Expressing harmonic sums as fractions

G.J.O. Jameson (*Math. Gazette* **101**, July 2017)

Introduction

As usual, write

$$H_n = 1 + \frac{1}{2} + \ldots + \frac{1}{n}.$$

In [Lo], it is shown, by an elegant method, that for all $n \geq 2$, H_n is not an integer. The method can be traced back at least to [PSz, Exercise 251, p. 159]. Actually, rather more is shown: if H_n is expressed as a fraction a_n/b_n in its lowest terms, then for all $n \geq 2$, b_n is even. One way of looking at this is to say that once the term $\frac{1}{2}$ has entered the sum, the factor 2 persists in the denominator from then on. This suggests the following question:

(Q1) for primes $p \geq 3$, is b_n a multiple of p for all $n \geq p$?

and more generally:

(Q2) for any prime p, is b_n a multiple of p^m for all $n \geq p^m$?

In other words, once $1/p^m$ has entered the sum, does the factor p^m persist in the denominator?

Let d_n denote the lowest common multiple of $1, 2, \ldots, n$. This number can be described as follows: for each prime $p \leq n$, let m_p be the largest integer m such that $p^m \leq n$. Then $d_n = \prod_{p \leq n} p^{m_p}$. Hence a positive answer to (Q2) would imply that b_n simply equals d_n.

At the same time, one can ask similar questions about the alternating sum

$$H_n^* = \sum_{r=1}^{n} \frac{(-1)^{r-1}}{r}.$$

Of course, H_n^* converges to $\log 2$ as $n \to \infty$, while H_n tends to infinity. However, these facts are not really relevant to our investigation. We note the following equivalent expressions for H_{2n}^* and H_{2n+1}^*:

$$H_{2n}^* = H_{2n} - 2\sum_{r=1}^{n} \frac{1}{2r} = H_{2n} - H_n = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n}, \tag{1}$$

and, with the term $\frac{1}{2n+1}$ added,

$$H_{2n+1}^* = H_{2n+1} - H_n = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n+1}. \tag{2}$$

We record the first few values of H_n and H_n^*:

n	1	2	3	4	5	6	7	8
H_n	1	$\frac{3}{2}$	$\frac{11}{6}$	$\frac{25}{12}$	$\frac{137}{60}$	$\frac{49}{20}$	$\frac{363}{140}$	$\frac{761}{280}$
H_n^*	1	$\frac{1}{2}$	$\frac{5}{6}$	$\frac{7}{12}$	$\frac{47}{60}$	$\frac{37}{60}$	$\frac{319}{420}$	$\frac{533}{840}$

For H_n, this is already enough to give a negative answer to both questions: the denominator 20 in H_6 is not a multiple of 3. So we need to modify the questions to something more discriminating. Are there at least some primes p for which (Q2) has a positive answer? For other primes p, can we at least establish some facts about the powers of p occurring in the denominator? We will present some answers, initially using only a slight generalisation of the method of [Lo].

Two lemmas

We give ourselves some notation to facilitate the discussion. Given a prime p, denote by N_p the set of integers that are not multiples of p. By a basic property of primes, if r and s are in N_p, then so is rs. Now consider a non-zero rational number $a = r/s$ in lowest terms (not excluding an integer $r = r/1$). By combining the prime factorisations of r and s, we see that there is a unique integer m (positive, 0 or negative) such that a is expressible as $p^m r_1/s_1$, where r_1 and s_1 are in N_p (here it is not essential for r_1 and s_1 to be coprime). We denote this m by $R_p(a)$. So a positive value of $R_p(a)$ means that p divides the numerator r, while a negative value means that p divides the denominator s. Clearly, $R_p(1) = 0$ for all p and $R_p(a/p) = R_p(a) - 1$. We now establish two simple facts about $R_p(a+b)$, which we will use repeatedly:

LEMMA 1. *Let p be prime. Then:*

(i) If $R_p(a) = R_p(b) = m$ and $a + b \neq 0$, then $R_p(a+b) \geq m$.

(ii) If $R_p(a) < R_p(b)$, then $R_p(a+b) = R_p(a)$.

Proof: (i) Then $a = p^m r_1/s_1$ and $b = p^m r_2/s_2$, where r_j, $s_j \in N_p$ for $j = 1, 2$. So

$$a + b = p^m \, \frac{r_1 s_2 + r_2 s_1}{s_1 s_2}.$$

Here $s_1 s_2 \in N_p$ and $r_1 s_2 + r_2 s_1$ may or may not be a multiple of p, hence $R_p(a+b) \geq m$.

(ii) We now have $a = p^m r_1/s_1$ and $b = p^n r_2/s_2$, where $m < n$ and r_j, $s_j \in N_p$ for $j = 1, 2$. Then

$$a + b = p^m \left(\frac{r_1}{s_1} + \frac{p^{n-m} r_2}{s_2} \right) = p^m \, \frac{t}{s_1 s_2},$$

where $t = r_1 s_2 + p^{n-m} r_2 s_1$. Since $r_1 s_2$ is in N_p, so is t. Hence $R_p(a+b) = m$. $\square$

Note that inequality can easily hold in (i): for example, $R_3(4) = R_3(5) = 0$, but $R_3(4+5) = R_3(9) = 2$.

