# A p-adic Study of the Partial Sums of the Harmonic Series

David W. Boyd

**CONTENTS**



Let $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ be the $n$-th partial sum of the harmonic series. A classical result of Wolstenholme states that, if $p > 3$ is prime, the numerator of $H_{p-1}$ is divisible by $p^2$. Here we consider, for a given prime $p$, the set $J_p$ of $n$ for which $p$ divides the numerator of $H_n$. This set $J_p$ had been previously determined for $p = 2, 3, 5, 7$. One of our results is that $J_{11}$ contains exactly 638 integers, the largest of which is a number of 31 decimal digits. We determine $J_p$ for all $p < 550$ with three exceptions: 83, 127 and 397.

The computation is based on a new $p$-adically convergent formula for the quantity $H_{pn} - H_n/p$. We describe a probabilistic model for the sets $J_p$, based on branching processes. The model predicts that $|J_p| = O(p^2(\log\log p)^{2+\varepsilon})$, and that there are infinitely many $p$ with $|J_p| \geq p^2(\log\log p)^2$. This strengthens an earlier conjecture of Eswarathasan and Levine that $|J_p|$ is finite for all $p$. Another prediction of the model is that there will be infinitely many pairs $(n, p)$ for which $p^3$ divides the numerator of $H_n$, but only finitely many for which $p^4$ divides $H_n$.

It has been conjectured that there are infinitely many $p$ for which $|J_p| = 3$. We give a probabilistic argument that suggests that such primes have a density $1/e$ in the set of all primes, and experimentally confirm this by a determination of all such $p \leq 10^5$.

## 1. INTRODUCTION

The sequence of partial sums $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ of the harmonic series has some interesting and well known arithmetic properties. For example, it is known that $H_n$ is an integer only for $n = 1$, a result that Sándor [1993] attributes to J. Kürschák. Another well known result is Wolstenholme's theorem [Hardy and Wright 1960, p. 89] saying that

This research was supported in part by an operating grant from NSERC.


1058-6458/96 $0.50 per page

the numerator of $H_{p-1}$ is divisible by $p^2$ if $p > 3$ is prime. There is a relationship between values of certain $H_n$ and Fermat's quotient $q_a = (a^{p-1}-1)/p$ mod $p$. For example, a result of Eisenstein from 1850 [Dickson 1952, p. 41] states that $H_{(p-1)/2} \equiv -2q_2$ mod $p$; this is easily seen from the binomial expansion of $(1+1)^p$. A more difficult result due to Glaisher [1901, p. 50] is that $H_{[p/3]} \equiv -\frac{3}{2}q_3$ mod $p$ if $p > 3$. These results are connected to the first case of Fermat's Last Theorem via the theorems of Wieferich and Mirimanoff [Ribenboim 1979, p. 151].

If $p$ is a prime, let $J_p$ be the set of $n \geq 1$ for which $p$ divides the numerator of $H_n$. Eswarathasan and Levine [1991] showed that, if $p > 3$, the set $J_p$ contains $p-1$, $p^2-p$, and $p^2-1$, and that, for certain primes, such as $p = 5, 13, 17, 23$ and 67, there are no other elements in $J_p$. They called the primes with this property "harmonic primes," and conjectured that the set of such primes is infinite. We will show in Section 3 that the criterion for a prime to be harmonic given in that paper suggests that the set of such primes has density $1/e$ in the set of all primes. There we give the results of a computation of all the harmonic primes $p < 10^5$ and the observed density in various intervals, which agrees quite well with this conjecture.

Eswarathasan and Levine provide a systematic method for generating $J_p$, based on the congruence $H_{pn} - H_n/p \equiv 0$ mod $p^2$. If $J_p$ is finite, their method will give a proof that it is finite, given sufficient computation. They list $J_3 = \{2, 7, 22\}$, $J_5 = \{4, 20, 24\}$ (so that 5 is harmonic), and $J_7$ (a set of 13 elements), but leave open the case of $J_{11}$. The determination of $J_{11}$ is given as a research problem in [Graham et al. 1989, p. 304], where some of these results are presented as a series of exercises. As part of the computations described in Section 5, we show that $J_{11}$ has 638 elements, the largest being 10118497718552149129684042117247. This may explain why previous studies stopped at $p = 7$. However, $J_{11}$ is by no means the largest set we encountered in our computation. For example, $|J_{109}| = 1273$, $|J_{521}| = 1763$, $|J_{127}| > 2713$, $|J_{83}| >$ 5870, and $|J_{397}| > 7718$. We have determined $J_p$ for all primes $p < 550$ for which $\max J_p < p^{100}$. This omits only the three primes 83, 127 and 397. The computation uses the $p$-adic routines in PARI [Batut et al. 1993] and is based on a $p$-adically convergent series for $H_{pn} - H_n/p$ (Theorem 5.2).

In Section 6 we provide a nonrigorous probabilistic explanation for these striking observations, based on the theory of branching processes. To each prime $p$ we associate a process that predicts the number of elements in the set

$$G_m = J_p \cap [p^{m-1}, p^m - 1]$$

from the number in $G_{m-1}$. This process turns out to be *critical*, that is, $E(|G_m|) = |G_{m-1}|$ for all $m \geq 3$. By a basic result of the theory, such processes become extinct in a finite time with probability one, but the expected time to extinction is infinite. This agrees with the conjecture already made in [Eswarathasan and Levine 1991] that $J_p$ is finite for all $p$.

As described in Section 7, this branching process model allows us to predict the maximum size of the sets $|J_p|$. It predicts that $|J_p| = O(p^2(\log\log p)^{2+\varepsilon})$ and that there are infinitely many $p$ with $|J_p| \geq p^2(\log\log p)^2$. This in turn gives a plausible prediction about the possibilities for $v_p(H_n)$, the exponent of the largest power of $p$ dividing $H_n$. On simple probabilistic grounds, one would expect the number of occurrences of $v_p(H_n) = k$, for a fixed $p$, to be about $|J_p|/p^{k-1}$. We already know from Wolstenholme's theorem that $v_p(H_n) = 2$ occurs infinitely often, but our model predicts that there should be primes for which the number of occurrences of $v_p(H_n) = 2$ is arbitrarily large. The model predicts also that there are primes $p$ for which the number of $n$ with $v_p(H_n) = 3$ is arbitrarily large but of order between $(\log\log p)^2$ and $(\log\log p)^{2+\varepsilon}$. On the other hand, $v_p(H_n) \geq 4$ should occur for only a finite number of pairs $(p, n)$. Probably $v_p(H_n) \geq 4$ never occurs.

As expected, our computation found many examples of $v_p(H_n) = 2$. We found only five examples of $v_p(H_n) = 3$: four for $p = 11$, namely $n = 848$, 9338, 10583 and 3546471722268916272; and one for $p = 83$. These results are not surprising, since $|J_{11}|/11^2 = 638/121 = 5.27\ldots$, while the number of enumerated elements of $J_{83}$ is $5870 = .85\ldots \times 83^2$. On the other hand, the number of enumerated elements of $J_{397}$ is only 7718, which is less than $.04\times 397^2$. These findings are entirely consistent with our model since $(\log\log 550)^2 < 4$. No examples of $v_p(H_n) \geq 4$ were found, which agrees with the conjecture that none exist.

In order to better appreciate the interaction between experiment, model building and conjecture in this study, we now depart from conventional practice and describe the genesis of the results in this paper. I was reminded of Wolstenholme's theorem while helping a student with readings in number theory, and it seemed natural to ask whether there were any other general results of a similar nature. This led to the central question of this paper: given a prime $p$, to study the set $J_p$ of $n$ for which $H_n$ is divisible by $p$. Initially we expected that there would be little difference between the various primes. An attempt to compute $J_p$ for some small primes showed the necessity of using $p$-adic methods, because the numerator and denominator of $H_n$ grow exponentially with $n$ (see the beginning of Section 5).

Computations for $p < 60$ revealed the recursive structure of the sets $J_p$ as described in Section 3. The peculiar behaviour for $p = 11$ was soon apparent, and initially it seemed possible that $J_{11}$ might be infinite. Up to this point, the computations had been based on the naïve $p$-adic method of Section 5.1, so it was only possible to compute $J_{11} \cap [1, 11^9]$. The structure of this set suggested some sort of pseudorandom behaviour for which branching processes apparently provided a suitable model. This model, described in Section 6.1, led to the conjecture that $J_p$ is finite for all $p$. A crucial test for this conjecture would be to prove that $|J_{11}| < \infty$.

At this time, we also noticed the heuristic argument, presented in Section 4, suggesting that the set of primes for which $J_p = \{p-1,\, p^2-p,\, p^2-1\}$ should have density $1/e$. (Our working name for such primes was "dull primes"). The test for this conjecture by the computation of the dull primes less than $10^5$ seems to be fairly convincing evidence of its correctness.

