MATHEMATICS OF COMPUTATION
Volume 71, Number 238, Pages 883–908
S 0025-5718(01)01355-2
Article electronically published on October 4, 2001

TWO CONTRADICTORY CONJECTURES CONCERNING CARMICHAEL NUMBERS

ANDREW GRANVILLE AND CARL POMERANCE

Dedicated to the two conjecturers, Paul Erdős and Dan Shanks. We miss them both.[1]

ABSTRACT. Erdős conjectured that there are $x^{1-o(1)}$ Carmichael numbers up to x, whereas Shanks was skeptical as to whether one might even find an x up to which there are more than $\sqrt{x}$ Carmichael numbers. Alford, Granville and Pomerance showed that there are more than $x^{2/7}$ Carmichael numbers up to x, and gave arguments which even convinced Shanks (in person-to-person discussions) that Erdős must be correct. Nonetheless, Shanks's skepticism stemmed from an appropriate analysis of the data available to him (and his reasoning is still borne out by Pinch's extended new data), and so we herein derive conjectures that are consistent with Shanks's observations, while fitting in with the viewpoint of Erdős and the results of Alford, Granville and Pomerance.

1. INTRODUCTION

Fermat's "little" theorem asserts that

$$a^n \equiv a \bmod n, \tag{1}$$

whenever n is prime. If (1) holds for a composite integer n then we call n a *pseudoprime to base a*. If a composite number n is a pseudoprime to every base a, then we call n a *Carmichael number*. One can identify Carmichael numbers fairly easily by using

Korselt's criterion (1899). *A composite odd number n is a Carmichael number if and only if n is squarefree and $p-1$ divides $n-1$ for every prime p dividing n.*

The smallest Carmichael number, 561 ($= 3 \times 11 \times 17$), was found by Carmichael in 1910. It was recently shown that there are infinitely many Carmichael numbers; in fact, that there are more than $x^{2/7}$ Carmichael numbers up to x, once x is sufficiently large (see [2]). Moreover, under certain (widely-believed) assumptions about the distribution of primes in arithmetic progressions, it is shown in Theorem 4 of [2] that there are $x^{1-o(1)}$ Carmichael numbers up to x, as had been conjectured by Erdős [8]. However, for $x = 10^n$ for n up to 16 (which is as far as has been computed [15]), there are less than $x^{0.337}$ Carmichael numbers up to x and, extrapolating

Received by the editor November 11, 1999 and, in revised form, July 25, 2000.

2000 *Mathematics Subject Classification.* Primary 11Y35; 11N60; Secondary 11N05, 11N37, 11N25, 11Y11.

The first author is a Presidential Faculty Fellow. Both authors were supported, in part, by the National Science Foundation.

[1]Dan Shanks passed away on September 6th, 1996, and Paul Erdős two weeks later.

the data to hand, it seems unlikely that there will be more than $x^{1/2}$ Carmichael numbers up to x for any $x < 10^{100}$.

In this article we are interested in this strange phenomenom, first discussed by Shanks [24]. He showed skepticism of Erdős's conjecture, based on the available data, and because he pointed out that it would be far easier to analyze the reliability of pseudoprime tests if there were very few pseudoprimes (however, the analysis in [6] is suitable for Shanks's requirements) – see Section 4 for details of Shanks's remarks.

So what explains this discrepancy between the computational evidence and the predicted asymptotic behavior, for the count of the number of Carmichael numbers up to x? In this paper we propose a conjecture which at least explains why the count should behave peculiarly. Our conjecture takes account of Shanks's observation that computed Carmichael numbers seem to have significantly fewer prime factors than those predicted by Erdős's heuristic. We separate the Carmichael numbers into two classes, *primitive* and *imprimitive*, suggesting that Shanks's intuition is more appropriately applied to imprimitive Carmichael numbers, while Erdős's thoughts are more appropriately applied to primitive Carmichael numbers, thus partially resolving their contradictory conjectures in a way that makes both of them right. We begin by examining the data made available in [12], [13], making several easy observations and recalling some known facts.

By computing the first few examples one quickly observes that Carmichael numbers seem to all have at least three prime factors. This is easily deduced as a consequence of Korselt's criterion, since if not then $n = pq$ where p and q are distinct primes (since n is squarefree), so that $q = n/p \equiv n/p \cdot p = n \equiv 1 \pmod{p-1}$ implying $p-1|q-1$, and similarly $q-1|p-1$, thus $p = q$ and so giving a contradiction.

Twelve of the thirteen Carmichael numbers up to 60,000 have exactly three prime factors. Eleven of the thirty Carmichael numbers between between 60,000 to 1,000,000 have three prime factors and eighteen have exactly four prime factors. Thus we begin to observe that "typical" Carmichael numbers have more prime factors as the number gets larger, though it is not clear how the count grows with the size of the Carmichael number. Pinch [15] has given a table of Carmichael numbers, which we reprint on the next page, showing how many Carmichael numbers up to 10^m have exactly k prime factors, for each $m \le 16$. The data in the table suggest that "typical" Carmichael numbers tend to have more prime factors as the number gets larger. In fact, the average number of prime factors of a Carmichael number $\le x$ goes from ≈ 3.49 for $x = 10^6$, to ≈ 4.00 for $x = 10^9$, to ≈ 5.04 for $x = 10^{13}$, to ≈ 5.91 for $x = 10^{16}$.

There are essentially two known ways to construct Carmichael numbers. The first, which we discuss in Section 2, studies Carmichael numbers $p_1p_2\dots p_k$ with a given number k of prime factors, where the ratios $p_1 - 1 : p_2 - 1 : \cdots : p_k - 1$ are given. We will show that the Carmichael numbers with k prime factors can be partitioned into "families", each conjecturally infinite, depending on these ratios; and we will show that such families exist for each k. Much of Shanks's analysis stems from such constructions, and from computational upper bounds. We will see that such constructions suggest that $C_k(x)$, the number of Carmichael numbers $\le x$ with exactly $k \ge 3$ prime factors, satisfies

$$C_k(x) \gg_k x^{1/k}/\log^k x, \tag{2}$$

x	$C_3(x)$	$C_4(x)$	$C_5(x)$	$C_6(x)$	$C_7(x)$	$C_8(x)$	$C_9(x)$	$C_{10}(x)$	$C(x)$
10^3	1								1
10^4	7								7
10^5	12	4							16
10^6	23	19	1						43
10^7	47	55	3						105
10^8	84	144	27						255
10^9	172	314	146	14					646
10^{10}	335	619	492	99	2				1547
10^{11}	590	1179	1336	459	41				3605
10^{12}	1000	2102	3156	1714	262	7			8241
10^{13}	1858	3639	7082	5270	1340	89	1		19279
10^{14}	3284	6042	14938	14401	5359	655	27		44706
10^{15}	6083	9938	29282	36907	19210	3622	170		105212
10^{16}	10816	16202	55012	86696	60150	16348	1436	23	246683

$C_k(x)$, the number of Carmichael numbers up to x with exactly k prime factors; $C(x) = C_3(x) + C_4(x) + \ldots$, the total number of Carmichael numbers up to x.

(we understand "suggests" to mean "under the assumption of a suitable conjecture, which will be stated later").