COROLLARY 1.1. *Suppose that $R_p(a_j) \geq m$ for $1 \leq j \leq n$ and $a_1 + a_2 + \cdots + a_n = s_n \neq 0$. Then $R_p(s_n) \geq m$.*

Proof: By Lemma 1, if $R_p(a) \geq m$, $R_p(b) \geq m$ and $a + b \neq 0$, then $R_p(a+b) \geq m$. The statement follows, by repeating this (if $a_1 + \cdots + a_r = 0$ for some r, choose the largest such r and start the sum at a_{r+1}). □

The following obvious restatement of Lemma 1 will be used several times:

COROLLARY 1.2. *Let $R_p(a - b) \geq m$. If $R_p(b) \geq m$, then $R_p(a) \geq m$, and if $R_p(b) \leq m - 1$, then $R_p(a) = R_p(b)$.* □

Where it results in tidier statements, we will use the notation A_p for the set of rationals a such that $R_p(a) \geq 1$. By Lemma 1, if a and b are in A_p, then so is $a+b$. Also, if $R_p(a) \leq 0$ and $b \in A_p$, then $R_p(a+b) = R_p(a)$. (Yet another alternative notation for $R_p(a-b) \geq 1$, which we will not use, is $a \equiv b \bmod p$.)

We record some immediate consequences for H_n and H_n^*. For any finite set of positive integers, let $H(E) = \sum_{r \in E} \frac{1}{r}$. If $E \subset N_p$, then $R_p(\frac{1}{r}) = 0$ for each $r \in E$, so by Corollary 1.1, $R_p[H(E)] \geq 0$. In particular, $R_p(H_n) \geq 0$ for $n < p$. Similarly for H_n^*. As the table shows, it is quite possible to have $R_p(H_n) > 0$ for such n: for example, $R_3(H_2) = 1$ and $R_5(H_4) = 2$.

No two consecutive numbers have $H_n \in A_p$: indeed, if $R_p(H_n) > 0$, then $R_p(H_{n+1}) \leq 0$, since $H_{n+1} = H_n + \frac{1}{n+1}$ and $R_p(\frac{1}{n+1}) \leq 0$. Similarly for H_{n-1} and for H_n^*.

Another very simple Lemma describes what happens if the numerator and denominator of a fraction are adjusted mod p^m:

LEMMA 2. *Let $a = r/s$ and $a' = r'/s'$, where $s, s' \in N_p$ and $r \equiv r'$ and $s \equiv s'$ mod p^m. Then $R_p(a - a') \geq m$.*

Proof. Then

$$a - a' = \frac{rs' - r's}{ss'},$$

in which $rs' \equiv r's \bmod p^m$ and $ss' \in N_p$. So $R_p(a - a') \geq m$. □

By Corollary 1.1, the statement extends to a sum of terms: if a_j relates to a_j' in the way stated for each j and $a = \sum_{j=1}^n a_j$, $a' = \sum_{j=1}^n a_j'$, then $R_p(a - a') \geq m$. Note that this is ready-made for an application of Corollary 1.2.

Remark: The stated congruences in Lemma 2 are not preserved if r'/s' is replaced by an equivalent fraction. This hypothesis needs to be treated with a bit of care!

Denote by $E + kp$ the translated set $\{kp + r : r \in E\}$. In $H(E + kp)$, the term $\frac{1}{r}$ has been replaced by $1/(kp + r)$, so as a case of Lemma 2 (with $m = 1$), we have:

COROLLARY 2.1. *Let $E \subset N_p$ and $k \geq 1$. Then $R_p[H(E) - H(E + kp)] \geq 1$. So if $R_p[H(E)] \geq 1$, then $R_p[H(E + kp)] \geq 1$, and if $R_p[H(E)] = 0$, then $R_p[H(E + kp)] = 0$. Similarly for $H^*(E)$.* □

We give a specific example of this, which we will apply later:

Example 1: Let $J_{k,r} = H_{5k+r} - H_{5k} = \sum_{j=1}^{r} 1/(5k + j)$. Since $R_5(H_r) = 0$ for $1 \leq r \leq 3$ and $R_5(H_4) = 2$, we have $R_5(J_{k,r}) = 0$ for $1 \leq r \leq 3$ and $R_5(J_{k,4}) \geq 1$.

The basic result, and applications

For n between p^m and p^{m+1}, the smallest possible value of $R_p(H_n)$ or $R_p(H_n^*)$ is $-m$. If the value is $-m$, then the factor p^m has persisted in the denominator. Larger values mean that some cancellation has occurred; non-negative values mean that p does not appear in the denominator at all. We now present our most basic result.

THEOREM 3. *Let p be prime and $kp^m \leq n < (k + 1)p^m$, where $k \geq 1$ and $m \geq 1$. Then:*

(i) If $R_p(H_k) \leq 0$, then $R_p(H_n) = R_p(H_k) - m$.

(ii) If $R_p(H_k) \geq 1$, then $R_p(H_n) \geq 1 - m$.

Similar statements hold for H_n^* for $p \geq 3$.