In November 1993, these results were presented in a lecture at the opening of the Centre for Experimental and Constructive Mathematics at Simon Fraser University. After the lecture, Peter Borwein made a remark about the possible relevance of Bernoulli polynomials. This led to a study of the classical [Glaisher 1901] and to Theorem 5.2, which provided means for continuing the $p$-adic computation of $H_n$ to much larger values of $n$, as described in Section 5.2. The idea of finding the coefficients in this expansion by solving a linear system, hence avoiding any computation of Bernoulli numbers, is natural from the point of view of numerical analysis. With this new method, we computed $J_p$ for all $p < 100$ except $p = 83$, verifying, in particular, that $|J_{11}| < \infty$. This was presented in December 1993 in a short lecture at the Western Number Theory Meeting in Asilomar.

In January 1994, we came across [Sándor 1993] and discovered that the conjecture that $J_p$ is always finite had been anticipated by Eswarathasan and Levine [1991]. Remarkably, their conjecture was apparently based only on a computation of $J_p$ for $p = 3$, 5 and 7. They had discovered the basic recursive structure of the sets $J_p$, defined the set of harmonic primes (our "dull primes"), and conjectured that the set of such primes is infinite, as is implied by our conjecture that they have density $1/e$. Regarding $p = 11$, they say: "it even seems quite difficult to show that $J_{11}$ is finite". (The reference to [Graham et al. 1989] in their paper might give the impression that their investigations were inspired by some of the problems in that book, but Knuth informs me that it was the other way around.)

Up to that point, we had thought in terms of different branching process being defined for each prime $p$, as described in Section 6.1. However, the apparent similarity of these processes for large $p$ suggested the universal branching process described in Section 6.2, which is independent of $p$. The success of the prediction of the finiteness of $J_p$ from the branching processes model suggested the stronger conjecture that the sets $J_p$ can be approximated by random samples from this universal branching process. As a test for this conjecture and to judge the effectiveness of our new computational approach, we decided to extend the computations to include the first 100 odd primes. A summary of the data is given in Table 2 of Section 5. Once it was realized that the deterministic initial conditions described in Section 6.3 should be included, the fit of the model to this data proved to be quite good, as is evident in Table 3. This suggests that the model is essentially correct, and we conjecture that the distribution of $|J_p|$ as $p$ varies over the primes will be exactly as described by this process, once the refinement mentioned in the last sentence of Section 6.3 is incorporated. The conjectured asymptotic bounds on $|J_p|$ already mentioned do not need this refinement since they depend only on the assumption that $c_1 r^{-1/2} < P(|J_p| > r) < c_2 r^{-1/2}$ for some constants $c_1 > 0$ and $c_2$.

Of course, the computational results of Section 6 also provide a proof of Eswarathasan and Levine's conjecture in 97 out of 100 cases. In itself, this could not be regarded as very strong evidence for their conjecture since 100 primes is a rather small sample of the set of all primes. Indeed, without the guidance of the probabilistic model, the behaviour of the numbers $|G_m|$ for $p = 82$ and 397 would tend to suggest that $|J_p| = \infty$ for these primes. However, the evidence does seem to favour our probabilistic model and hence to suggest that our quantitative version of their conjecture is correct.

## 2. SOME PRELIMINARY MATERIAL

For our computations and proofs, we will need some background on the $p$-adic numbers. This can all be found in the classic book [Mahler 1981].

If $p$ is a prime and if $x \neq 0$ is a rational number, we may write $x = p^k a/b$, where $a$ and $b$ are relatively prime integers not divisible by $p$. We define the $p$-adic order (or additive valuation) of $x$ to be $v_p(x) = k$, and the $p$-adic norm (or multiplicative valuation) of $x$ by $|x|_p = p^{-k}$. We define $v_p(0) = -\infty$ and $|0|_p = 0$. Then $|\cdot|_p$ is a norm on the rationals.

The set of $p$-adic numbers $\mathbb{Q}_p$ is the completion of the rationals in the metric $d_p(x, y) = |x - y|_p$. The additive and multiplicative valuations extend to $\mathbb{Q}_p$ by continuity. Each $x \in \mathbb{Q}_p$ has a unique $p$-adic expansion $x = \sum_{k=v_p(x)}^{\infty} a_k p^k$, with $0 \leq a_k < p$, where the series converges in the $p$-adic norm. We will use the standard notation $O(p^s)$ for any $x$ for which $v_p(x) \geq s$.

An element $x \in \mathbb{Q}_p$ with $v_p(x) \geq 0$ is called a $p$-adic integer. Every ordinary integer is a $p$-adic integer, as is every rational with denominator prime to $p$.

The *ultrametric inequality* states that $v_p(x+y) \geq \min(v_p(x), v_p(y))$, with equality if $v_p(x) \neq v_p(y)$. An easy consequence of this is that $v_p(H_{p^k}) = -k$, so $\liminf_{n\to\infty} v_p(H_n) = -\infty$. As we will see in Proposition 3.3, the conjecture that $J_p$ is finite is equivalent to $\lim_{n\to\infty} v_p(H_n) = -\infty$, that is, $\lim_{n\to\infty} |H_n|_p = \infty$, analogous to the well-known $\lim_{n\to\infty} H_n = \infty$.

We will also need some facts about the Bernoulli numbers $B_n$ and Bernoulli polynomials $B_n(x)$. Following Euler [1738], the Bernoulli polynomials are defined by the exponential generating function

$$\sum_{n=0}^{\infty} \frac{B_n(x)t^n}{n!} = \frac{te^{xt}}{e^t - 1},$$

and the Bernoulli numbers by $B_n = B_n(0)$. From this it is obvious that $B_n(x)$ is a polynomial of degree $n$ whose coefficients are given explicitly by

$$B_n(x) = \sum_{k=0}^{n} \binom{n}{k} B_{n-k} x^k.$$

It follows immediately from the generating function that $B_n(x+1) - B_n(x) = nx^{n-1}$ if $n \geq 1$, so we obtain the classical formula of Jakob Bernoulli [1713, pp. 95–98]:

$$\sum_{k=1}^{m-1} k^{n-1} = \frac{B_n(m) - B_n}{n} = \sum_{k=1}^{n} \frac{1}{n}\binom{n}{k} B_{n-k} m^k \tag{2.1}$$

It is an elementary fact that $B_n = 0$ for $n$ odd, except for $n = 1$. A deeper result is the theorem of Clausen and von Staudt, which implies that the denominator of $B_{2n}$ is square-free and is divisible by $p$ if and only if $p-1$ divides $2n$ [Mahler 1981, p. 291]. Hence $|B_n|_p \le p$ for all $p$ and $|B_n|_p \le 1$ if $p-1$ does not divide $2n$.

If $A = (a_{ij})$ is a matrix with $p$-adic entries, we define $|A|_p = \max |a_{ij}|_p$ and $v_p(A) = \min v_p(a_{ij})$. Clearly $|AB|_p \le |A|_p |B|_p$. Given variables $x_1, \dots, x_n$, the Vandermonde matrix $V(x_1, \dots, x_n)$ is defined as the $n \times n$ matrix whose $(i,j)$-th entry is $x_i^{j-1}$, for $1 \le i, j \le n$. We shall need the following estimate:

**Lemma 2.1.** *The matrix* $V = V(0^2, 1^2, 2^2, \dots, n^2)$ *is invertible, and* $|V^{-1}|_p < p^{2n/(p-1)}$ *for any prime* $p$.

*Proof.* Consider any $V = V(x_1, \dots, x_n)$ where the $x_i$ are distinct. Let $c$ be a column vector with components $c_0, \dots, c_{n-1}$. Then the entries of $Vc$ are $P(x_i)$, where $P(x) = c_0 + c_1 x + \dots + c_{n-1}x^{n-1}$. Thus, if $b$ has components $b_1, \dots, b_n$, we have $Vc = b$ if and only if $P(x)$ is the Lagrange interpolating polynomial defined by $P(x_i) = b_i$. This polynomial is given explicitly by

$$P(x) = \sum_{i=1}^{n} b_i \frac{\prod_{k\ne i}(x_k - x)}{\prod_{k \ne i}(x_k - x_i)}.$$

Thus, the denominator of the $(i,j)$-th entry of $V^{-1}$ is a factor of $\prod_{k\ne i}(x_k - x_i)$.

Specializing to $x_i = i^2$, for $0 \le i \le n$, we find that the denominator of an entry of the $i$-th row of $V^{-1}$ is a factor of $\left|\prod_{k\ne i}(k-i)(k+i)\right|$, which divides $(n-i)!\,(n+i)!$. As is well known, $v_p(n!) = \sum_{k=1}^{\infty}[n/p^k] < n/(p-1)$, and hence

$$v_p((n-i)!\,(n+i)!) < \frac{n-i}{p-1} + \frac{n+i}{p-1} = \frac{2n}{p-1}.$$

Thus each entry of $V^{-1}$ is of the form $a/b$ with $v_p(b) < 2n/(p-1)$ and $v_p(a) \ge 0$, so $|V^{-1}|_p \le p^{2n/(p-1)}$. □

## 3. RECURSIVE GENERATION OF THE SETS $J_P$

The next result, from [Eswarathasan and Levine 1991], is the basis for the recursive construction of the sets $J_p$. Notice that we define $H_0 = 0$.