The second method for constructing Carmichael numbers, which we discuss in Section 3, was first developed by Erdős [8], and was the basis of the proof of the infinitude of Carmichael numbers [2]. The idea is to first rewrite "$p-1|n-1$ for all $p|n$" in Korselt's criterion as "$L|n-1$ where $L := \operatorname{lcm}_{p|n}(p-1)$", and then to focus on the number L. In fact Erdős picks L first, then finds all primes p for which $p-1$ divides L and then tries to find a product of some of those primes which is $\equiv 1 \pmod{L}$. As Erdős showed, this suggests that

$$C(x) = x^{1-o(1)}. \tag{3}$$

Evidently

$$C(x) = C_3(x) + C_4(x) + C_5(x) + \cdots + C_{k(x)}(x), \tag{4}$$

where $k(x)$ is the maximum number of distinct prime factors of any integer $\leq x$. Since $k(x) \ll \log x/\log\log x$, this suggests by (3) that $C_k(x) \geq x^{1-o(1)}$ for some k, $3 \leq k \leq k(x)$. This is substantially larger than the lower bound given in (2), and one might thus believe that the lower bound given in (2) is typically far from the correct number of Carmichael numbers. We however do not think that this is the case for *fixed* k. Instead we conjecture that it is of exactly the correct order of magnitude:

Conjecture 1. *For any given integer $k \geq 3$, there are $x^{1/k+o_k(1)}$ Carmichael numbers up to x with exactly k prime factors.*

(This conjecture, due to the first author, was first stated in print by the second author in [20]. In this article we are trying to explain the reasoning that led to this conjecture, especially since there have now been several papers with partial results towards Conjecture 1; most recently, Balasubramanian and Nagaraj [4] have shown that $C_3(x) \leq x^{5/14+o(1)}$.)

In Theorem 7 we prove $C_k(x) \le x^{2/3+o_k(1)}$, though we would like to improve this to $C_k(x) \le x^{1/2+o_k(1)}$, for each fixed k. Note that our Conjecture 1 implies, for x sufficiently large,

$$C_3(x) > C_4(x) > C_5(x) > \cdots > C_k(x);$$

however, this is obviously not borne out by the data so far. One, perhaps attainable, objective is to show that $C_4(x) < x^{1/3-\delta}$ for some $\delta > 0$, for all sufficiently large x, which would suggest that at least the first inequality here is correct. To get this estimate we will need to show that there are few Carmichael numbers with four prime factors, whose smallest prime factor is given (in fact, we guess that there are finitely many — this is true for Carmichael numbers with three prime factors).

Evidently we need to understand how Conjecture 1 can possibly make sense if (3) and (4) hold. The point is that the estimate in our conjecture, if true, cannot hold with much uniformity, so that one may have $C_k(x)$ bigger than, say, $\sqrt{x}$ but only with $k \to \infty$ as $x \to \infty$. In particular, an immediate modification of Erdős's heuristic (see Section 3) leads us to the following:

Conjecture 2a. *For any fixed $0 < \nu < 1$, we have, for all integers k with $k = \log^{\nu+o(1)} x$, that there are $x^{\nu+o(1)}$ Carmichael numbers up to x with exactly k prime factors, once x is sufficiently large.*

This implies (3). In Section 3 we refine this argument so as to also conjecture appropriate estimates when $\nu = 1$.

Conjecture 2b. *If integer $k = \delta \log x/\log\log x$ with $0 < \delta \le 1$, then $C_k(x) = x^{1-\delta+o(1)}$.*

We also believe that there are few Carmichael numbers in the missing range:

Conjecture 2c. *If $k \to \infty$ as $x \to \infty$, but $k = \log^{o(1)} x$, then $C_k(x) = x^{o(1)}$.*

Combining Conjectures 1 and 2, we can draw the graph of $\log C_k(x)/\log x$ as k varies: At first (Conjecture 1) it decreases like $1/k$, until it gets to $o(1)$; it then looks like $\log k/\log\log x$ (Conjectures 2a and 2c), until k is very close to the end of its domain, when it rapidly drops to 0 (Conjecture 2b). We expect the minimum to occur with $k \approx \log\log x$. Thus Carmichael numbers with k prime factors are distributed very differently from the integers with k prime factors; for in that case the maximum occurs with $k \approx \log\log x$, the peak of a "Bell curve" (see [9]).

In Section 7 we give a heuristic to estimate $C_k(x)$ for different values of k. This leads to the following single formula which implies Conjectures 1 and 2:

Conjecture 3. *If k is an integer in the range $3 \le k \le y := \log x/\log\log x$, then*

$$C_k(x) = \frac{x^{1/k}}{k!}\ k^y\ (\log\log x)^{O(y)}.$$

In [17] a heuristic argument is given that

$$C(x) = x^{1-\epsilon(x)},$$

where $\epsilon(x) = (1+o(1))\log\log\log x/\log\log x$. Further, it is seen from the argument in [17] that $C_k(x)$ is of similar magnitude when $k \sim \log x/(\log\log x)^2$. Conjecture 3 is not strong enough to imply these results, but see Corollary 5 in Section 7 (which is conditional on Conjecture 4 in that section), which is strong enough.

We now give an overview as to how we justify making these conjectures, in light of the work of Erdős and Shanks. Consider Carmichael numbers n with the ratio of the $p-1$'s fixed, for primes p dividing n. By Korselt's criterion we know that n is squarefree, so we can write $n = p_1 p_2 \dots p_k$. Let $g = g(n) := \gcd(p_1 - 1, p_2 - 1, \dots, p_k - 1)$, and write $p_i - 1 = g a_i$ for some integer a_i. Finally, let $\lambda = \lambda(n) := \operatorname{lcm}[p_1 - 1, p_2 - 1, \dots, p_k - 1] = g[a_1, \dots, a_k]$; this is, in fact, the order of the largest cyclic subgroup of $(\mathbb{Z}/n\mathbb{Z})^*$.

Korselt's criterion states that n is a Carmichael number if and only if

$$\frac{1}{g}\left(\prod_{i=1}^{k}(ga_i + 1) - 1\right) \equiv 0 \pmod{[a_1, \dots, a_k]}.$$

Since the left side is a polynomial in g, we see that the congruence is satisfied for g if and only if it is satisfied for the least positive residue of $g \pmod{[a_1, \dots, a_k]}$. For example, if $a_i = i$ and $k = 3$, then

$$-g \equiv 6g^2 + 11g + 6 = \frac{1}{g}\left(\prod_{i=1}^{k}(gi + 1) - 1\right) \equiv 0 \pmod 6,$$

which works only for $g \equiv 6 \pmod 6$. If n is a Carmichael number and g is itself a least positive residue $(\mathrm{mod}[a_1, \dots, a_k])$, then we call n a *primitive* Carmichael number; otherwise n is *imprimitive*. In our example, $7 \times 13 \times 19$ is a primitive Carmichael number, whereas the next Carmichael number with these ratios, $37 \times 73 \times 109$, is imprimitive. Note that Carmichael number n is primitive exactly when $g(n) \le [a_1, \dots, a_k]$ or, equivalently, $g(n) \le \lambda(n)^{1/2}$.

We believe that the main reason behind the very different conclusions reached by Erdős and Shanks is that most Carmichael numbers are, in fact, primitive, whereas most Carmichael numbers with a fixed number of prime factors, such as those mostly found in computations, are imprimitive. Our conjectures suggest that *most* Carmichael numbers have $(\log x)^{1-o(1)}$ distinct prime factors, while Theorem 3 implies that there are $x^{o(1)}$ such imprimitive Carmichael numbers, and Theorem 4 that there are no imprimitive Carmichael numbers with $\gg \log x / \log\log x \log\log\log x$ prime factors.

Actually Corollary 3 gives a strong version of the upper bound implicit in the analogue of Conjecture 1 for imprimitive Carmichael numbers; that is, $C_k^0(x) \le x^{1/k+o(1)}/k!$ uniformly, where $C_k^0(x)$ denotes the number of imprimitive Carmichael numbers $\le x$ which have exactly k prime factors. We conjecture that $C_k^0(x) \sim C_k(x)$ for each fixed $k \ge 3$; we do *not* conjecture that this holds uniformly in k.