Proof. Separating out multiples of p^m, we have

$$H_n = \sum_{r=1}^{k} \frac{1}{rp^m} + s_n = \frac{H_k}{p^m} + s_n, \tag{3}$$

where s_n is the sum of terms $\frac{1}{r}$ with $r \leq n$ and r not a multiple of p^m. For each such r, $R_p(\frac{1}{r}) \geq 1 - m$, so by Corollary 1.1, $R_p(s_n) \geq 1 - m$. Now $R_p(H_k/p^m) = R_p(H_k) - m$. In case (i), it follows from Lemma 1(ii) that $R_p(H_n)$ equals this value. In case (ii), it follows from Corollary 1.1 again that $R_p(H_n) \geq 1 - m$.

Now consider H_n^*. For prime $p \geq 3$, rp^m has the same parity as r, so $(-1)^{rp^m-1} =$

$(-1)^{r-1}$, and the identity for H_n^* analogous to (3) is

$$H_n^* = \sum_{r=1}^{k} \frac{(-1)^{r-1}}{rp^m} + t_n = \frac{H_k^*}{p^m} + t_n, \tag{4}$$

where $R_p(t_n) \geq 1 - m$ for the same reason as s_n. The statements now follow as before. □

We can read off numerous applications and special cases. First, we restate the case $k = 1$ (note that $H_1 = H_1^* = 1$):

COROLLARY 3.1. *For all primes p, if $p^m \leq n < 2p^m$, then $R_p(H_n) = -m$. If $p \geq 3$, then $R_p(H_n^*) = -m$.* □

Next, we show that for both H_n and H_n^*, the answer to (Q2) is positive for $p = 2$.

COROLLARY 3.2. *For $2^m \leq n < 2^{m+1}$, we have $R_2(H_n) = R_2(H_n^*) = -m$.*

Proof. For H_n, this is Corollary 3.1. For H_n^*, this Corollary does not apply, but we can deduce the result from the statement for H_n, as follows. Let $m \geq 2$. By (1) and (2), for n equal to $2k$ or $2k+1$, we have $H_n^* = H_n - H_k$. Then $R_2(H_n) = -m$ and $R_2(H_k) = 1 - m$, so $R_2(H_n^*) = -m$. □

Next, we restate the case $k = 2$:

COROLLARY 3.3. *Let p be prime and $2p^m \leq n < 3p^m$. Then $R_p(H_n) = -m$ for all $p \geq 5$ and $R_p(H_n^*) = -m$ for all $p \geq 3$.*

Proof. Take $k = 2$ in Theorem 3. Since $H_2 = \frac{3}{2}$ and $H_2^* = \frac{1}{2}$, we have $R_p(H_2) = 0$ for all $p \geq 5$ and $R_p(H_2^*) = 0$ for $p \geq 3$. □

Combining Corollaries 3.2 and 3.3, we see at once that $p = 3$ satisfies (Q2) for H_n^*:

COROLLARY 3.4. *For $3^m \leq n < 3^{m+1}$, we have $R_3(H_n^*) = -m$.* □

However, $R_3(H_6) = R_3(\frac{49}{20}) = 0$, so the corresponding statement for H_n is false. Later, at the cost of some work, we will give a full description of $R_3(H_n)$ for all n. For now, we summarise what follows easily from Theorem 3:

Example 2: $R_3(H_n)$. By Corollary 3.1, $R_3(H_n) = -m$ for $3^m \leq n < 2.3^m$. However, $R_3(H_2) = R_3(\frac{3}{2}) = 1$, so for $2.3^m \leq n < 3^{m+1}$, we have $R_3(H_n) \geq 1 - m$: the interval $[3^m, 3^{m+1})$ divides into two blocks with contrasting results. In the case $m = 1$, this says that $R_3(H_n) \geq 0$ for $n = 6, 7, 8$. In fact, as seen in the table, $R_3(H_6) = R_3(H_8) = 0$ and $R_3(H_7) = 1$. By Theorem 3 again, it follows, for example, that $R_3(H_n) = -1$ for $18 \leq n \leq 20$, while $R_3(H_n) \geq 0$ for $21 \leq n \leq 23$.

Example 3: $R_5(H_n)$. Since $R_5(H_k) = 0$ for $k = 1, 2, 3$, we have $R_5(H_n) = -m$ for $5^m \le n < 4.5^m$. However, $R_5(H_4) = 2$, hence $R_5(H_n) \ge 1 - m$ for $4.5^m \le n < 5^{m+1}$. In particular, $R_5(H_n) = -1$ for $5 \le n \le 19$, while $R_5(H_n) \ge 0$ for $20 \le n \le 24$. Meanwhile, $R_5(H_3^*) = 1$, so $R_5(H_n^*) \ge 0$ for $15 \le n \le 19$.

Example 4: $R_7(H_n)$. Note that $R_7(H_k) = 0$ for $1 \le k \le 5$, while $R_7(H_6) = 2$. Hence $R_7(H_n) = -1$ for $7 \le n \le 41$, while $R_7(H_n) \ge 0$ for $42 \le n \le 48$.