**Lemma 3.1.** *If* $p > 3$ *is prime*, $n \ge 1$, *and* $0 \le k \le p-1$, *then*

$$H_{pn} = \frac{1}{p}H_n + O(p^2) \tag{3.1}$$

*and*

$$H_{pn+k} = H_{pn} + H_k + O(p). \tag{3.2}$$

*If* $p = 3$, *the first of these congruences holds with* $O(p^2)$ *replaced by* $O(p)$ *and the second holds as stated.*

*Proof.* The difference $H_{pn} - H_n/p$ is the sum of $1/m$ over all $m \le pn$ that are relatively prime to $p$. It can be written as a sum of $n$ sums of the form $\sum_{j=1}^{p-1} 1/(kp+j)$, each of which is $O(p^2)$ by the same argument as for Wolstenholme's theorem. Equation (3.2) is obvious. □

**Proposition 3.2.** *If* $p$ *is an odd prime*, $n \ge 1$ *and* $0 \le k \le p-1$, *then* $v_p(H_{pn+k}) > 0$ *if and only if* $v_p(H_n) > 0$ *and* $H_{pn} = -H_k + O(p)$.

*Proof.* Observe that $v_p(H_k) = 0$ for $1 \le k \le p-1$. Thus, if $k \ne 0$ and $v_p(H_{pn+k}) > 0$, equation (3.2) implies that $H_{pn} = -H_k + O(p)$ and so $v_p(H_{pn}) \ge 0$. Then (3.1) implies $v_p(H_n) > 0$. The case $k = 0$ is obvious. □

**Remark.** This proposition has an amusing interpretation in terms of the expansion of $n$ in base $p$. It says that, if $n = a_0\, a_1 \dots a_m = a_0 + a_1 p + \dots + a_m p^m$ with $0 \le a_i \le p-1$, the condition $v_p(H_n) > 0$

implies that $v_p(H_{n'}) > 0$ for each of the numbers $n' = a_0, a_0a_1, \ldots, a_0 \ldots a_{m-1}$.

As in [Eswarathasan and Levine 1991], Proposition 3.2 gives a method for generating the sets $J_p$. For fixed $p$, let

$$G_m = \{p^{m-1} \le n < p^m : p \text{ divides } H_n\}. \qquad (3.3)$$

So $G_1$ contains $p-1$ and possibly some other $n$. Obviously $J_p = \bigcup_{m\ge1} G_m$. We define $G_0 = \{0\}$ with $H_0 = 0$, and $J_p^0 = J_p \cup \{0\} = \bigcup_{m\ge0} G_m$. Given $G_m$, we can generate $G_{m+1}$ as follows: let $n \in G_m$ with $H_n = ap + O(p^2)$. Then (3.1) and (3.2) give $H_{pn+k} = a + H_k + O(p)$ for $0 \le k \le p-1$. Thus $pn+k$ will be in $G_{m+1}$ if and only if $a + H_k = O(p)$; this can be tested by running through a table of the values of $-H_k \bmod p$. Clearly, $J_p$ is finite if and only if $G_m$ is empty for some $m$. We denote by $M_p$ the smallest $m$ for which $G_m$ is empty (and $M_p = \infty$ if $J_p$ is infinite).

The elements of $J_p^0 = J_p \cup \{0\}$ can be arranged in a tree. The elements of the $m$-th *generation* $G_m$ are the nodes at height $m$. If $m \ge 1$, there is an edge labelled $k \in [0, p-1]$ from $n \in G_m$ with $H_n = ap + O(p^2)$ to $pn+k \in G_{m+1}$ if and only if $H_k \equiv -a$. If $m = 0$ there are edges from the root node $G_0$ to $G_1$ for each $1 \le k \le p-1$ with $H_k \equiv 0$. If we think of this as a family tree, $M_p$ is the *extinction time* of the family.

The set of residues $R = \{H_0 \bmod p, \ldots, H_{p-1} \bmod p\}$ clearly plays an important part in the structure of the tree just described. One obvious property of $R$ is that it is symmetric with respect to $\frac{1}{2}(p-1)$, that is, $H_{p-1-k} \equiv H_k \bmod p$. This follows from Wolstenhome's theorem. Our heuristic probabilistic arguments will be based on the assumption that this is essentially the only general property possessed by $R$.

**Proposition 3.3.** *For any prime $p \ge 3$, the set $J_p$ is finite if and only if $v_p(H_n) \to -\infty$ as $n \to \infty$.*

*Proof.* If $v_p(H_n) \to -\infty$ then $v_p(H_n) \le 0$ for sufficiently large $n$. For such $n$ we have $p \nmid H_n$, hence $J_p$ is finite. On the other hand, if $J_p$ is finite, $G_m$ is empty for sufficiently large $m$. Let $M_p$ denote the smallest such $m$, so $v_p(H_n) \le 0$ for $p^{m-1} \le n < p^m$. Then, by induction, using Proposition 3.2, we see that $v_p(H_n) \le -l$ for all $n$ with $p^{m+l-1} \le n < p^{m+l}$. Thus $\lim_{n\to\infty} v_p(H_n) = -\infty$. $\square$

**Remark.** Denoting $|x|_\infty$ the usual absolute value of a rational $x$, we have $\prod_{p\le\infty} |x|_p = 1$ for any rational $x$ (this is called the product formula). Thus $\prod_p |H_n|_p = 1$ for each $n$, and hence

$$\lim_{n\to\infty} \prod_p |H_n|_p = 1.$$

It is well known that $|H_n|_\infty \to \infty$, and it is easy to see that $|H_n|_2 = 2^k$ for $2^k \le n < 2^{k+1}$. If $J_p$ is finite for each $p \ge 3$, Proposition 3.3 implies that $\lim_{n\to\infty} |H_n|_p = \infty$ for each $p \le \infty$, so $\prod_p \lim_{n\to\infty} |H_n|_p = \infty$. There is no inconsistency here, just a lack of uniform convergence.

## 4. HARMONIC PRIMES

For $p > 3$ prime, we define the *Wolstenholme quotient* $w_p$ to be the integer with $0 \le w_p \le p-1$ such that $H_{p-1} = w_p p^2 + O(p^3)$.

**Proposition 4.1.** *$H_{p^2-p} = w_p p + O(p^2)$ and $H_{p^2-1} = w_p p + O(p^2)$ for any prime $p > 3$, so $J_p$ always contains $p-1$, $p^2-p$ and $p^2-1$. The set $J_p$ consists of exactly these three integers if and only if there are no solutions to $H_k \equiv 0 \bmod p$ and $H_k \equiv -w_p \bmod p$ for $1 \le k \le p-2$.*

In this case we say that $p$ is *harmonic*.

*Proof.* This follows directly from Lemma 3.1. See [Eswarathasan and Levine 1991] or [Graham et al. 1989, pp. 531–532] for details. $\square$

As observed earlier, the set of residues $R = \{H_k \bmod p : 0 \le k \le p-1\}$ is symmetric, so the values of $H_1, \ldots, H_{(p-1)/2} \bmod p$ determine the other $H_k \bmod p$ for $k < p$. According to Proposition 4.1, $p$ is harmonic provided the set of residues $H_1, \ldots, H_{(p-1)/2} \bmod p$ misses the values $0$ and $-w_p$. If we make the heuristic assumption that these residues are independent random integers in

$\{0,\ldots,p-1\}$, the probability that none of $H_k$, for $1 \le k \le \frac{1}{2}(p-1)$, equals 0 or $-w_p$ is

$$\left(\frac{p-2}{p}\right)^{(p-1)/2}.$$

This tends to $1/e = .368\ldots$ as $p \to \infty$, suggesting that the density of harmonic primes is $1/e$. Table 1 contains a survey of the harmonic primes $p < 10^5$ that tends to confirm this conjecture, although the number of harmonic primes in a given interval is perhaps somewhat higher than expected. The computation was done $p$-adically, as we explain in Section 5.

| range | all | harmonic | ratio |
|---|---|---|---|
| $[5,100]$ | 23 | 8 | .348 |
| $[5,1000]$ | 166 | 60 | .361 |
| $[5,10000]$ | 1227 | 447 | .364 |
| $[5,100000]$ | 9590 | 3622 | .378 |
| $(10000,20000]$ | 1033 | 374 | .362 |
| $(20000,30000]$ | 983 | 390 | .397 |
| $(30000,40000]$ | 958 | 356 | .372 |
| $(40000,50000]$ | 930 | 351 | .377 |
| $(50000,60000]$ | 924 | 345 | .373 |
| $(60000,70000]$ | 878 | 354 | .403 |
| $(70000,80000]$ | 902 | 342 | .379 |
| $(80000,90000]$ | 876 | 340 | .388 |
| $(90000,100000]$ | 879 | 323 | .367 |

**TABLE 1.** For each range of values of $p$, we give the number of primes in this range, the number of harmonic primes in this range, and the ratio of these two counts.