In Corollary 4 we establish that $C^0(x) \ll x^{1/3}/\log^3 x$, where $C^0(x)$ denotes the total number of imprimitive Carmichael numbers $\le x$; so, if there are $x^{1-o(1)}$ Carmichael numbers up to x, as we believe, then we see that very few of them are imprimitive. This supports our conjecture that $C^0(x) = o(C(x))$. Moreover Theorems 3 and 5b together suggest that $C_k^0(x) = o(C_k(x))$ if $k \gg \log\log x$.

Using Richard Pinch's data [14], [15], [16] and unpublished calculations of Chick and Davies and of Williams, we can see how these last two conjectures compare with the known Carmichael numbers. For $x = 10^6, 10^7, \dots, 10^{16}$ we write the number of imprimitive Carmichael numbers up to x as a percentage of the total number of Carmichael numbers up to x; and also do the same thing for Carmichael numbers with exactly three prime factors going up to highest limit computed so far of $x = 10^{20}$.

x	$C_3^0(x)$	$C_3(x)$	%age	$C^0(x)$	$C(x)$	%age
10^6	4	23	17.4	4	43	9.3
10^7	11	47	23.4	11	105	10.5
10^8	25	84	29.8	25	255	9.8
10^9	59	172	34.3	63	646	9.8
10^{10}	127	335	37.9	134	1547	8.7
10^{11}	252	590	42.7	268	3605	7.4
10^{12}	471	1000	47.1	508	8241	6.2
10^{13}	928	1858	49.9	1023	19279	5.3
10^{14}	1734	3284	52.8	1911	44706	4.3
10^{15}	3462	6083	56.9	3783	105212	3.6
10^{16}	6615	10816	61.2	7218	246683	2.9
10^{17}	12725	19539	65.1			
10^{18}	24396	35586	68.6			
10^{19}	46877	65309	71.8			
10^{20}	89751	120459	74.5			

These data do lend credence to our conjectures that $C_3^0(x) \sim C_3(x)$ (that is, $C_3^0(x)/C_3(x) \to 1$ as $x \to \infty$) and $C^0(x) = o(C(x))$ (that is, $C^0(x)/C(x) \to 0$ as $x \to \infty$).

2. Constructing Carmichael numbers with exactly $k \geq 3$ prime factors

Chernick (1939). *If* $6m+1$, $12m+1$, $18m+1$ *are all prime, then their product is a Carmichael number.*

This produces Carmichael numbers for $m = 1, 6, 35, 45, 51, 55, 56$ and 100 of the integers $m \leq 100$. One might ask how frequently this can happen as m grows larger. Prime triplets such as these were considered back into the last century by Sylvester and Dickson, and although little is yet proved we do now have a good *conjectural* understanding of how often these are prime, thanks to Hardy and Littlewood [12] and Schinzel and Sierpinski [22], [23]:

Prime Triplets Conjecture. *Let* a_1t+b_1, a_2t+b_2 *and* a_3t+b_3 *be distinct linear polynomials, with integer coefficients, where each* a_i *is positive and coprime to* b_i. *If there is an integer* r *such that* $(a_1r+b_1)(a_2r+b_2)(a_3r+b_3)$ *is coprime to* 6, *then*

$$\#\{m \leq x : a_1m+b_1, a_2m+b_2, a_3m+b_3 \text{ are all prime}\} \sim \kappa\tau \frac{x}{\log^3 x},$$

where τ *is some absolute positive constant, and* $\kappa = \kappa_{a_1,b_1,a_2,b_2,a_3,b_3}$ *is a rational number satisfying* $1 < \kappa \ll (n/\phi(n))^3$, *with*

$$n = a_1a_2a_3|(a_1b_2 - a_2b_1)(a_1b_3 - a_3b_1)(a_2b_3 - a_3b_2)|.$$

Can we generalize the Chernick construction? We take the perspective that Chernick's construction comes from a *family* of Carmichael numbers, generated by $7 \times 13 \times 19$. If we look at the first few Carmichael numbers, $3 \times 11 \times 17$ and $5 \times 13 \times 17$, and even $7 \times 11 \times 13 \times 41$, we might ask whether each of these also generates a plausible family?

If $3 \times 11 \times 17$ comes from a family, then it must be of the form

$$(2g+1)(10g+1)(16g+1).$$

If these factors are all primes, then we can easily verify that Korselt's criterion is satisfied if and only if $g \equiv 1 \pmod{20}$. Thus there are infinitely many Carmichael numbers from this family if there are infinitely many integers m for which $(40m+3), (200m+11)$, and $(320m+17)$ are all simultaneously prime. This is predicted to be so by the Prime Triplets Conjecture.

We may analyze the other two examples analogously and find that $5 \times 13 \times 17$ comes from a family $(4g+1)(12g+1)(16g+1)$ with $g \equiv 1 \pmod 3$, whereas $7 \times 11 \times 13 \times 41$ comes from a family $(6g+1)(10g+1)(12g+1)(40g+1)$ with $g \equiv 1 \pmod{30}$, but now we are asking for 4 linear expressions to be simultaneously prime. Our first result shows that any Carmichael number generates an infinite family of Carmichael numbers provided the full prime k-tuples conjecture is true. (We state this conjecture in a quantitative form, due to Hardy and Littlewood.)

Let $a_1t+b_1, a_2t+b_2, \ldots, a_kt+b_k$ be distinct linear polynomials with integer coefficients, where each a_i is positive and coprime to b_i. Clearly, a necessary condition for the existence of infinitely many integers m such that each a_im+b_i is prime is that for each prime p there should exist at least one integer m such that p does not divide any a_im+b_i, that is, p does not divide $\prod(a_im+b_i)$. For example, if each $b_i = 1$, then the linear polynomials have this condition, since $\prod(a_i0+b_i) = 1$. We call the set of linear polynomials *admissible* if for every prime p there exists an integer m such that p does not divide any a_im+b_i. The prime k-tuples conjecture of L. E. Dickson is that if the set of linear polynomials is admissible, then there are indeed infinitely many integers m such that each a_im+b_i is prime, while the Hardy-Littlewood conjecture is an assertion about the asymptotic distribution of such integers m. For each prime p define $\nu(p)$ to be the number of distinct $m \pmod p$ for which p does divide some a_im+b_i. Note that a set is admissible if and only if $\nu(p) < p$ for all primes p.

Hardy-Littlewood Conjecture. *If $a_1t+b_1, a_2t+b_2, \ldots, a_kt+b_k$ comprise an admissible set of linear polynomials, with each $a_i > 0$, then*

$$\#\{m \le x : \text{ each } a_im+b_i \text{ is prime}\} \sim \left\{\prod_p \left(1-\frac{\nu(p)}{p}\right)\left(1-\frac{1}{p}\right)^{-k}\right\} \frac{x}{\log^k x}.$$

Theorem 1. *Suppose that $n = p_1p_2\ldots p_k$ is a Carmichael number. Then $q_1\ldots q_k$ is a Carmichael number whenever each $q_i := 1+m(p_i-1)$ is prime, where $m \equiv 1 \pmod L$ and $L = \operatorname{lcm}(p_i-1)$.*

Proof. Obviously $N = q_1\ldots q_k$ is squarefree. Also,

$$\operatorname{lcm}(q_i-1) = \operatorname{lcm}(m(p_i-1)) = mL.$$

To show that N is a Carmichael number it is then sufficient to show that $N \equiv 1 \pmod{mL}$. Now clearly each $q_i \equiv 1 \pmod m$, so that $N \equiv 1 \pmod m$. But $(m, L) = 1$, so it is sufficient to show that $N \equiv 1 \pmod L$. But $m \equiv 1 \pmod L$, so that

$$N = \prod(1+m(p_i-1)) \equiv \prod(1+(p_i-1)) = n \equiv 1 \pmod L,$$

since n is a Carmichael number, so the theorem is proved. □

To be able to apply the Hardy-Littlewood Conjecture to the Carmichael number family constructed in Theorem 1, it is necessary to ensure that the set of linear polynomials $(p_i - 1)m + 1$, where $m \equiv 1 \pmod{L}$, is admissible. That is, is the set of polynomials $(p_i - 1)Lt + p_i$ admissible? The product of these polynomials when t is 0 is n, and the product of these polynomials when $t \equiv -1/L \pmod{n}$ is $\equiv 1 \pmod{n}$. Thus no prime always divides the product of the polynomials, and so the set is admissible. Note that for this argument to make sense we need L to be invertible modulo n. However, since n is a Carmichael number, we have $L|n-1$ by Korselt's criterion, so that L is indeed coprime to n.