Let $H_n = a_n/b_n$ and $H_n^* = a_n^*/b_n^*$, in lowest terms. Clearly, a prime p fails property (Q1) for H_n whenever p appears as a factor of a_k for some $k < p$ (and similarly for H_n^*). For H_n, our short list of values is enough to show that 3, 5, 7 and 11 fail (Q1). For H_n^*, the values at 3, 4, 7 and 8 show, respectively, that 5, 7, 11 and 13 fail (Q1). Later, we will establish general results implying that no primes beyond 3 satisfy (Q1) for either H_n or H_n^*.

Are we in a position to determine b_n and b_n^* for any given n? Not in all cases, because of the uncertainty in case (ii) of Theorem 3, but we can do so for a number of particular values. We need to consider each prime $p \le n$. By Corollaries 3.1 and 3.3, we can dispose simultaneously of quite a lot of the primes in question:

COROLLARY 3.5. *Let $n \ge 10$. Then for all primes p such that $\frac{n}{3} < p \le n$, we have $R_p(H_n) = R_p(H_n^*) = -1$.*

Proof. The assumptions ensure that $p \ge 5$ and $p \le n < 3p$, so the statement is the case $m = 1$ of Corollaries 3.1 and 3.3. □

Some examples follow. It is natural to compare the answer with d_n, the lowest common multiple of $1, 2, \ldots, n$.

Example 5: H_{21}. First, note that

$$d_{21} = 2^4 \times 3^2 \times 5 \times 7 \times 11 \times 13 \times 17 \times 19.$$

By Corollary 3.2, $R_2(H_{21}) = -4$. By Examples 1, 2 and 3, $R_3(H_{21}) \ge 0$, $R_5(H_{21}) \ge 0$ and $R_7(H_{21}) = -1$. By Corollary 3.5, $R_p(H_{21}) = -1$ for $11 \le p \le 19$. So

$$b_{21} = 2^4 \times 7 \times 11 \times 13 \times 17 \times 19.$$

All this applies equally to H_{22}.

Example 6: H_{29}, H_{30}. We have

$$d_{29} = d_{30} = 2^4 \times 3^3 \times 5^2 \times 7 \times 11 \times 13 \times 17 \times 19 \times 23 \times 29.$$

We show that $b_{29} = b_{30} = d_{30}$: no factors cancel. We refer to Corollary 3.1 for $p = 2$, 3 and 5, Example 3 for $p = 7$, and Corollary 3.5 for $11 \le p \le 29$.

The reader may care to check that b^*_{30} has all the same factors except 7, and to investigate some other cases, for example H^*_{21} and H_{42}.

The pairing method; Wolstenholme's theorem

Recall that $H_4 = \frac{25}{12}$ and $H_6 = \frac{49}{20}$. One is tempted to suspect that $R_p(H_{p-1}) \ge 2$ for all primes $p \ge 5$. This is indeed true: it is a theorem of Wolstenholme, dating from 1862. For the moment, we will only prove the weaker statement $R_p(H_{p-1}) \ge 1$. This can be done very neatly by combining terms in pairs, as follows:

PROPOSITION 4. *For all primes $p \ge 3$, we have $R_p(H_{p-1}) \ge 1$.*

Proof. The numbers $1, 2, \ldots, p-1$ can equallly well be listed as $p-r$ for $1 \le r \le p-1$. Hence $2H_{p-1} = \sum_{r=1}^{p-1} c_r$, where

$$c_r = \frac{1}{r} + \frac{1}{p-r} = \frac{p}{r(p-r)}. \tag{5}$$

Since $R_p(c_r) \ge 1$ for each r, it follows that $R_p(2H_{p-1}) \ge 1$, hence $R_p(H_{p-1}) \ge 1$. □

Applying Theorem 3 with $k = p-1$, we deduce at once:

COROLLARY 4.1. *For all primes $p \ge 3$, we have $R_p(H_n) \ge 0$ for $p(p-1) \le n < p^2$, so p fails (Q1).* □

The statement extends easily to blocks of terms between multiples of p. Write

$$H_{j,p} = \sum_{r=1}^{p-1} \frac{1}{jp+r}$$

and $\tilde{H}_{kp} = \sum_{j=0}^{k-1} H_{j,p}$. This equates to the sum H_{kp} with multiples of p excluded.

PROPOSITION 5. *Let $p \ge 3$ be prime. For integers j, $k \ge 1$, we have $R_p(\tilde{H}_{kp}) \ge 1$ and $R_p(H_{j,p}) \ge 1$. Also, $R_p(\tilde{H}_{kp^2}) \ge 2$.*

Proof. The members of $\tilde{H}_{kp}$ can equally be written as $kp - r$ for $r \in \tilde{H}_{kp}$. The first statement follows as in Proposition 4, using

$$\frac{1}{r} + \frac{1}{kp-r} = \frac{kp}{r(kp-r)}.$$

Similarly for $\tilde{H}_{kp^2}$, with kp replaced by kp^2, and for $H_{j,p}$, using $jp+r$ and $jp+p-r$. □

COROLLARY 5.1. *Let $p \geq 3$ be prime. Then:*

(i) *If $R_p(H_k) \leq 1$, then $R_p(H_{kp}) = R_p(H_k) - 1$;*

(ii) *If $R_p(H_k) \geq 2$, then $R_p(H_{kp}) \geq 1$.*

The same applies to $H_{(k+1)p-1}$.