The argument in the previous paragraph implicitly assumes that $w_p \ne 0$. An article by Gardiner [1988] considers various equivalent formulations of the condition $w_p = 0$, one of the more interesting being that it is equivalent to $p$ dividing the numerator of the Bernoulli number $B_{p-3}$. There are only two known primes that satisfy this condition, namely $p = 16843$, noticed by Wells Johnson in his computation of irregular primes, and $p = 2124679$, found independently by Richard McIntosh [Guy 1993] and by Buhler et al. [1993]. Therefore it seems safe to ignore this possibility in our heuristic argument. The argument is also not overly sensitive to the fact that $H_1 = 1$ can hardly be considered to be random.

## 5. COMPUTATION OF THE SERIES

Since we wish to consider the factorization of the rational numbers $H_n$, an obvious way to proceed would be to compute the $H_n$ exactly using rational arithmetic and to compute $v_p(H_n)$ by trial division. This cannot succeed for even moderately large $n$, for the following reason: if we write $H_n = a_n/b_n$ in lowest terms, both $a_n$ and $b_n$ grow exponentially with $n$. To see this, note that $b_n \le \operatorname{lcm}(1,2,\ldots,n) \sim e^{(1+o(1))n}$, by the prime number theorem [Hardy and Wright 1960, p. 362]. In the other direction, let $k \ge 2$ and let $p$ be a prime satisfying $n/k < p \le n$. Then the sum of the terms $1/m$ of $H_n$ with $p|m$ is exactly $p^{-1}H_{k-1}$. Thus $p|b_n$ unless $p$ is one of the finitely many primes dividing $H_{k-1}$. So $b_n \ge \prod p$, where the product is over all $p$ with $n/k < p \le n$ that do not divide $H_{k-1}$. Again by the prime number theorem, we have

$$\prod p \sim \exp((1-k^{-1}+o(1))n),$$

and since $k$ is arbitrary, $b_n \sim e^{(1+o(1))n}$. Since $a_n/b_n \sim \log n$, we also have $a_n = e^{(1+o(1))n}$.

Thus, even for $n$ as small as $10^4$, we would need over 4000 digits to represent each of $a_n$ and $b_n$ exactly. Since we will find it necessary to deal with $n$ as large as $397^{100}$, this is clearly not a feasible approach.

### 5.1. First p-adic Method

Since we are concerned with the question of divisibility of $H_n$ by $p$, it is natural to represent $H_n$ $p$-adically. For a given precision $s$ we can represent $H_n$ by the truncated expansion

$$H_n = \sum_{k=v_p(H_n)}^{s-1} a_k p^k + O(p^s), \tag{5.1}$$

where $0 \le a_k < p$ are the $p$-adic digits of $H_n$. The successive terms $H_n$ can be computed from

$H_n = H_{n-1} + 1/n$, and the truncated expansion of $H_n$ can be computed accurately by adding the truncated expansions of $H_{n-1}$ and $1/n$. This is not an exact representation of $H_n$, but, in contrast to the decimal expansion, carries propagate to higher-order digits so round-off error does not accumulate. Provided $s > v_p(H_n)$, the value of $v_p(H_n)$ can be determined accurately. The computational number theory system PARI has an efficient implementation of arithmetic using truncated $p$-adic expansions, which we used throughout our computations.

This naïve or direct $p$-adic method was used to compute the complete set $J_p$ for all $p < 60$ except for $p = 11$. The computation of $H_n$ mod 11 for $n \le 11^9$ ran for several days on a Sun Sparcstation 10. However, for $p = 11$, we must consider $n$ as large as $11^{30}$, so this approach cannot settle even the case $p = 11$.

### 5.2. Second p-adic Method

In order to deal with $H_n$ for $n$ larger than about $10^{10}$, we need the following refinement of the congruence (3.1).

**Theorem 5.2.** *Let $p$ be an odd prime. Then there is a sequence $c_k \in \mathbb{Q}_p$ such that, for all $n \ge 1$,*

$$H_{pn} - \frac{1}{p}H_n = \sum_{k=1}^{\infty} c_k p^{2k} n^{2k}, \tag{5.2}$$

*where the series converges in the $p$-adic norm. The $c_k$ are $p$-adic integers unless $(p-1)|2k$ or $p|k$. In general, $v_p(c_k) = -1 + v_p(1/k)$ if $(p-1)|2k$, and $v_p(c_k) = v_p(1/k)$ otherwise.*

*Proof.* By Bernoulli's formula (2.1),

$$\sum_{k=1}^{m-1} k^{r-1} = \sum_{k=1}^{r} \binom{r-1}{k} B_{r-k} m^k$$

for any $r, m \ge 1$. If we take $r = \varphi(p^{s+1}) = p^s(p-1)$, Euler's formula gives $k^r \equiv 1 \bmod p^{s+1}$ if $(k,p) = 1$, and $k^r = O(p^r)$ if $p|k$. Setting $m = pn$ in the sum above we get

$$\begin{aligned} H_{pn} - \frac{1}{p}H_n &= \sum_{\substack{k=1 \\ (k,p)=1}}^{pn-1} k^{-1} = \sum_{k=1}^{pn-1} k^{r-1} + O(p^s) \\ &= \sum_{k=1}^{r} \frac{1}{r}\binom{r}{k} B_{r-k}(pn)^k + O(p^s) \\ &= \sum_{k=1}^{r} (-1)^{k-1}\frac{1}{k} B_{r-k} p^k n^k + O(p^s), \end{aligned} \tag{5.3}$$

where we have used the identity

$$\frac{1}{r}\binom{r}{k} = \frac{1}{k}\prod_j \frac{r-j}{j} = \frac{1}{k}(-1)^{k-1} + O(p^s).$$

Except for $k = r-1$, the terms with $k$ odd in this sum vanish. Let $c_k(s) = -B_{r-2k}/2k$, for $k \ge 1$. Then, by the theorem of Clausen and von Staudt, $c_k(s)$ is a $p$-adic integer unless $(p-1)|2k$ or $p|k$. Clearly $v_p(c_k) = -1 + v_p(1/k)$ if $(p-1)|2k$ and $v_p(c_k) = v_p(1/k)$ otherwise. Thus, retaining only terms in the final sum in (5.3) that are of order less than $O(p^s)$, we have, for $n \ge 1$,

$$H_{pn} - \frac{1}{p}H_n = \sum_{k=1}^{N} c_k(s) p^{2k} n^{2k} + O(p^s), \tag{5.4}$$

where $N = \frac{1}{2}(s+l)$ for $l \sim \log s/\log p$.

We wish to let $s \to \infty$ in (5.4). For convenience, let $c_0(s) = 0$. Then, from (5.4), we have

$$\sum_{k=0}^{N} (c_k(s+1) - c_k(s)) p^{2k} n^{2k} = O(p^s).$$

We can solve these $N+1$ equations with $0 \le n \le N$ for the coefficients $(c_k(s+1) - c_k(s))p^{2k}$. The matrix of this system of equations is the Vandermonde $V = V(0^2, 1^2, \ldots, N^2)$. By Lemma 2.1 we have $v_p(V^{-1}) > -2N/(p-1)$, so we obtain

$$(c_k(s+1) - c_k(s))p^{2k} = O(p^{s-2N/(p-1)}).$$

Hence, for each $k$, the $p$-adic limit $\lim_{s\to\infty} c_k(s) = c_k$ exists, and

$$c_k - c_k(s) = O(p^{-2k+(\lambda+o(1))s})$$

with $\lambda = 1 - 1/(p-1) > 0$. Replacing $c_k(s)$ by $c_k$ in (5.4) changes the error from $O(p^s)$ to $O(p^{\lambda s})$. Letting $s \to \infty$ yields (5.2). The final statements about $c_k$ follow from the corresponding results for $c_k(s)$. □

**Remark 1.** For $n = 1$ and $p > 3$, we have $H_p - H_1/p = H_{p-1} = w_p p^2 + O(p^3)$, by the definition of the Wolstenholme quotient, so $c_1 \equiv w_p \bmod p$. For $p = 3$ we have $v_3(c_1) = -1$ since $(3-1)|2$.

**Remark 2.** The proof of the theorem shows that

$$c_k = \lim_s \frac{B_{\varphi(p^s)-2k}}{2k}.$$

However, this is not an efficient formula for calculating $c_k$. Instead, as in the proof, one should compute $b_n = H_{pn} - H_n/p$ to precision $s > 2Np/(p-1)$ and then solve the linear system

$$\sum_{k=1}^{N} c_k p^{2k} n^{2k} = b_n + O(p^s)$$

for $c_1p^2, \dots, c_N p^{2N}$. Using Lemma 2.1 as in the proof of the theorem we see that $c_k p^{2k}$ is obtained to precision $s - 2N/(p-1) \geq 2N$.