Thus we can deduce

Corollary 1. *Assume that the Hardy-Littlewood Conjecture holds. If there exists one Carmichael number with k prime factors, then this generates a family of Carmichael numbers with k prime factors, and there are $\gg x^{1/k}/\log^k x$ such Carmichael numbers up to x.*

We may partition the set of all Carmichael numbers into families depending on the set of ratios of $p-1$ for those primes p dividing the Carmichael number. Corollary 1 says that, assuming the Hardy-Littlewood Conjecture, each family is infinite.

The consequence in Corollary 1 rests on there being at least one Carmichael number with exactly k prime factors, which is by no means guaranteed, a priori (though it is known for $3 \le k \le 1,000,000$ [1]). We now construct a family with exactly k prime factors (assuming the Hardy-Littlewood Conjecture holds), for each $k \ge 3$, by modifying an idea of Euclid:

Choose $n \ge 2$ so that $k = 2n-1$ or $2n$. Let $a_i = 2^{i-1}$ for $1 \le i \le n$, and $a_i = 2^{i-1-n}(2^n - 1)$ for $n+1 \le i \le k$. Then $L = \operatorname{lcm}[a_1, \dots, a_k] = 2^{n-1}(2^n - 1)$ and $\sum_i a_i = 2^{k-n}(2^n - 1)$, so that $L | \sum_i a_i$.

We claim that if $q_i = 1 + La_i m$ is prime for each i, $1 \le i \le k$, then $q_1 q_2 \dots q_k$ is a Carmichael number. To verify this using Korselt's criterion, first note that $q_1 q_2 \dots q_k$ is squarefree. Secondly, $q_1 q_2 \dots q_k \equiv 1 + Lm \sum_i a_i \equiv 1 \pmod{mL^2}$, and since each $a_i | L$ we deduce that $q_i - 1 = La_i m | L^2 m | q_1 q_2 \dots q_k - 1$, and therefore Korselt's criterion is satisfied.

This set of linear polynomials $La_i m + 1$ is admissible, and so the Hardy-Littlewood Conjecture implies

Corollary 2. *Assume that the Hardy-Littlewood Conjecture holds. Then for each integer $k \ge 3$ there are $\gg_k x^{1/k}/\log^k x$ Carmichael numbers up to x which have exactly k prime factors.*

3. Constructing Carmichael numbers with many prime factors

In 1956, Erdős [8] showed how to construct Carmichael numbers with very many prime factors. The idea behind his construction was to attack the problem the other way round. Instead of starting with primes p and then studying $L := \operatorname{lcm}(p-1)$, he chose to begin with a highly composite integer L and then consider the set of primes p for which $p-1$ divides L. In fact, if for some subset $p_1, p_2, \dots, p_k$ we have $p_1 p_2 \dots p_k \equiv 1 \pmod{L}$, then $p_1 p_2 \dots p_k$ is a Carmichael number by Korselt's criterion.

Example. If $L = 120$, then $p - 1 | L, p \nmid L$ for $p = 7, 11, 13, 31, 41, 61$. If we sort through all the subsets of this set of primes, we find that $41041 = 7 \cdot 11 \cdot 13 \cdot 41$,

$172081 = 7 \cdot 13 \cdot 31 \cdot 61$, $852841 = 11 \cdot 31 \cdot 41 \cdot 61$ are all $\equiv 1 \pmod{120}$ and so are all Carmichael numbers.

Alford (see [11]) took a large value for L, determined many primes p for which $p-1$ divides L, and then established that there are at least $2^{128}-1$ Carmichael numbers made up from them — this was the inspiration for [2]. Recently Alford and Grantham [1] have modified this construction to show that there exists a Carmichael number with k prime factors for each integer k in the range $3 \le k \le 1,000,000$. (Moreover, they have constructed a Carmichael number N with 125,458 prime factors that is divisible by a Carmichael number n_k with exactly k prime factors, for each k in the range $50 \le k \le 125,000$.)

In [8], Erdős used this construction to try to get, at least heuristically, a lower bound for the number of Carmichael numbers up to x. Making assumptions about

- the proportion of $d|L$ with $d+1=p$ prime,
- the "equidistribution" of products $p_1 \dots p_k \mod L$,

he deduced that $C(x) \gg_\epsilon x^{1-\epsilon}$, where $C(x)$ is the number of Carmichael numbers up to x (or, put another way, $C(x) = x^{1-o(1)}$).

In [2] Alford, Granville and Pomerance modified Erdős's heuristic argument to show that $C(x) \ge x^{2/7}$ once x is sufficiently large. Although this does not help resolve the dispute over the asymptotic behavior of $C(x)$, or even $\log C(x)/\log x$, the following result, also in that paper, does help in this regard.

Theorem 4 of [2]. *Let $\epsilon > 0$. Suppose there is a number x_ϵ such that*

$$\#\{p \le x : p \equiv 1 \pmod{d}\} \ge \frac{\pi(x)}{2\varphi(d)} \tag{5}$$

for all positive integers $d \le x^{1-\epsilon}$, once $x \ge x_\epsilon$. Then there is a number x'_ϵ such that $C(x) \ge x^{1-2\epsilon}$ for all $x \ge x'_\epsilon$. In particular, if such an x_ϵ exists for each $\epsilon > 0$, then $C(x) = x^{1-o(1)}$ for $x \to \infty$.

When the first author discussed our results with Shanks he noted that he certainly believed (5) holds in the range described, that he had far more extensive data on the distribution of primes in arithmetic progressions, and agreed that this result showed that Erdős was surely correct after all. However he reiterated his frequent request (to both authors) for an investigation into the smallest x for which $C(x) > \sqrt{x}$. As Pinch [14], [15] produces more and more data on Carmichael numbers, it becomes clearer that this is a particularly relevant question:

x	10^3	10^4	10^5	10^6	10^7	10^8	10^9
$C(x)$	1	7	16	43	105	255	646
β	0	.21127	.24082	.27224	.28874	.30082	.31225

x	10^{10}	10^{11}	10^{12}	10^{13}	10^{14}	10^{15}	10^{16}
$C(x)$	1547	3605	8241	19279	44706	105212	246683
β	.31895	.32336	.32633	.32962	.33217	.33480	.33700

$C(x) = x^\beta$, the number of Carmichael numbers up to x, as a power of x

The speed of convergence to the asymptotic behavior (of $\log C(x)/\log x$) is evidently agonizingly slow, and one might guess that this is a consequence of some interesting phenomenon.

Let us now review, in detail, Erdős's heuristic so that we can give some justification for (3) and the predicted lower bounds in Conjectures 2a and 2b.