Proof. Separating multiples and non-multiples of p, we have

$$H_{kp} = \frac{H_k}{p} + \tilde{H}_{kp}, \qquad H_{(k+1)p-1} = \frac{H_k}{p} + \tilde{H}_{(k+1)p}.$$

Since $R_p(H_k/p) = R_p(H_k) - 1$, the statements follow, by Lemma 1. □

The pairing idea has further applications. Observe from the table of values that $R_{11}(H_3) = 1$ and $R_{11}(H_7) = 2$. This is a special case of the following result:

PROPOSITION 6. *Let $p \geq 5$ be prime and $1 \leq n < p$. Then:*

(i) *If $R_p(H_n) = 0$, then $R_p(H_{p-n-1}) = 0$;*

(ii) *If $R_p(H_n) \geq 1$, then $R_p(H_{p-n-1}) \geq 1$.*

Proof. Let

$$W_n = \sum_{r=1}^{n} \frac{1}{p-r} = H_{p-1} - H_{p-n-1}.$$

With the notation of (8), $H_n + W_n = \sum_{r=1}^{n} c_r$, so $R_p(H_n + W_n) \geq 1$. Now

$$H_{p-n-1} = H_{p-1} - W_n = H_{p-1} - (H_n + W_n) + H_n.$$

Since $R_p(H_{p-1}) \geq 1$, the statements follow by Lemma 1. □

In particular, $R_p(H_{p-2}) = R_p(H_{p-3}) = 0$. Also, $R_p(H_{p-2} - 1) \geq 1$.

Example 7. From the table, we see that $R_{17}(H_n) = 0$ for $1 \leq n \leq 8$. So we can conclude, without further calculation, that the same applies for all $n \leq 15$.

The pairing method also delivers a rather striking result for H_n^*. As observed earlier, the original table shows that for $p = 5$, 7, 11 and 13, $R_p(H_n^*) = 1$ for $n = 3$, 4, 7 and 8 respectively. The fact that $R_{13}(H_8^*) \geq 1$ is seen much more clearly from (1):

$$H_8^* = \tfrac{1}{5} + \tfrac{1}{6} + \tfrac{1}{7} + \tfrac{1}{8} = (\tfrac{1}{5} + \tfrac{1}{8}) + (\tfrac{1}{6} + \tfrac{1}{7}) = \tfrac{13}{40} + \tfrac{13}{42}.$$

This points the way to the proof of the general result, which is as follows. As usual, $[x]$ denotes the largest integer not greater than x.

THEOREM 7. *Let $p \geq 5$ be prime, and let $n = [\frac{2}{3}p]$. Then $R_p(H_n^*) \geq 1$.*

Proof. First, suppose p is of the form $3k+1$. Then $n = 2k$. By (1), $H_n^* = \sum_{r=k+1}^{2k} \frac{1}{r}$. But the numbers from $k+1$ to $2k$ can equally be written as $3k+1-r = p-r$ for $k+1 \leq r \leq 2k$. Hence

$$2H_n^* = \sum_{r=k+1}^{2k} \left(\frac{1}{r} + \frac{1}{p-r} \right).$$

So $R_p(2H_n^*) \geq 1$.

Now suppose that $p = 3k+2$. Then $n = 2k+1$ and by (2), $H_n^* = \sum_{r=k+1}^{2k+1} \frac{1}{r}$. The numbers r from $k+1$ to $2k+1$ can be listed as $p-r$ for the same range of r, and the statement follows in the same way. □

COROLLARY 7.1. *With this notation, $R_p(H_r^*) \geq 0$ for $np \leq r < (n+1)p$. So no primes $p \geq 5$ satisfy (Q1) for H_n^*.* □

For readers with the appetite for it, we now prove Wolstenholme's theorem itself by a further development of the pairing method. This proof folllows [Gar]: a different method can be seen in [HWr, p. 88–90]). We will need the fact, derived from Bezout's identity, that the numbers $1, 2, \ldots, p-1$ form a group G_p under multiplication mod p. Each element r has an inverse r^{-1} in this group, so that $rr^{-1} \equiv 1 \bmod p$ (note: in this context, r^{-1} does *not* mean $\frac{1}{r}$!).

Write $N_{k,p} = \prod_{r=1}^{p-1}(kp+r)$, so that, in particular, $N_{0,p} = (p-1)!$.

THEOREM 8. *Let $p \geq 5$ be prime. Then*

$$H_{p-1} = \frac{p^2 M_p}{(p-1)!}, \tag{6}$$

$$H_{k,p} = \frac{p^2 M_{k,p}}{N_{k,p}} \tag{7}$$

for some integers M_p and $M_{k,p}$. Hence $R_p(H_{p-1})$, $R_p(H_{k,p})$ and $R_p(\tilde{H}_{kp})$ are at least 2.