For example, if $p = 3$ one obtains

$$\begin{aligned}
c_1 &= 2\,3^{-1} + 1 + 3^4 + O(3^5),\\
c_2 &= 3^{-1} + 2 + 3 + 3^3 + 2\,3^4 + O(3^5),\\
c_3 &= 2\,3^{-2} + 2\,3^{-1} + 2 + 2\,3 + 3^4 + O(3^5),\\
c_4 &= 2\,3^{-1} + 2 + 3 + 3^2 + 2\,3^3 + O(3^5)
\end{aligned}$$

For $p = 5$ we have

$$\begin{aligned}
c_1 &= 3 + 3\,5^2 + 2\,5^3 + 5^4 + O(5^5),\\
c_2 &= 4\,5^{-1} + 4 + 5 + 3\,5^2 + 4\,5^3 + 2\,5^4 + O(5^5),\\
c_3 &= 3 + 4\,5 + 5^2 + 3\,5^4 + O(5^5),\\
c_4 &= 2\,5^{-1} + 2 + 2\,5 + 5^2 + 2\,5^3 + O(5^5),
\end{aligned}$$

all easily derived without the computation of any Bernoulli numbers.

**Remark 3.** For a given $N$, the sum $\sum_{k=1}^{N} c_k p^{2k} n^{2k}$ represents $H_{pn} - H_n/p$ with precision

$$s = \min_{k>N}(v_p(c_k) + 2k),$$

which is typically $2N+2$ and certainly no smaller than $2N + 2 - [\log_p(N+1)]$, which only occurs if $N+1$ is a power of $p$. In our computations, the largest value of $N$ used was $N = 50$, for which $s \geq 101$ for all primes $p > 100$.

Now the method of computation is easily described. Given the prime $p$, one chooses a value of $N$, determines the precision $s$ from Remark 3 and computes the coefficients $c'_k = c_k p^{2k}$ to precision $s$ as explained in Remark 2 above. In the process, one will have computed $H_n$ for $1 \leq n \leq p-1$ to precision at least $s$ and hence will know $G_1$. Once one has computed $G_m$ and $H_n$ for each $n \in G_m$ to a precision $r \leq s$, one computes $G_{m+1}$ as follows: For $n \in G_m$, compute $H_{pn}$ from equation (5.2) to precision $r-1$. Then compute successively $H_{pn+k} = H_{pn+k-1} + 1/(pn+k)$ for $k = 1, \dots, p-1$, thus determining $G_{m+1}$, and $H_n$ for each $n \in G_{m+1}$, to precision $r-1$. Notice that here, in contrast to the method of Section 5.1, the precision decreases by 1 in passing from $m$ to $m+1$, so a given initial precision $s$ will only allow one to compute $G_m$ up to $m = s$. If it turns out that $G_s$ is not empty, one must begin the computation again with a larger value of $N$. In practice we used the values $N = 10$, 20, 30, 40 and 50 in succession. Since one may reduce the precision as $m$ increases, the speed of computation actually increases with increasing $m$ in spite of the compensating increase in the size of $n$.

It should be clear that the computation is in general very much faster than the direct method of Section 5.1 since the number of terms $H_n$ with $p^m \leq n < p^{m+1}$ that are computed to determine $G_{m+1}$ is $p\,|G_m|$ rather than $p^m(p-1)$. For example, the largest value of $|G_m|$ appearing in Table 2 is $|G_{86}| = 228$ for $p = 397$, so in computing $|G_{87}|$ we needed $397 \times 228 = 90516$ terms $H_n$, as compared with $397^{86} \times 396 = 1.24 \times 10^{226}$ terms required

| $p$ | $M_p$ | $\lvert J_p\rvert$ | values of $\lvert G_m\rvert$ for $1 \le m < M_p$ |
|---|---|---|---|
| 3 | 4 | 3 | 1, 1, 1 |
| 7 | 7 | 13 | 1, 2, 4, 2, 3, 1 |
| 11 | 30 | 638 | 3, 8, 10, 11, 18, 38, 24, 26, 27, 26, 35, 39, 33, 40, 40, 40, 32, 39, 47, 34, 20, 10, 12, 6, 4, 4, 4, 5, 3 |
| 19 | 7 | 19 | 1, 2, 4, 3, 6, 3 |
| 29 | 5 | 18 | 3, 8, 5, 2 |
| 31 | 7 | 26 | 1, 2, 4, 6, 8, 5 |
| 37 | 4 | 15 | 3, 4, 8 |
| 43 | 5 | 27 | 3, 8, 8, 8 |
| 47 | 5 | 11 | 1, 2, 4, 4 |
| 53 | 6 | 17 | 3, 4, 4, 4, 2 |
| 59 | 6 | 17 | 1, 2, 4, 6, 4 |
| 61 | 4 | 13 | 3, 6, 4 |
| 71 | 8 | 45 | 1, 2, 8, 8, 10, 9, 7 |
| 83 | * | * | 1, 2, 4, 6, 8, 6, 4, 6, 6, 6, 6, 8, 6, 8, 12, 10, 10, 12, 10, 11, 10, 18, 20, 24, 31, 30, 25, 21, 26, 25, 26, 26, 27, 28, 38, 43, 51, 54, 72, 62, 59, 66, 65, 66, 58, 56, 48, 54, 68, 77, 60, 51, 60, 49, 65, 72, 79, 70, 66, 71, 74, 77, 79, 72, 75, 80, 91, 86, 87, 77, 81, 89, 92, 80, 68, 72, 64, 60, 64, 78, 90, 117, 101, 94, 108, 118, 114, 100, 102, 96, 108, 113, 108, 125, 147, 155, 141, 163, 171, 173, ... |
| 89 | 4 | 7 | 1, 2, 4 |
| 97 | 11 | 74 | 3, 6, 6, 14, 6, 6, 10, 8, 12, 3 |
| 101 | 10 | 44 | 1, 2, 8, 7, 4, 6, 6, 6, 4 |
| 103 | 14 | 63 | 1, 2, 4, 6, 4, 8, 10, 8, 4, 4, 6, 4, 2 |
| 109 | 47 | 1273 | 7, 18, 22, 16, 14, 20, 38, 57, 48, 58, 58, 48, 39, 34, 39, 46, 32, 38, 37, 42, 40, 24, 36, 50, 44, 36, 35, 42, 40, 32, 22, 20, 18, 17, 12, 12, 18, 16, 12, 6, 4, 8, 8, 6, 2, 2 |
| 127 | * | * | 1, 2, 4, 6, 4, 4, 8, 10, 12, 14, 18, 25, 18, 16, 18, 30, 35, 38, 36, 26, 34, 36, 30, 53, 46, 36, 24, 26, 25, 30, 34, 26, 34, 26, 26, 18, 28, 34, 35, 46, 40, 40, 42, 40, 32, 25, 26, 21, 19, 22, 20, 16, 10, 10, 14, 19, 16, 16, 16, 19, 26, 37, 26, 20, 22, 20, 22, 24, 24, 18, 31, 42, 26, 28, 22, 24, 30, 26, 30, 26, 40, 34, 41, 36, 40, 33, 34, 50, 46, 47, 34, 32, 40, 44, 34, 40, 36, 31, 34, 28, ... |
| 131 | 4 | 7 | 1, 2, 4 |
| 137 | 8 | 38 | 3, 6, 6, 4, 10, 4, 5 |
| 151 | 4 | 7 | 1, 2, 4 |
| 163 | 20 | 74 | 1, 2, 4, 5, 2, 4, 4, 6, 6, 6, 6, 4, 2, 4, 4, 4, 4, 4, 2 |
| 167 | 49 | 526 | 1, 2, 4, 2, 1, 2, 8, 8, 18, 14, 12, 10, 10, 8, 10, 14, 12, 8, 6, 8, 14, 8, 8, 10, 12, 16, 22, 19, 14, 20, 23, 23, 26, 14, 10, 8, 10, 6, 7, 10, 6, 4, 4, 10, 18, 20, 16, 10 |
| 173 | 33 | 288 | 3, 6, 6, 7, 10, 10, 14, 14, 15, 16, 14, 10, 10, 12, 6, 10, 12, 16, 16, 14, 14, 8, 4, 4, 8, 7, 10, 4, 2, 2, 2, 2 |
| 181 | 6 | 19 | 1, 2, 8, 4, 4 |
| 197 | 11 | 41 | 1, 2, 4, 4, 4, 4, 6, 6, 6, 4 |
| 199 | 5 | 11 | 3, 4, 2, 2 |
| 211 | 12 | 59 | 1, 2, 4, 2, 2, 8, 6, 8, 12, 8, 6 |
| 227 | 6 | 31 | 5, 10, 8, 4, 4 |
| 229 | 17 | 65 | 1, 2, 4, 4, 4, 2, 2, 4, 4, 4, 4, 2, 4, 6, 10, 8 |
| 233 | 21 | 176 | 1, 2, 4, 8, 6, 15, 12, 14, 10, 16, 12, 16, 8, 8, 6, 8, 12, 12, 4, 2 |
| 257 | 5 | 20 | 3, 8, 7, 2 |
| 269 | 20 | 106 | 3, 6, 6, 6, 4, 4, 4, 6, 12, 11, 6, 4, 6, 6, 6, 6, 4, 2, 4 |
| 271 | 18 | 55 | 3, 4, 2, 2, 2, 2, 6, 2, 4, 4, 4, 6, 2, 4, 2, 4, 2 |
| 283 | 14 | 89 | 1, 2, 8, 4, 2, 4, 4, 8, 12, 10, 12, 14, 8 |
| 313 | 11 | 79 | 3, 6, 4, 12, 14, 8, 14, 12, 4, 2 |
| 347 | 10 | 47 | 3, 6, 6, 4, 6, 4, 10, 4, 4 |
| 353 | 6 | 21 | 3, 6, 2, 6, 4 |
| 359 | 35 | 253 | 1, 2, 4, 8, 4, 2, 2, 2, 8, 10, 8, 8, 8, 6, 10, 14, 10, 12, 14, 8, 10, 12, 12, 6, 6, 8, 14, 18, 12, 6, 2, 2, 2, 2 |
| 367 | 8 | 29 | 1, 2, 4, 6, 6, 4, 6 |
| 373 | 4 | 7 | 1, 2, 4 |
| 379 | 13 | 79 | 3, 8, 4, 4, 14, 14, 6, 8, 6, 4, 6, 2 |
| 383 | 11 | 41 | 1, 2, 12, 4, 4, 2, 2, 4, 6, 4 |
| 389 | 8 | 19 | 1, 2, 4, 6, 2, 2, 2 |
| 397 | * | * | 3, 6, 2, 4, 4, 8, 14, 4, 6, 12, 14, 17, 16, 14, 24, 30, 40, 38, 38, 33, 40, 44, 40, 42, 50, 52, 42, 44, 44, 58, 52, 50, 28, 24, 26, 34, 30, 26, 24, 34, 40, 29, 30, 42, 30, 44, 38, 48, 60, 86, 86, 66, 68, 80, 63, 60, 56, 68, 78, 60, 56, 46, 50, 70, 68, 72, 82, 74, 105, 90, 94, 94, 130, 111, 78, 85, 82, 93, 116, 135, 151, 184, 180, 208, 180, 228, 223, 197, 156, 131, 144, 126, 152, 184, 158, 162, 140, 120, 118, 126, ... |