Let L be the least common multiple of the integers up to $\log x/\log\log x$. Let m be arbitrarily large, but fixed, and let S be the set of primes $q \le \log^m x$ for which $q-1$ divides L but q does not. Since the number of integers up to $\log^m x$ which divide L is known to be $\gg_m \log^m x$, it may be reasonable to assume that (see [2]) the number of elements of S is $\gg_m \pi(\log^m x)$. (This is proved for m slightly beyond 3.) Consider now the set T of all squarefree numbers up to x whose prime factors come from S. If $k = [\log x/\log(\log^m x)]$, then every subset of k primes from S corresponds to a number in T. Thus,

$$\#T \ge \binom{\#S}{k} \ge \left(\frac{\#S}{k}\right)^k \ge \left(\frac{c_m \log^m x/\log(\log^m x)}{\log x/\log(\log^m x)}\right)^{[\log x/\log(\log^m x)]}$$
$$\ge \left(c_m \log^{m-1} x\right)^{[\log x/\log(\log^m x)]} = x^{1-1/m+o_m(1)}.$$

Now each number in T is coprime to L, and it may be reasonable to assume that about 1 out of each $\phi(L)$ members of T is congruent to 1 modulo L. Now $L = x^{o(1)}$, so we shall go ahead and conjecture that there are $\ge x^{1-1/m+o_m(1)}$ members of T which are $\equiv 1 \pmod L$. But every composite member n of T which is $\equiv 1 \pmod L$ is a Carmichael number, since it is squarefree, and for each prime factor q of n we have $q-1|L$ and so $q-1|n-1$.

This heuristic argument of Erdős immediately gives (3), and also Conjecture 2b. By a small modification we get Conjecture 2a. We choose $k = \log^\nu x$ and let S be the set of primes $q \le x^{1/k}$ with $q-1|L$, $q \nmid L$. Now the conjecture is that the number of members of S is $\gg N/\log(x^{1/k})$, where N is the number of integers $\le x^{1/k}$ which divide L. Thus, we conjecture that $\#S \ge x^{1/k}/u^{(1+o(1))u}$, where $u = \log(x^{1/k})/\log(\log x/\log\log x)$, see [5]. A calculation like the one above gives Conjecture 2a.

With more care one can optimize the above argument, and conjecture [17], [19]

$$C(x) = x^{1-\{1+o(1)\}\log\log\log x/\log\log x};$$

one can also prove that the implicit upper bound here holds (see [17], [19]). We expect that the same estimate holds for $\pi_2(x)$, the number of base 2 pseudoprimes up to x, see [17].

4. Shanks's objections

We cannot do better than to essentially reproduce Shanks's own words (section 69 of [24]). We have edited out some remarks and changed notation.

Let $\pi_2(x)$ be the number of integers $n \le x$ which are pseudoprimes to base 2. Note that $C(x) \le \pi_2(x)$. Shanks notes that if we want to study how effective a base 2 pseudoprime test is, as a primality test, then we need to study the ratio $\pi_2(x)/\pi(x)$ as $x \to \infty$. He begins by producing a table (which we extend here in the next table with Pinch's published computations [16], and some unpublished computations).

x	$C(x)$	$\pi_2(x)$	$\pi(x)$	$\pi_2(x)/\sqrt{\pi(x)}$
10^3	1	3	168	0.231
10^4	7	22	1229	0.628
10^5	16	78	9592	0.796
10^6	43	245	78498	0.874
10^7	105	750	664579	0.920
10^8	255	2057	5761455	0.857
10^9	646	5597	50847534	0.785
10^{10}	1547	14884	455052511	0.698
10^{11}	3605	38975	4118054813	0.607
10^{12}	8241	101629	37607912018	0.524
10^{13}	19279	264239	346065536839	0.449

The number of Carmichael numbers, 2-pseudoprimes and primes up to x

Shanks notes that Erdős proved that $\pi_2(x)/\pi(x) \to 0$ as $x \to \infty$; that is, almost all integers which satisfy (1) are prime (Pomerance [17] has subsequently shown that $\pi_2(x) \leq x^{1-\{1/2+o(1)\}\log\log\log x/\log\log x}$ and conjectures [19] that $\pi_2(x) \leq x^{1-\{1+o(1)\}\log\log\log x/\log\log x}$). Shanks goes on to write

> "*Erdős has repeatedly conjectured that $\pi_2(x)/x^{1-\epsilon}$ and even $C(x)/x^{1-\epsilon}$ will increase without bound for every positive ϵ. If he is correct, $\pi_2(x)/\sqrt{\pi(x)}$ will stop decreasing at some x and then will increase without bound. What is that x?*
>
> *The matter is of interest. If a 40-digit n is a pseudoprime to base 2, and if $\pi_2(x) < \sqrt{\pi(x)}$ the probability that n is composite is less than 10^{-19}. But if $\pi_2(x)/\sqrt{\pi(x)}$ increases without bound starting at some unknown x, we lose that estimate. Erdős's "conjecture" remains controversial...*"

Shanks goes on to remark that, then, it was not even known that $\pi_2(x)/\log x \to \infty$ as $x \to \infty$; though now we know this and substantially more ([2]). He then proceeds to conjecture that there are more than $x^{1/2-\epsilon}$ base 2 pseudoprimes up to x if x is sufficiently large, and that this should also hold for base a pseudoprimes for any base a. He justifies this by noting that, for every a, if both factors of $n = (6am+1)(12am+1)$ are prime, then n is a pseudoprime to base 3 and to base a "*so that there is little doubt that our conjecture is true*" (here he is assuming that something like the Hardy-Littlewood Conjecture is true). In Section 8 we will discuss further such base 2 pseudoprimes and a recent conjecture by Will Galway which may be compared with our own conjectures.

> "*If $\pi_2(x) < \sqrt{\pi(x)}$ remains true (or nearly true) as $x \to \infty$, then (the Hardy-Littlewood conjecture) shows that that $\pi_2(x)$ is neatly trapped between $\sqrt{x}/\log^2 x$ and $\sqrt{x}/\sqrt{\log x}$. However, there is insufficient evidence to designate $\pi_2(x) < \sqrt{\pi(x)}$ a conjecture, and we are aware of Erdős's opinion. Numbers at infinity are quite different from those that we see down here: the average number of their prime divisors increases as $\log\log x$ and, while that increases very slowly, it increases without bound. People say that Erdős understands these numbers. We do note that the Erdős construction* [8] *that is said to yield so many Carmichael*

numbers is decidely peculiar in that they all are products of primes r_i for which each $r_i - 1$ is squarefree. That is most untypical of the known Carmichael numbers; among the first 300 only three have that character, namely:

$$67 \cdot 331 \cdot 463, \qquad 23 \cdot 43 \cdot 131 \cdot 859, \qquad 131 \cdot 571 \cdot 1871.$$

All told, we regard the Erdős conjecture as an (unlisted) Open Question."

Shanks then goes on to reproduce Chernick's construction and to apply, as we did, the Prime Triplets Conjecture. "*Therefore although it remains unproved that there are infinitely many Carmichael numbers there is little doubt that $C(x)$ increases at least as fast as $\kappa x^{1/3}/\log^3 x$ for some constant κ.*"

Shanks's criticism of Erdős's construction that it is "*decidely peculiar in that they all are products of primes r_i for which each $r_i - 1$ is squarefree*" is a misunderstanding on Shanks's part. Erdős forced this to happen in his construction so as to simplify the analysis; but, as we did in Section 3 above, one can develop Erdős's heuristic without this restriction. The rest of Shanks's analysis is, to our minds, valid, and worthy of further exploration. Certainly all the subsequent data do little to refute Shanks's reasoning, even if the theoretical evidence does.

5A. A heuristic upper bound for $C_3(x)$

The estimate (2) implies the lower bound of Conjecture 1. In this subsection we study upper bounds, based on the ideas of Section 3. We begin by studying the case $k = 3$ and then generalize that argument.

