr>
<td>7</td>
<td>3 \cdot 11^2/2^2 \cdot 5 \cdot 7</td>
</tr>
<tr>
<td>8</td>
<td>761/2^4 \cdot 5 \cdot 7</td>
</tr>
<tr>
<td>9</td>
<td>7129/2^3 \cdot 3^2 \cdot 5 \cdot 7</td>
</tr>
<tr>
<td>10</td>
<td>11^2 \cdot 61/2^3 \cdot 3^2 \cdot 5 \cdot 7</td>
</tr>
<tr>
<td>11</td>
<td>97 \cdot 863/2^3 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11</td>
</tr>
</tbody>
</table>

**Table 3.** Harmonic sums fully factored

**Theorem 5** (Wolstenholme [5], 1862). For each prime \( p \geq 5 \), \( H_{p-1} \in p^2\mathbb{Z}_p \).

**Proof.** We collect terms in \( H_{p-1} \) that are equidistant from the middle of the sum:

\[
H_{p-1} = 1 + \frac{1}{2} + \cdots + \frac{1}{p-1}
\]

\[
= \left(1 + \frac{1}{p-1}\right) + \left(\frac{1}{2} + \frac{1}{p-2}\right) + \cdots + \left(\frac{1}{(p-1)/2} + \frac{1}{(p+1)/2}\right)
\]

\[
= \sum_{k=1}^{(p-1)/2} \left(\frac{1}{k} + \frac{1}{p-k}\right)
\]

\[
= \sum_{k=1}^{(p-1)/2} \frac{p}{k(p-k)}
\]

\[
= p \sum_{k=1}^{(p-1)/2} \frac{1}{k(p-k)}.
\]
Since a $p$ has been pulled out, we want to show this last sum is in $p\mathbb{Z}_p$. The terms in the sum are $p$-adic units, and reducing the terms modulo $p$ yields

$$
\frac{(p-1)/2}{\sum_{k=1}^{(p-1)/2} \frac{1}{k(p-k)}} \equiv -\frac{(p-1)/2}{\sum_{k=1}^{(p-1)/2} k^2} \pmod{p\mathbb{Z}_p}.
$$

The numbers $1^2, \ldots, ((p - 1)/2)^2$ represent all the nonzero squares modulo $p$, so their reciprocals also represent the nonzero squares modulo $p$. Therefore

$$
\sum_{k=1}^{(p-1)/2} \frac{1}{k^2} \equiv \sum_{k=1}^{(p-1)/2} k^2 \pmod{p\mathbb{Z}_p}.
$$

Using the formula $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$ with $n = (p - 1)/2$,

$$
\sum_{k=1}^{(p-1)/2} k^2 = \frac{p(p^2 - 1)}{24}.
$$

Since $p > 3$, $(p, 24) = 1$, so this sum is in $p\mathbb{Z}_p$ and we’re done. \qed

Theorem 5 says $\text{ord}_p(H_{p-1}) \geq 2$. If one calculates $H_{p-1}$ for primes $3 < p < 10000$, always $\text{ord}_p(H_{p-1}) = 2$. But at $p = 16843$ and at $p = 2124679$, $\text{ord}_p(H_{p-1}) = 3$. No further examples where $\text{ord}_p(H_{p-1}) > 2$ are known.

References