**TABLE 2.** Nonharmonic odd primes $p < 550$ (continued on next page).

| $p$ | $M_p$ | $\lvert J_p\rvert$ | values of $\lvert G_m\rvert$ for $1 \le m < M_p$ |
|---|---|---|---|
| 401 | 5 | 13 | $3, 6, 2, 2$ |
| 409 | 4 | 9 | $3, 4, 2$ |
| 419 | 52 | 703 | $1, 2, 4, 4, 8, 10, 10, 12, 18, 22, 30, 22, 38, 25, 29, 20, 24, 22, 16, 10, 18, 12, 12, 18, 12, 10, 12, 12, 10, 8, 14, 14, 12, 6, 10, 16, 20,$ $10, 12, 10, 14, 14, 16, 16, 22, 16, 14, 6, 2, 6, 2$ |
| 421 | 5 | 23 | $3, 10, 6, 4$ |
| 433 | 30 | 205 | $3, 6, 10, 14, 8, 8, 6, 4, 9, 8, 6, 6, 2, 2, 2, 7, 6, 2, 4, 6, 12, 14, 14, 14, 8, 6, 8, 6, 4$ |
| 439 | 11 | 105 | $3, 4, 10, 16, 18, 20, 16, 10, 4, 4$ |
| 457 | 36 | 323 | $1, 2, 4, 4, 8, 16, 10, 18, 34, 28, 18, 10, 4, 10, 18, 14, 14, 20, 10, 8, 8, 8, 4, 6, 8, 4, 2, 2, 2, 2, 4, 4, 6, 8, 4$ |
| 463 | 4 | 7 | $1, 2, 4$ |
| 521 | 61 | 1763 | $3, 4, 8, 8, 16, 18, 16, 30, 32, 28, 22, 38, 58, 61, 60, 68, 72, 72, 62, 72, 60, 50, 53, 54, 70, 46, 32, 32, 40, 26, 26, 28, 30, 26, 18,$ $26, 24, 20, 20, 24, 34, 34, 28, 34, 20, 14, 6, 4, 8, 10, 12, 14, 16, 8, 12, 18, 16, 10, 10, 2$ |
| 523 | 4 | 7 | $1, 2, 4$ |

**TABLE 2.** Nonharmonic odd primes $p < 550$ (continued). For each $p$ we list the total size of the set $J_p$ and the size of its generations $G_m$, as defined in (3.3). The smallest value of $m$ for which $|G_m| = 0$ is $M_p$. An asterisk indicates that $M_p > 100$. For these primes, the sum of the sizes of the $G_m$ up to $m = 100$ is a lower bound for $|J_p|$, equal to 5870 for $p = 83$, to 2713 for $p = 127$, and to 7718 for $p = 397$.

| $p$ | $M_p$ | $\lvert J_p\rvert$ | $\lvert G_m\rvert$ |
|---|---|---|---|
| $5, 13, 17, 23, 41, 67, 73, 79, 107, 113, 139, 149, 157, 179, 191, 193, 223, 239, 241, 251, 263, 277,$ $281, 293, 307, 311, 317, 331, 337, 349, 431, 443, 449, 461, 467, 479, 487, 491, 499, 503, 541, 547$ | 2 | 3 | $1, 2$ |

**TABLE 3.** Harmonic primes $p < 550$.

by the direct method. The individual steps in the more elaborate method based on Theorem 5.2 are slightly more time-consuming than for the method of Section 5.1 since the required precision is higher, but the considerable difference in the number of steps more than compensates for this.

We applied this method to the first 100 odd primes. Computations were done using PARI 1.38 on a Sun Sparcstation 10 with 48 Mbytes of main memory, with precision $2s \le 100$. The results are summarized in Tables 2 and 3.

## 6. A PROBABILISTIC MODEL

### 6.1. Branching Processes

We have seen in Section 3 that the set $J_p^0 = J_p \cup \{0\}$ has the structure of a tree. We say that $n \in G_m$ is a node at height $m$, and that $n$ has type $a$ if $H_n = ap + O(p^2)$. Each node of type $a$ at height $m$ gives birth to $j$ children at height $m+1$, where $j$ is the number of $0 \le k \le p-1$ for which $H_k \equiv -a \bmod p$. The type of the child $pn+k$ is not determined by $a$, but rather by the higher-order digits in the $p$-adic expansions of $H_n$ and $H_k$. If we wish to model the generation of the tree $J_p$ by a random process, it would seem reasonable to regard the type of $n \in J_p$ as being essentially random, at least for large $n$. So we can think of each member of $G_m$ as giving birth to a random number of children in $G_{m+1}$. The probability distribution is determined by the distribution of $H_k \bmod p$: If there are $n_j$ values of $a$ for which $H_k \equiv -a$ has $j$ solutions, the probability that $n \in G_m$ has $j$ children should be given by $p_j = n_j/p$.

An example should make this clear. Let $p = 11$. Then the sequence $-H_k \bmod p$ for $0 \le k \le 10$ is $0, 10, 4, 0, 8, 10, 8, 0, 4, 10, 0$. Thus, an $n$ of type 0 has 4 children, one of type 10 has 3 children, one of type 4 or 8 has 2 children, and one of any of the remaining types has no children. The corresponding probabilities are thus $p_0 = \frac{7}{11}$, $p_2 = \frac{2}{11}$, $p_3 = p_4 = \frac{1}{11}$, and $p_j = 0$ for all other $j$. It is worth observing that the empirical values of $11p_0$, $11p_2$, $11p_3$ and $11p_4$ are 7.052, 2.034, .896 and 1.017, since the $638 = 11 \times 58$ nodes of $J_{11}$ are distributed as follows:

| type $a$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 59 | 56 | 59 | 63 | 61 | 63 | 61 | 63 | 57 | 44 | 52 |

Notice in this example that the expected number of children of a given parent is $\sum_j jp_j = 1$. This is easily seen to be true for all $p$: For each $a = 0, 1, \ldots, p-1$, let $S_a$ denote the set of $0 \le k \le p-1$ for which $H_k \equiv -a$. Then each $k = 0, 1, \ldots, p-1$ appears in exactly one $S_a$, so $\sum_a |S_a| = p$. Clearly $\sum_j jn_j = \sum_a |S_a|$ so $\sum_j j(n_j/p) = 1$.

Therefore, for each $p$, we have defined a simple branching process or Galton–Watson process. Such processes were originally considered by Galton and Watson in modeling the extinction of families, but they also apply to, among other things, bacterial growth and nuclear chain reactions. The standard reference is [Harris 1963].