Any squarefree number with exactly three prime factors must be of the form $(1 + ag)(1 + bg)(1 + cg)$, where a, b, c are coprime. To satisfy Korselt's criterion we must have that abc divides $g(ab + ac + bc) + a + b + c$, in which case we see that a, b, c are pairwise coprime. Thus g satisfies

$$\begin{aligned} g &\equiv -(1/b + 1/c) \mod a, \\ g &\equiv -(1/a + 1/c) \mod b, \\ g &\equiv -(1/a + 1/b) \mod c. \end{aligned} \tag{6}$$

Let $g_0 = g_0(a, b, c)$ be the least positive integer satisfying these congruences, so that $g \equiv g_0 \pmod{abc}$. Then

$$C_3(x) \le \#\{a < b < c,\ g : abcg^3 \le x \text{ and } g \equiv g_0(a, b, c) \pmod{abc}\}.$$

We begin by bounding the number of imprimitive Carmichael numbers: If $g > g_0$ then $g > abc$, so $(abc)^4 \le x$ and $g \le (x/abc)^{1/3}$. Therefore the number of such g is $\le x^{1/3}/(abc)^{4/3}$. In total this gives

$$C_3^0(x) \le x^{1/3} \sum_{a,b,c} \frac{1}{(abc)^{4/3}} \le x^{1/3}\zeta(4/3)^3.$$

Therefore

$$C_3(x) \le \#\{a < b < c : abcg_0^3 \le x\} + O(x^{1/3}).$$

To estimate the number of primitive Carmichael numbers with three prime factors we need to get a good lower bound on the least solution g_0 of (6), at least on average. That is, we need some strong "explicit Chinese Remainder Theorem" to make headway. Since no such result is available, we might *assume* that, as a, b, c vary, $\{\frac{g_0(a,b,c)}{abc} : N < abc < 2N\}$ is distributed uniformly in $[0, 1)$. If even roughly

true, this then implies that $C_3(x) \ll x^{1/3}$. Combining this with the arguments of Section 2, we understand why one might be led to Conjecture 1, in the case $k = 3$.

There have been several results which imply good upper bounds for $C_3(x)$. In 1980, Pomerance, Selfridge and Wagstaff [21] showed that $C_3(x) \ll x^{2/3}$. In 1993 Damgård, Landrock and Pomerance [6] gave an explicit estimate of the shape $C_3(x) \ll x^{1/2+o(1)}$. In 1995, S.W. Graham (unpublished) showed that $C_3(x) \ll x^{2/5+o(1)}$, and most recently, in 1997, Balasubramanian and Nagaraj [4] showed that $C_3(x) \ll x^{5/14+o(1)}$. We shall look at this problem in more detail in Section 8.

5B. A heuristic upper bound for $C_k(x)$ for fixed $k \geq 3$

We now generalize the methods at the start of the previous subsection. In Section 3 we saw that every Carmichael number belongs to a (unique) family

$$(1 + a_1 g)(1 + a_2 g) \dots (1 + a_k g), \tag{7}$$

where

$$(a_1, a_2, \dots, a_k) = 1. \tag{8}$$

To study $C_k(x)$ we examine such families with exactly k terms. Things are a little more complicated now than for $k = 3$, since for a family like (7), we can assume (8) but not that the a_i's are pairwise coprime. Moreover the families themselves are not as easy to deal with. Two reasons are:

- Some k-tuples of integers $a_1, a_2, \dots, a_k$, even pairwise coprime integers, do not lead to any g that can pass Korselt's criterion. For example, for the form in (7) arising from the 4-tuple $\{1, 2, 3, 5\}$ to pass Korselt's criterion we would need
$$\begin{aligned} 1 &\equiv (1 + g)(1 + 2g)(1 + 3g)(1 + 5g) \\ &= 1 + 11g + 41g^2 + 61g^3 + 30g^4 \pmod{5g} \end{aligned}$$
for some positive integer g, so that $1 + g + g^2 \equiv 0 \pmod 5$. However, this congruence has no integral solutions.
- Some k-tuples of integers $a_1, a_2, \dots, a_k$, even pairwise coprime integers, lead to more than one arithmetic progression of values of g. For example, for the 5-tuple $\{1, 2, 3, 5, 193\}$ we have that $(1 + g)(1 + 2g)(1 + 3g)(1 + 5g)(1 + 193g)$ satisfies the Korselt criterion (if all the factors are prime), exactly when $g \equiv 1536$ or $3726 \pmod{5790}$.

In our argument to get an upper bound, we ignore these problems by bounding the possible number of arithmetic progressions g can belong to.

For each set of k distinct integers satisfying (8), define A to be the product of the primes dividing $a_1 a_2 \dots a_k$. That is, $A = \gamma(a_1 a_2 \dots a_k)$, where $\gamma(m)$ is the largest squarefree divisor of m. If the integer in (7) is a Carmichael number with each factor prime, then, by (7), it is $\equiv 1 \pmod{a_i g}$ for each i, and thus $\equiv 1 \pmod p$ for any prime p which divides A. Now for any such prime p, there exists an index j such that p does not divide a_j by (8). This implies that the polynomial $(1 + a_1 t)(1 + a_2 t) \dots (1 + a_k t) - 1$ is not identically zero $\pmod p$, and therefore has at most k distinct roots $\pmod p$. (Actually, there are at most $k - 1$ roots, since at least one a_i is divisible by p.) Therefore, by the Chinese Remainder Theorem, g belongs to one of at most $\prod_{p|A} \min\{k, p\} \leq k^{\omega(A)}$ residue classes $\pmod A$. (The function $\omega(m)$ is the number of distinct prime factors of m.) Let S be the set

consisting of the least positive integer in each one of those residue classes. Thus, if (7) is a Carmichael number, then $g \equiv s \pmod{A}$ for some $s \in S$. If $s \in S$ and (7) is a Carmichael number $\le x$ with $g = mA + s$, then we deduce that

$$\begin{aligned} x &\ge (1 + a_1 g)(1 + a_2 g) \dots (1 + a_k g) \\ &\ge (mA + s)^k a_1 a_2 \dots a_k \\ &\ge m^k A^k a_1 a_2 \dots a_k. \end{aligned}$$

Thus, the number of choices for $m \ge 1$ for a given $s \in S$ is $\le (x/a_1 \dots a_k)^{1/k}/A$. However, this neglects the possibility $m = 0$. In the case that $(x/a_1 \dots a_k)^{1/k}/A \ge 1/2$, we can easily allow the possibility $m = 0$ by putting a factor 2 in front of the expression. But if $(x/a_1 \dots a_k)^{1/k}/A$ is small, we will sometimes have $x \ge s^k a_1 \dots a_k$ and sometimes not. We might guess that on average the residue class $s \pmod{A}$ will satisfy the inequality with "probability" $(x/a_1 \dots a_k)^{1/k}/A$. Thus, we believe that

$$C_k(x) \ll \sum_{1 < A \le x^{1/k}} \frac{\mu^2(A) k^{\omega(A)}}{A} \sum_{\substack{a_1 < \dots < a_k \\ \gamma(a_1 \dots a_k) = A}} \left(\frac{x}{a_1 \dots a_k} \right)^{1/k}$$

This leads to

$$\begin{aligned} C_k(x) &\ll x^{1/k} \sum_{A > 1} \frac{\mu^2(A) k^{\omega(A)}}{A} \left(\left(\sum_{p \mid a \Longrightarrow p \mid A} \frac{1}{a^{1/k}} \right)^k - 1 \right) \\ &= x^{1/k} \sum_{A > 1} \frac{\mu^2(A) k^{\omega(A)}}{A} \left(\prod_{p \mid A} (1 - p^{-1/k})^{-k} - 1 \right) \\ &= x^{1/k} \prod_p \left(1 + \frac{k}{p} \left((1 - p^{-1/k})^{-k} - 1 \right) \right). \end{aligned}$$

We write this last product as $P_1 P_2 P_3$, where in P_1 we consider primes $p < e^k$, in P_2 we consider primes p with $e^k < p < k^k$, and in P_3 we consider primes $p > k^k$.