Proof. With c_r as in (5), we have $2H_{p-1} = \sum_{r=1}^{p-1} c_r$, so $2(p-1)!H_{p-1} = p\sum_{r=1}^{p-1} u_r$, where

$$u_r = \frac{(p-1)!}{r(p-r)}.$$

Note that u_r is an integer: in fact, it is the product of the numbers from 1 to $p-1$ leaving out r and $p-r$. Hence, mod p, we have

$$u_r \equiv u_r.r(p-r)r^{-1}(p-r)^{-1} \equiv (p-1)!r^{-1}(p-r)^{-1} \equiv -(p-1)!(r^{-1})^2.$$

The numbers r^{-1}, for $1 \leq r \leq p-1$, are simply the group elements, in other words the numbers $1, 2, \ldots, p-1$ in a different order. Hence

$$\sum_{r=1}^{p-1} (r^{-1})^2 = \sum_{r=1}^{p-1} r^2 = \tfrac{1}{6}(p-1)p(2p-1).$$

Now $p-1$ is even and either $p-1$ or $2p-1$ is a multiple of 3, since p is congruent to either 1 or 2 mod 3. Hence $\frac{1}{6}(p-1)(2p-1)$ is an integer, and $\sum_{r=1}^{p-1} u_r$ is a multiple of p. So $(p-1)!H_{p-1}$ is a multiple of p^2.

The expression in (6) is certainly not in its lowest terms, since the denominator can at least be reduced to d_{p-1}. However, p does not divide into $(p-1)!$, so p^2 is preserved when we move to the lowest terms, and hence $R_p(H_{p-1}) \geq 2$.

The proof of (7) is similar. In the product expression for u_r, the integer j is replaced by $kp + j$, which is the same mod p. □

Remark. By Wilson's theorem, $(p-1)! \equiv -1 \bmod p$, but our proof did not need this fact.

Corollary 5.1 can now be strengthened, as follows (with proof as before):

COROLLARY 8.1. *Let $p \geq 5$ be prime. Then:*

(i) If $R_p(H_k) \leq 2$, then $R_p(H_{kp}) = R_p(H_k) - 1$;

(ii) If $R_p(H_k) \geq 3$, then $R_p(H_{kp}) \geq 2$.

The same applies to $H_{(k+1)p-1}$. □

COROLLARY 8.2. *For primes $p \geq 5$, we have $R_p(H_{(p-1)p}) \geq 1$ and $R_p(H_{p^2-1}) \geq 1$. Furthermore, if $R_p(H_{p-1}) = 2$, then both these values are exactly 1.* □

Are there actually any primes for which $R_p(H_{p-1}) \geq 3$? According to [B, p. 293], just two such primes are known: 16,843 and 2,124,679. Remarkably, this is equivalent to p dividing the numerator of the Bernoulli number B_{p-3} (see [Gar]).

By Theorem 3, we can deduce the following further Corollary, showing that cancellation of the factor p always extends at least to a block ending with $p^3 - 1$:

COROLLARY 8.3. *For primes $p \geq 5$, we have $R_p(H_n) \geq 0$ for $p^3 - p \leq n < p^3$.* □

Example 8. We can now clear up the uncertain area in Example 3 for $R_5(H_n)$. By Corollary 8.2, $R_5(H_{20}) = R_5(H_{24}) = 1$. By Example 1, for $n = 21, 22, 23$, we have $R_5(H_n - H_{20}) = 0$, hence $R_5(H_n) = 0$.

Full description of $R_3(H_n)$

We now establish a complete description of $R_3(H_n)$ for all n. As we will see, the pattern is quite intricate. It also gives a fair indication of what happens for larger primes.

We proceed step by step in powers of 3. We start by restating the values for $3 \le n \le 8$, seen in our original table:

n	$3,4,5$	6	7	8
$R_3(H_n)$	-1	0	1	0

By Theorem 3, we deduce for $3^2 \le n < 3^3$:

n	9-17	18-20	21-23	24-26
$R_3(H_n)$	-2	-1	≥ 0	-1

By Corollary 5.1, $R_3(H_{21}) = R_3(H_{23}) = 0$. (In the same way, we could have deduced the values $R_3(H_6) = R_3(H_8) = 0$ from $H_2 = \frac{3}{2}$ without calculating H_6 and H_8.) For H_{22}, a quick partial answer is delivered by the following slight variation of Corollary 5.1: we have $H_{22} = \tilde{H}_{21} + a$, where $a = \frac{1}{3}H_7 + \frac{1}{22} = \frac{121}{140} + \frac{1}{22}$. Now $22 \times 121 + 140 \equiv 1 + 2 = 0 \bmod 3$ and $22 \times 140 \in N_3$. So $R_3(a) \ge 1$, hence $R_3(H_{22}) \ge 1$.

But is the value 1, or more than 1? To answer this, we will have to consider the numerator mod 9 and split H_{22} in a different way. The following method is a special case of one described in [EL] for the purpose of establishing a general criterion for H_n to be in A_p. Note that by Proposition 5, $R_3(\tilde{H}_{18}) \ge 2$. Now $H_{22} = \tilde{H}_{18} + b$, where

$$b = \tfrac{121}{140} + \tfrac{1}{19} + \tfrac{1}{20} + \tfrac{1}{22}.$$

By Lemma 2, adjusting numerators and denominators mod 9, we have $R_3(b - b') \ge 2$, where

$$b' = \tfrac{4}{5} + \tfrac{1}{1} + \tfrac{1}{2} + \tfrac{1}{4} = \tfrac{51}{20} = 3 - \tfrac{9}{20}.$$

So $H_{22} = \tilde{H}_{18} + (b - b') - \frac{9}{20} + 3$. This shows that $R_3(H_{22}) = 1$, completing our list of values of $R_3(H_n)$ for $3^2 \le n < 3^3$. More specifically, it shows that H_{22} can be expressed in the form $9r/s + 3$, where $s \in N_3$.