A Galton–Watson process can be completely described by the probability generating function

$$f(s) = \sum_j p_j s^j,$$

which in our case is a polynomial. Define the iterates of $f$ by $f_0(s) = s$ and $f_{m+1}(s) = f(f_m(s))$, and let $f_m(s) = \sum_j p_{m,j} s^j$. The basic and delightful result of Watson is that $P(|X_m| = j) = p_{m,j}$, where $X_m$ denotes the $m$-th generation. Thus, the probability of extinction by the $m$-th generation is $p_{m,0} = f_m(0)$, and can be found by simply iterating the function $f(s)$ starting at $s = 0$. Also, $E(|X_m|) = f'_m(1) = f'(1)^m$ from the chain rule. The size of $E(|X_1|) = f'(1)$ thus governs the dynamics of the process, the *supercritical* case $E(|X_1|) > 1$ corresponding to exponential growth in the population (of persons, bacteria or neutrons) and the *subcritical* case $E(|X_1|) < 1$ corresponding to exponential decay.

We have $E(|X_1|) = f'(1) = \sum_j jp_j = 1$, as shown above, so $E(|X_m|) = 1$ for all $n$ and hence we have a *critical* process. The generating function $f(s)$ has a single fixed point in $0 \le s \le 1$ at the point $s = 1$ and the curves $t = f(s)$ and $t = s$ are tangent there. Thus it is clear geometrically (from the standard "cobweb" diagram) that $f_m(0) \to 1$ as $m \to \infty$; in other words, the process will become extinct with probability 1.

In fact, it was shown by Kolmogorov in 1938 [Harris 1963, p. 21], that, if we let $V_p = f''(1)$ be the variance of $|X_1|$,

$$P(|X_m| > 0) = 1 - f_m(0) \sim \frac{2}{V_p m}$$

as $m \to \infty$. This shows that $P(|X_m| > 0) \to 0$. If we let $M_p$ be the extinction time, i.e., the minimal value of $m$ so that $|X_m| = 0$ (or $\infty$ if this never occurs), the distribution function of $M_p$ is given by

$$\begin{aligned} P(M_p \le m) &= P(|X_m| = 0) \\ &= f_m(0) \sim 1 - \frac{2}{V_p m}. \end{aligned} \tag{6.1}$$

Thus $M_p < \infty$ with probability $1 = \lim f_m(0)$, but the expected time to extinction,

$$E(M_p) = \sum P(|X_m| > 0),$$

is infinite because the harmonic series diverges.

The process $\{X_m\}$ we have just defined describes the progeny of a single individual: the initial condition is $|X_0| = 1$ with probability one. We will modify this slightly below to incorporate more appropriate initial conditions.

We can regard this process as producing a random tree that is finite with probability one. Our particular tree $J_p^0$ is completely deterministic, but it is a possible outcome of the process just described. It seems more reasonable to suppose that it falls into the set of probability one consisting of the finite trees produced by $\{X_m\}$, rather than into the set of probability zero consisting of the infinite trees so produced. This tends to support the conjecture that $J_p$ should be finite. Since the expected value of $M_p$ is infinite, we should expect as we vary $p$ that we will encounter trees for which $M_p$ is arbitrarily large. The data of Table 2 seem to support this conjecture.

Note that a different branching process has been associated with each $p$, so (6.1) might lead us to expect that primes with a larger variance $V_p$ should tend to have smaller $M_p$. It is true that $p = 83$, with $M_p > 100$, has the comparatively small $V_{83} =$

$1.229\ldots$. The largest value of $V_p$ for $p < 1000$ is attained by the harmonic prime 179, for which $V_{179} = 2.335\ldots$ and $M_{179} = 3$. However, there are many harmonic primes with small variance, and both 109 and 521 have variance greater than 2 (which is the "typical" value, as we will see below), so it does not seem possible to detect a relationship between $V_p$ and $M_p$ from the data of Tables 2 and 3.

Another factor of plausible significance for the size of $M_p$ is $|G_1|$. Indeed, $p = 109$ with $|G_1| = 7$ does turn out to be rather special, having $M_{109} = 47$ and $|J_{109}| = 1273$. However, $p = 227$ with $|G_1| = 5$ has only $M_{227} = 6$ and $|J_{227}| = 31$, while $p = 83$ and 127 with $M_p > 100$ both have $|G_1| = 1$, so it does not seem that this has a significant effect on the size of $M_p$.

### 6.2. A Universal Distribution

Examining the probability distributions for $p < 1000$, one observes that they do not differ much from one another. An argument like that used to guess the density of harmonic primes suggests that for large $p$ the probabilities $p_j$ are well approximated by the following distribution:

$$p_j = \begin{cases} 0 & \text{for } j \text{ odd,} \\ e^{-1/2}\dfrac{2^{-j/2}}{(j/2)!} & \text{for } j \text{ even} \end{cases} \tag{6.2}$$

(that is, a Poisson distribution supported on the even integers). The parity distinction is due to the symmetry about $\frac{1}{2}(p-1)$ of the set of residues

$$\{H_0 \bmod p, \ldots, H_{p-1} \bmod p\}.$$

For these probabilities $p_j$, the generating function is $f(s) = \exp(\frac{1}{2}(s^2-1))$. This distribution has mean 1 and variance 2. See Table 4 for a comparison with observed values.

For the limiting process, whatever the initial conditions, all $|X_m|$ will be even for sufficiently large $m$, but this will not be the case for the processes defined for the individual primes as in the first paragraph of this section. For these, if $j$ is the (odd) number of $0 \le k \le p-1$ with $H_k \equiv H_{(p-1)/2} \bmod p$, then $p_j = 1/p$. For all other odd $i$, $p_i = 0$. Thus the probability that $|X_m|$ is odd will be positive but small. Inspection of Table 2 verifies that most of the entries there are even numbers.

| $j$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| $p_j$ | .607 | 0 | .303 | 0 | .076 | 0 | .013 | 0 | .002 |
| $p = 83$ | .530 | 0 | .434 | .012 | .024 | 0 | 0 | 0 | 0 |
| $p = 499$ | .625 | 0 | .273 | .002 | .078 | 0 | .020 | 0 | .002 |
| $p = 677$ | .606 | .001 | .319 | 0 | .081 | 0 | .007 | 0 | 0 |

**TABLE 4.** Universal probabilities predicted by (6.2) and corresponding observed values for $p = 83$, 499 and 677.

It would be of interest to establish this limiting distribution rigorously from the analytic theory of prime numbers.

### 6.3. Initial Conditions

The branching processes described above assume as initial conditions a single node at height 0. Then $E(|X_m|) = 1$ for all $m$. However, we know that $|G_1| \ge 1$ and $|G_2| \ge 2$, since the nodes $p-1$, $p^2-p$ and $p^2-1$ always appear in the tree $J_p^0$. Therefore $E(|X_1|) = E(|X_2|) = 1$ seem inappropriate. If we regard the four nodes 0, $p-1$, $p^2-p$ and $p^2-1$ as the only nonrandom nodes in the tree, then $J_p^0$ is the union of four random trees, each generated by a branching process starting at the four nodes. For simplicity, assume that the generating function for each of these processes is $f(s) = \exp(\frac{1}{2}(s^2-1))$. Then the generating function for nodes at height $m$ is $g_m(s) = f_{m-2}^2(s)f_{m-1}(s)f_m(s)$, for $m \ge 2$, where $f_m$ is the $m$-th iterate of $f$. For $m = 1$, $g_1(s) = sf(s)$. If $Y_m$ denotes the set of nodes at height $m$ in the resulting tree, we have $E(|Y_1|) = 2$, $E(|Y_2|) = 3$ and $E(|Y_m|) = 4$ for $m \ge 3$. Note that, as we have described it, the process $\{Y_m\}$ does not depend on $p$.

Let $M$ denote the extinction time of this process. Then, as above, $P(M = m) = g_m(0) - g_{m-1}(0)$. It is instructive to compare the observed distribution

| range of $M_p$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11–15 | 16–20 | 21–25 | 26–30 | 31–40 | 41–50 | $>50$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| predicted count of $p$'s | 35.0 | 12.9 | 8.3 | 5.9 | 4.5 | 3.5 | 3.8 | 2.3 | 7.4 | 4.0 | 2.5 | 1.7 | 2.2 | 1.4 | 5.7 |
| observed count of $p$'s | 42 | 11 | 7 | 5 | 3 | 4 | 0 | 2 | 9 | 4 | 1 | 3 | 2 | 2 | 5 |

**TABLE 5.** Distribution of extinction times for the first 100 odd primes. For each range of values of $M_p$, the last row gives the number of $p$ with $3 \le p < 550$ such that $M_p$ lies in that range, and the middle row gives the value predicted from (6.3).

of the 100 values of $M_p$ found in Tables 2 and 3 with the predicted distribution of $M$, which is

$$100(g_m(0) - g_{m-1}(0)). \tag{6.3}$$

This comparison is performed in Table 5.