For $p < e^k$, we have $p^{-1/k} < 1 - \frac{\log p}{2k}$. Then $(1 - p^{-1/k})^{-k} < (2k/\log p)^k$, so that $P_1 \le k^{k\pi(e^k)} e^{O(k)}$.

For $e^k < p < k^k$, we have $(1 - p^{-1/k})^{-k} < e^{1.5kp^{-1/k}} < e^k/k$, so that

$$P_2 < \left(\prod_{e^k < p < k^k} (1 + 1/p) \right)^{e^k} = (O(\log k))^{e^k}.$$

For $p > k^k$, we have

$$(1 - p^{-1/k})^{-k} - 1 < e^{1.5kp^{-1/k}} - 1 \ll kp^{-1/k}.$$

Thus,

$$\log P_3 \ll \sum_{p > k^k} k^2 p^{-1-1/k} \ll \frac{k^3}{(k^k)^{1/k} \log(k^k)} = \frac{k}{\log k}.$$

Putting these estimates together, we have that $P_1P_2P_3 \le k^{O(e^k)}$. Thus, we have the heuristic argument that $C_k(x) \ll_k x^{1/k}$, where the implied constant is $\le k^{O(e^k)}$. This implies the upper bound in Conjecture 1.

6. Primitive and imprimitive Carmichael numbers

The heuristic argument in subsection 5b can actually be interpreted as proving a theorem about imprimitive Carmichael numbers:

Theorem 2. *For each integer $k \ge 3$ there is a number c_k such that $C_k^0(x) < c_k x^{1/k}$ for all $x > 0$. Further, $c_k = k^{O(e^k)}$.*

We remark that in fact a stronger theorem is proved in subsection 5b. We have the upper bound $c_k x^{1/k}$ for the number of Carmichael numbers up to x with k prime factors for which $g(n) \ge \gamma(\lambda(n)/g(n))$, where, as before, $\gamma(\ )$ records the largest squarefree divisor of its argument.

Theorem 3. *There is an absolute constant x_0 such that if $x \ge x_0$ and k is any integer ≥ 3, then*

$$C_k^0(x) \le \frac{1}{k!} x^{1/k} e^{\log x/\log\log(x^{1/k})}.$$

Proof. The result follows from Theorem 2 for $k \le 100$, so we may henceforth assume that $k > 100$. Writing a primitive Carmichael number as $p_1 \dots p_k$ with each $p_i - 1 = ga_i$ as in the introduction, we see that $C_k^0(x)$ is at most the number of choices of positive integers $g, a_1, \dots, a_k$ where

$$a_1 < \dots < a_k, \quad g^k a_1 \dots a_k \le x, \quad g > A := [a_1, \dots, a_k],$$

and

$$(ga_1 + 1) \dots (ga_k + 1) \equiv 1 \pmod{g[a_1, \dots, a_k]}. \tag{9}$$

Therefore

$$x \ge g^k a_1 \dots a_k > A^{k+1},$$

so that $A < x^{1/(k+1)}$. Also, $g \le (x/a_1 \dots a_k)^{1/k} \le (x/A)^{1/k}$. If g is a solution to the congruence (9), then

$$\frac{1}{g}\left(\prod (ga_i + 1) - 1\right) \equiv 0 \pmod{\gamma(A)}, \tag{10}$$

where $\gamma(A)$ is the largest squarefree divisor of A. Since $\gamma(A)$ is squarefree and since the expression on the left side of (10) is a polynomial in g of degree $k-1$, the congruence (10) has $\le (k-1)^{\omega(\gamma(A))} = (k-1)^{\omega(A)}$ solutions. For each solution g_0 of (10), the number of integers $g \equiv g_0 \pmod{\gamma(A)}$ with $A < g \le (x/A)^{1/k}$ is $\le (x/A)^{1/k}/\gamma(A)$. Thus,

$$\begin{aligned} C_k^0(x) &\le \sum_{A \le x^{1/(k+1)}} \sum_{\substack{a_1 < \dots < a_k \\ a_i \mid A}} \left(\frac{x}{A}\right)^{1/k} \frac{(k-1)^{\omega(A)}}{\gamma(A)} \\ &\le \frac{1}{k!} x^{1/k} \sum_{A \le x^{1/(k+1)}} \tau(A)^k \frac{(k-1)^{\omega(A)}}{A^{1/k}\gamma(A)} \\ &\le \frac{1}{k!} x^{1/k} \sum_{A \le x^{1/k}} \tau(A)^{k+\lg k} \frac{1}{A^{1/k}\gamma(A)}, \end{aligned}$$

where $\lg k = \log k/\log 2$ is the base 2 logarithm of k. Note that

$$M_k := \max_{A\le x^{1/k}} \tau(A) = 2^{(1+o(1))\log(x^{1/k})/\log\log(x^{1/k})} \tag{11}$$

as $x^{1/k}\to\infty$. We may assume that $k \le (1+o(1))\log x/\log\log x$, since otherwise there are no integers $\le x$ with k distinct prime factors, and so $C_k^0(x)=0$. Thus $x^{1/k} \ge (\log x)^{1+o(1)}$, and so as $x\to\infty$, we have $x^{1/k}\to\infty$.

By the above, we have

$$C_k^0(x) \le \frac{1}{k!} x^{1/k} M_k^{k+\lg k} \sum_{A\le x^{1/k}} \frac{1}{A^{1/k}\gamma(A)}.$$

Now

$$\begin{aligned}\sum_{A\le x^{1/k}} \frac{1}{A^{1/k}\gamma(A)} &\le \sum_{B\le x^{1/k}} \frac{\mu^2(B)}{B^{1+1/k}} \prod_{p\mid B} \frac{p^{1/k}}{p^{1/k}-1} = \sum_{B\le x^{1/k}} \frac{1}{B} \prod_{p\mid B} \frac{1}{p^{1/k}-1} \\ &< \sum_{B\le x^{1/k}} \frac{1}{B} \prod_{p\mid B} \frac{k}{\log p} < 2 \sum_{B\le x^{1/k}} \frac{1}{B} k^{\omega(B)} \ll M_k^{\lg k}\log(x^{1/k}).\end{aligned}$$

We thus conclude, using (11), that

$$C_k^0(x) \ll \frac{1}{k!} x^{1/k} M_k^{k+2\lg k}\log(x^{1/k}) \ll \frac{1}{k!} x^{1/k} 2^{\{4/5+o(1)\}\log x/\log\log(x^{1/k})},$$

since $k>100$, which implies the theorem. □

Note that $\log\log(x^{1/k}) \ge \log\log\log x + o(1)$ since $k \le (1+o(1))\log x/\log\log x$, and so the last displayed equation, along with Theorem 2 for $k\le 100$, implies the following result:

Corollary 3. *There is an absolute constant x_1 such that if $x\ge x_1$ and k is any integer ≥ 3, then*

$$C_k^0(x) \le \frac{1}{k!} x^{1/k} e^{\log x/\log\log\log x}.$$

Let $C^0(x)$ denote the total number of all imprimitive Carmichael numbers $\le x$.