Not much more work is now needed to determine the complete pattern of $R_3(H_n)$. By the values now found, Theorem 3 and (for 66 and 68) Corollary 5.1, we have the following distribution for $3^3 \le n < 3^4$:

n	27-53	54-62	63-65	66	67	68	69-71	72-80
$R_3(H_n)$	-3	-2	-1	0	≥ 0	0	-1	-2

.

To resolve H_{67}, we apply the expression $H_{22} = 9r/s + 3$ just found:

$$H_{67} = \tilde{H}_{66} + \tfrac{1}{3}H_{22} + \tfrac{1}{67} = \tilde{H}_{66} + \frac{3r}{s} + 1 + \tfrac{1}{67}.$$

Now $R_3(1+\frac{1}{67}) = 0$, hence $R_3(H_{67}) = 0$. We can substitute this value in the tabulation above.

We have now reached the point where there are no strictly positive values of $R_3(H_n)$ in the table. By Theorem 3, multiplication by 3 will result in a similar pattern for $3^4 \le n < 3^5$, consisting of six blocks with corresponding values of $R_3(H_n)$ reduced by 1, and similarly for all higher powers of 3.

In particular, 68 is the largest number n for which $R_3(H_n) \ge 0$ - hardly a fact that one would have guessed at the outset.

Some corresponding results for sums of $1/r^2$

Only slight modifications of our reasoning are needed to establish similar results for the sums

$$S_n = \sum_{r=1}^{n} \frac{1}{r^2}, \qquad S_n^* = \sum_{r=1}^{n} \frac{(-1)^{r-1}}{r^2}.$$

Again we record the first few values:

n	1	2	3	4	5	6
S_n	1	$\frac{5}{4}$	$\frac{49}{36}$	$\frac{205}{144}$	$\frac{5269}{3600}$	$\frac{5369}{3600}$
S_n^*	1	$\frac{3}{4}$	$\frac{31}{36}$	$\frac{115}{144}$	$\frac{3019}{3600}$	$\frac{973}{1200}$

For n between p^m and p^{m+1}, the least possible value is now $-2m$. The corresponding results are generally easier than before. We outline some of them, more briefly.

THEOREM 9. *Let p be prime and $kp^m \le n < (k+1)p^m$. Then:*

(i) If $R_p(S_k) \le 1$, then $R_p(S_n) = R_p(S_k) - 2m$.

(ii) If $R_p(S_k) \ge 2$, then $R_p(S_n) \ge 2 - 2m$.

Similar statements hold for S_n^* for $p \ge 3$.

Proof. The identity corresponding to (1) is

$$S_n = \sum_{r=1}^{k} \frac{1}{r^2 p^{2m}} + s_n = \frac{S_k}{p^{2m}} + s_n,$$

where s_n is a sum of terms $\frac{1}{r^2}$, each with $R_p(\frac{1}{r^2}) \ge 2 - 2m$. The statements follow from Lemma 1, as in Theorem 3. □

Since $S_1 = S_1^* = 1$, the case $k = 1$ says: if $p^m \le n < 2p^m$, then $R_p(S_n) = -2m$ for all p, and $R_p(S_n^*) = -2m$ for $p \ge 3$. Reasoning as in Corollary 3.2, we deduce:

COROLLARY 9.1. *For $2^m \le n < 2^{m+1}$, we have $R_2(S_n) = R_2(S_n^*) = -2m$.* $\square$

Since $S_2 = \frac{5}{4}$, the case $k = 2$ says: if $2p^m \le n < 3p^m$ and $p \ne 5$, then $R_p(S_n) = -2m$, while $R_5(S_n) = 1 - 2m$. Similarly, since $S_2 = \frac{3}{4}$, we have then $R_p(S_n^*) = -2m$ if $p \ne 3$, while $R_3(S_n^*) = 1 - 2m$. Combining these facts, we can state:

COROLLARY 9.2. *For $3^m \le n < 3^{m+1}$, we have $R_3(S_n) = -2m$.* $\square$

COROLLARY 9.3. *We have $R_3(S_n^*) = -2m$ for $3^m \le n < 2.3^m$, and $R_3(S_n^*) = 1 - 2m$ for $2.3^m \le n < 3^{m+1}$.* $\square$

Note that this has been achieved with much less work than the description of $R_3(H_n)$! The reader might care to write out the distribution of $R_5(S_n)$ and $R_5(S_n^*)$.