Note that this model has

$$g_3(0) = .350\ldots < f_1(0)^2 = 1/e = .368\ldots,$$

so that $M = 3$ occurs with probability $< 1/e$. On the other hand, the model of Section 4 predicts that the probability that $p$ is harmonic is $1/e$. Certainly harmonic primes have $M_p = 3$, so our new model predicts a slightly smaller percentage of harmonic primes than the earlier model. The reason for this is that we have ignored the fact that the types of the four nonrandom nodes are known to be 0, 0, $w_p$, and $w_p$. Incorporating this will produce a more complicated but probably more accurate model.

## 7. PREDICTIONS FROM THE PROBABILISTIC MODEL

The model of the preceding section allows us to make precise predictions about the possible size of $J_p$, the extinction time $M_p$ and the distribution of $(n,p)$ for which $v_p(H_n) = k$. In this section, we assume that the $J_p$ are independent samples from the trees produced by the branching process $\{Y_m\}$ described in Sections 6.2 and 6.3. As noted there, this cannot be exactly true since that process produces trees with an even number of nodes at any level $m > 1$, while this is not true for each individual $J_p$. However, we expect that the approximation will be sufficiently good to produce accurate predictions about quantities such as $|J_p|$ and $M_p$.

The size of the random set $J_p$ is $\sum_m |Y_m|$. An asymptotic result of Otter [Harris 1963, p. 32] gives

$$P(|J_p| > r) \sim br^{-1/2}, \tag{7.1}$$

where $b = 4\pi^{-1/2}$. Our results concerning $|J_p|$ will be based on this estimate and an application of the Borel–Cantelli lemmas [Lamperti 1966, pp. 26–27]. Since it would suffice for this to have bounds $c_1 r^{-1/2} \le P(|J_p| > r) \le c_2 r^{-1/2}$, our predictions do not depend in a serious way on the distributions of the $J_p$ being identical.

From (7.1), if $\varepsilon > 0$, we get

$$P(|J_p| > p^2(\log\log p)^{2+2\varepsilon}) = O\left(\frac{1}{p(\log\log p)^{1+\varepsilon}}\right),$$

and since

$$\sum_p \frac{1}{p(\log\log p)^{1+\varepsilon}} < \infty,$$

the first Borel–Cantelli lemma implies that, with probability one, only finitely many inequalities

$$|J_p| > p^2(\log\log p)^{2+2\varepsilon}$$

can hold, and hence that

$$|J_p| = O_\varepsilon(p^2(\log\log p)^{2+\varepsilon}) \tag{7.2}$$

with probability one. As we decided earlier to ignore sets of probability zero, we thus conjecture that this equation holds for all $p$.

In the other direction, we have

$$P(|J_p| > (p\log\log p)^2) \sim \frac{b}{p\log\log p}.$$

Since the sum of this series diverges and we are assuming the $J_p$ are independent, the second Borel–Cantelli lemma implies that, with probability one,

$$|J_p| > p^2(\log\log p)^2 \tag{7.3}$$

for infinitely many primes $p$.

Now we make some heuristic deductions about the possibilities for $v_p(H_n)$. Note that $v_p(H_n) \geq k$ is equivalent to the vanishing of the digits $a_0, \ldots, a_{k-1}$ in the $p$-adic expansion of $H_n$. In the set $J_p$, $a_0$ is always 0, so $v_p(H_n) = k$ should occur for about $|J_p|/p^{k-1}$ of the numbers in $J_p$. If $p$ is one of the primes for which (7.3) holds, we have

$$|\{n : v_p(H_n) = 2\}| \gg p(\log\log p)^2,$$

so we should expect the number of occurrences of $v_p(H_n) = 2$ to be large, as was indeed observed in our computations.

If $k = 3$ and $p$ is one of the primes for which (7.3) holds, we have

$$|\{n : v_p(H_n) = 3\}| \gg (\log\log p)^2,$$

so we should expect there to be primes for which the number of occurrences of $v_p(H_n) = 3$ is arbitrarily large. Note, however, that the maximum rate of growth from (7.2) is $(\log\log p)^{2+\varepsilon}$. As described in Section 1, our computation revealed only 5 pairs $(n,p)$ with $v_p(H_n) = 3$ for the primes $p < 550$, which is consistent with these predictions.

Next, for $k \geq 5$, the estimate (7.2) shows that

$$|\{(n,p) : v_p(H_n) \geq 5\}| \leq \sum_p \frac{(\log\log p)^3}{p^2} < \infty,$$

so there are only a finite number of such pairs. It seems reasonable to conjecture that there are none.

The remaining case, $k = 4$, is slightly more delicate. We use Kolmogorov's three series theorem [Lamperti 1996, p. 34] to show that $\sum_p |J_p|/p^3$ converges with probability one. Writing $T_p = |J_p|/p^3$, one must check, for some $c > 0$, that the following three series converge: $\sum_p P(T_p > c)$, $\sum_p E(T_p^{(c)})$, and $\sum_p \operatorname{var} T_p^{(c)}$. Here $T_p^c = T_p$ if $T_p \leq c$ and 0 otherwise. Since (7.1) implies that

$$P(T_p > c) \sim \frac{b}{c^{1/2}p^{3/2}},$$

the convergence of the three series is clear for every $c > 0$ from the convergence of $\sum_p 1/p^{3/2}$. Thus, with probability one, there are only a finite number of $(n,p)$ with $v_p(H_n) = 4$. Again we conjecture that there are none.

One can treat the extinction time $M_p$ in a similar way. Since here $P(M_p > r) \sim 4/r$, the analogues to (7.2) and (7.3) are that $M_p = O(p(\log\log p)^{1+\varepsilon})$ for all $p$ and that there should be infinitely many $p$ with $M_p > p\log\log p$. This also follows from the observation that

$$E(|Y_m| \mid |Y_m| \neq 0) = E(|Y_m|)/P(|Y_m| = 0) \sim m$$

and that $|J_p| = \sum_{m=1}^{M_p} |Y_m|$, so $|J_p| \sim \frac{1}{2}M_p^2$.

In conclusion, the model described in Section 6 leads to precise predictions about the size of the $J_p$ and the possible values for $v_p(H_n)$. The agreement of the model with the computations summarized in Tables 2 and 3 seems good enough to give some confidence in these predictions. The goal now is to find a rigorous proof of some of these results.

## REFERENCES

[Batut et al. 1993] C. Batut, D. Bernardi, H. Cohen and M. Olivier, *User's Guide to Pari-GP*. This manual is part of the program distribution, available by anonymous ftp from the host pari@ceremab.u-bordeaux.fr.

[Bernoulli 1713] J. Bernoulli, *Ars Conjectandi*, Basel, 1713. Reprinted in *Die Werke von Jakob Bernoulli*, Birkhäuser, Basel, 1975, vol. 3, pp. 163–167. English translation in D. E. Smith, *A Source Book in Mathematics*, McGraw-Hill, New York, 1929; reprinted by Dover, New York, 1959, pp. 85–90.

[Buhler et al. 1993] J. Buhler, R. Crandall, R. Ernvall, and T. Metsänkylä, "Irregular primes and cyclotomic invariants to four million", *Math. Comp.* **61** (1993), 151–153.

[Dickson 1952] L. E. Dickson, *History of the Theory of Numbers*, vol. 1, Chelsea, New York, 1952.

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

[Euler 1738] L. Euler, "Methodus generalis summandi progressiones", *Comment. Acad. Sci. Petrop.* **6** (1738), 68–97.

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

[Glaisher 1901] J. W. L. Glaisher, "A general congruence theorem relating to the Bernoullian function", *Proc. London Math. Soc.* **33** (1900/01), 27–56.

[Guy 1993] R. K. Guy, "A quarter century of *Monthly* unsolved problems, 1969–1993", *American Math. Monthly*, **100** (1993), 945–949.

[Graham et al. 1989] R. L. Graham, D. E. Knuth and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, MA, 1989.

[Hardy and Wright 1960] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 4th ed., Oxford University Press, Oxford, 1960.

[Harris 1963] T. E. Harris, *The Theory of Branching Processes*, Springer, Berlin, 1963.

[Lamperti 1966] John Lamperti, *Probability: A Survey of the Mathematical Theory*, Benjamin, New York, 1966.

[Mahler 1981] K. Mahler, *p-adic Numbers and Their Functions*, 2nd ed., Cambridge Univ. Press, Cambridge, 1981.

[Ribenboim 1979] P. Ribenboim, 13 *Lectures on Fermat's Last Theorem*, Springer, New York, 1979.

[Sándor 1993] J. Sándor, review of [Eswarathasan and Levine 1991], *Zentralblatt Math.* **764** (1993), 11018.

David W. Boyd, Department of Mathematics, University of British Columbia, Vancouver, Canada V6T 1Z2 (boyd@math.ubc.ca)

Received May 23, 1994; accepted in revised form January 22, 1995