The pairing argument in Proposition 4 extends to S_n^* rather than S_n:

PROPOSITION 10. *For all primes $p \ge 3$, we have $R_p(S_{p-1}^*) \ge 1$.*

Proof. Since $(-1)^{p-r-1} = -(-1)^{r-1}$, we have

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

The result follows by combining pairs as in Proposition 4. $\square$

The method of Theorem 8 establishes the same property for S_{p-1} (for an alternative proof, see [HWr, Theorem 117]):

PROPOSITION 11. *For all primes $p \ge 5$, we have $R_p(S_{p-1}) \ge 1$.*

Proof: We have $[(p-1)!]^2 S_{p-1} = \sum_{r=1}^{p-1} v_r$, where $v_r = [(p-1)!]^2/r^2$. As in Theorem 8, $v_r \equiv [(p-1)!]^2(r^{-1})^2 \bmod p$, and the proof finishes as before. $\square$

By Theorem 9, it follows that for all primes $p \ge 5$, we have $R_p(S_n) \ge -1$ and $R_p(S_n^*) \ge -1$ for $p(p-1) \le n < p^2 - 1$ (so the property seen in Corollaries 9.1 and 9.2 does not extend to any other primes).

As mentioned after Theorem 8, only two primes are known for which $R_p(H_{p-1}) \ge 3$. It is shown in [Gar] that this rare property is equivalent to $R_p(S_{p-1}) \ge 2$. We present a rather more direct proof.

LEMMA 12. *We have*

$$\sum_{r=1}^{p-1} \frac{1}{r(p-r)} + S_{p-1} = p^2 \sum_{r=1}^{p-1} \frac{1}{2r^2(p-r)^2}.$$

Proof. Observe that
$$\frac{1}{r(p-r)} + \frac{1}{r^2} = \frac{p}{r^2(p-r)}.$$
Substituting $p-r$ for r, we have also
$$\frac{1}{(p-r)r} + \frac{1}{(p-r)^2} = \frac{p}{(p-r)^2 r}.$$
Adding, we obtain
$$\sum_{r=1}^{p-1} \frac{2}{r(p-r)} + 2S_{p-1} = \sum_{r=1}^{p-1} \frac{p(p-r)+pr}{r^2(p-r)^2} = p^2 \sum_{r=1}^{p-1} \frac{1}{r^2(p-r)^2}. \qquad \square$$

PROPOSITION 13. *$R_p(H_{p-1}) \geq 3$ if and only if $R_p(S_{p-1}) \geq 2$.*

Proof. By Lemma 12,
$$2H_{p-1} = p\sum_{r=1}^{p-1} \frac{1}{r(p-r)} = -pS_{p-1} + p^3 \sum_{r=1}^{p-1} \frac{1}{2r^2(p-r)^2}.$$
The stated equivalence follows, by Lemma 1. $\square$

Summary of further results and problems concerning H_n

Further results on H_n (but not H_n^*) can be seen, for example, in [EL], [B] and [Sh]. Denote by J_p the set of integers n such that $R_p(H_n) \geq 1$. We have shown that $J_3 = \{2, 7, 22\}$ and that for all $p \geq 5$, J_p contains at least the numbers $p-1$, $p(p-1)$ and p^2-1. The prime p is called *harmonic* if J_p contains only these three numbers. In [EL], a criterion for primes to be harmonic is described, and 16 such primes below 200 are listed, starting with 5. By contrast, J_7 has 13 members, finishing with 102,728 (the calculations are already fairly elaborate for this case). In [B], computational methods are used to determine the number of members of J_p is computed for primes up to 500 (with three awkward exceptions), and harmonic primes are identified up to 10^5. The case $p = 11$ is particularly striking: J_{11} has no fewer than 638 members, the largest one lying between 11^{28} and 11^{29}.

It is conjectured that J_p is finite for all p, and also that there are infinitely many harmonic primes. More precisely, Boyd conjectures that the density of harmonic primes among all primes is $1/e$, and suggests a formula bounding the number of members of J_p. He gives plausible heuristic reasoning supporting these conjectures, but none of them has actually been proved.

The results of [Sh] include one determining the harmonic density of the set of n such that $R_p(H_n) \geq 1-m$ (where $p^m \leq n < p^{m+1}$). Also, Shiu identifies 2641 numbers up to 10,000 for which $b_n = d_n$, and conjectures that this occurs for infinitely many n.

Acknowledgement. I am grateful to Robin Chapman for drawing my attention to the pairing method and to [Sh], and to Peter Shiu for directing me to references [EL] and [B].

References

[B] D. W. Boyd, A p-adic study of the partial sums of the harmonic series, *Experimental Math.* **3** (1994), 287–302.

[EL] A. Eswarathasan and E. Levine, p-integral harmonic sums, *Discrete Math.* **91** (1991), 249–257.

[Gar] A. Gardiner, Four problems on prime power divisibility, *American Math. Monthly* **95** (1988), 926–931.

[HWr] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford Univ. Press (1979).

[Lo] Nick Lord, Quick proofs that certain sums of fractions are not integers, *Math. Gazette* **99** (2015), 128–130.

[PSz] Georg Pólya and Gabor Szegö, *Aufgaben und Lehrsätze aus der Analysis II*, Springer (1925, 1971).

[Sh] Peter Shiu, The denominators of harmonic numbers, arXiv:1607.02863vl (2016).

21 October 2016