Corollary 4. *For all $x>1$ we have $C^0(x) \ll x^{1/3}/(\log x)^3$.*

Proof. From a sieve argument like that in subsection 8a below, though now noting that the three factors must each be prime, we can improve the bound of subsection 5a to $C_3^0(x) \ll x^{1/3}/(\log x)^3$. But then the result follows, since

$$C^0(x) = C_3^0(x) + \sum_{k=4}^{\infty} C_k^0(x),$$

and $\sum_{k\ge 4} C_k^0(x) \ll x^{1/4} e^{\log x/\log\log(x^{1/4})}$ by Theorem 3. □

It is probably true that $C^0(x) < x^{1/3}$ for all $x>0$, but the above arguments will need some more work to get this. In particular, note that $x^{1/4}e^{\log x/\log\log(x^{1/4})} > x^{1/3}$ for all $x < 10^{282734}$.

We have the following elementary result on the number of prime factors of imprimitive Carmichael numbers:

Theorem 4. *If n is an imprimitive Carmichael number with k prime factors, then*

$$k \leq (\log 2 + o(1)) \log n / (\log \log n \log \log \log n).$$

Proof. Let $n = p_1 \dots p_k$ with each p_i prime with $p_i = ga_i + 1$, where $g = g(n)$. Let $A = [a_1, \dots, a_k]$, so that $n^{1/k} > g > A$. Also $k \leq \tau(A) \leq 2^{\{1+o(1)\} \log A / \log \log A}$, so that $(\log n)/k > \log A > \{1 + o(1)\} \log k \log \log k / \log 2$; and the result follows. □

The proof in [2] that there are infinitely many Carmichael numbers does not distinguish between primitive and imprimitive Carmichael numbers. However, tracing through the proof, it is shown that there are $> x^{2/7}$ Carmichael numbers up to x, and that these Carmichael numbers all have $> \log x / (\log \log x)^{1+\epsilon}$ prime factors, for each fixed $\epsilon > 0$, once x is sufficiently large, depending on the choice of ϵ. Corollary 3 above implies that there are at most $x^{o(1)}$ imprimitive Carmichael numbers up to x with so many prime factors, so it follows that almost all of the Carmichael numbers produced by the proof in [2] are primitive. By making some small changes to the proof in [2], one can show that there is a positive number c such that for all sufficiently large x, there are $> x^{2/7}$ Carmichael numbers up to x with $> c \log x / \log \log x$ prime factors. It thus follows from Theorem 4 that for large x, these Carmichael numbers are all primitive. It is still not proved that there are infinitely many imprimitive Carmichael numbers, though, as in Corollary 1, this follows from the Hardy-Littlewood Conjecture.

Using his data base of Carmichael numbers, Richard Pinch has kindly computed for us counts of primitive and imprimitive Carmichael numbers up to various levels and with various numbers of prime factors.

x	$C_3^0(x)$	$C_4^0(x)$	$C_5^0(x)$	$C^0(x)$
10^6	4			4
10^7	11			11
10^8	25			25
10^9	59	4		63
10^{10}	127	7		134
10^{11}	252	16		268
10^{12}	471	37		508
10^{13}	928	93	2	1023
10^{14}	1734	174	3	1911
10^{15}	3462	312	9	3783
10^{16}	6615	573	30	7218

$C_k^0(x)$, the number of imprimitive Carmichael numbers up to x with exactly k prime factors; $C^0(x)$, the total number of imprimitive Carmichael numbers up to x.

The smallest imprimitive Carmichael number is $294409 = 37 \times 73 \times 109$. There are no imprimitive Carmichael numbers up to 10^{16} with more than 5 prime factors, though Pinch found the imprimitive Carmichael number

$$62411762908817281 = 113 \times 337 \times 449 \times 673 \times 2017 \times 2689$$

with 6 prime factors, and he believes, though he hasn't checked, that this is the only one below 10^{17}.

7. Conjecture 3

Let us now work out a heuristic argument for an estimate of $C_k(x)$, the number of Carmichael numbers (both primitive and imprimitive) with exactly k prime factors. For a composite number n to be a Carmichael number it is necessary and sufficient that n is squarefree and that $n \equiv 1 \pmod{\lambda(n)}$. (Note that if the congruence holds, then n must be squarefree.) However, for every n, we have $n \equiv 1 \pmod{g(n)}$. One might say then that a random squarefree number n is a Carmichael number with "probability" $g(n)/\lambda(n)$. And so we might expect that $C_k(x)$ is approximately the sum of $\mu^2(n)g(n)/\lambda(n)$ for $n \le x$ with n having exactly k prime factors. We throw in an error factor so as to allow a more precise conjecture.

Conjecture 0. *Let $y = \log x/\log\log x$. Then, uniformly for $3 \le k \le y$,*

$$C_k(x) = e^{O(y)} \sum_{\substack{n\le x\\ \omega(n)=k}} \frac{\mu^2(n)g(n)}{\lambda(n)}.$$

The trouble with Conjecture 0 is that it does not seem so easy to estimate the sum. We thus try to "simulate" $\lambda(n)$, which leads us to considerations that are very similar to what we considered above for imprimitive Carmichael numbers.

As before, we will associate to each Carmichael number with k prime factors integers $g, a_1 < \cdots < a_k$, where the k primes are $ga_i + 1$ for $i = 1, \dots, k$. If we further assume that $(a_1, \dots, a_k) = 1$, then the association is well-defined. In studying $C_k^0(x)$ we also assumed that $g > [a_1, \dots, a_k]$, but we cannot assume this in the general case. Let

$$N_k(x) := \sum_{\substack{a_1<a_2<\cdots<a_k\\ a_1a_2\dots a_k\le x\\ \gcd(a_1,a_2,\dots,a_k)=1}} \left(\frac{x}{a_1\dots a_k}\right)^{1/k} \frac{1}{\operatorname{lcm}[a_1,\dots,a_k]},$$

where the a_i's run over positive integers. We will conjecture shortly that $N_k(x)$ is a fairly good approximation for $C_k(x)$. The reasoning goes as follows. To get a Carmichael number bounded by x out of a k-tuple $a_1, \dots, a_k$, we shall need an integer g that satisfies

$$\prod(ga_i + 1) \le x, \quad \prod(ga_i + 1) \equiv 1 \pmod{g[a_1, \dots, a_k]},$$

and each $ga_i + 1$ is prime.

- The inequality is about the same as $g \le (x/a_1 \dots a_k)^{1/k}$.
- The congruence may be rewritten as $g^{-1}(\prod(ga_i + 1) - 1) \equiv 0 \pmod{p^\ell}$, for each prime power p^ℓ exactly dividing $A := [a_1, \dots, a_k]$, and then combining the results by the Chinese remainder theorem. Given $g, a_1, a_2, \dots, a_{k-1}$, there exists $a_k \pmod{p^\ell}$ so that the congruence is satisfied if and only if p does not divide any $1 + ga_i$, and in that case the congruence class $a_k \pmod{p^\ell}$ is unique. Thus we "expect" there to be $\prod_{p|A}\{(1-1/p)^k + 1/p\}$ values of $g \pmod{A}$ satisfying the congruence. In other words the "probability" that the congruence is satisfied is $(1/A)\prod_{p|A}\{(1-1/p)^k + 1/p\} = (\log\log x)^{O(k)}/A$.

Therefore we guess that there are $(x/a_1 \dots a_k)^{1/k}(\log\log x)^{O(k)}/[a_1, \dots, a_k]$ values of $g \le (x/a_1 \dots a_k)^{1/k}$ which satisfy the congruence. Notice that for many choices of $a_1, \dots, a_k$ this expression is < 1. We thus are making the heuristic assumption

that when added, the fractions give a good estimate for the total number of choices for $g, a_1, \dots, a_k$; that is, the estimate is correct on average.

- If we randomly select an integer close to X, then the probability that our selection is prime is around $1/\log X$. We need each $ga_i + 1$ to be prime, and we might suppose, heuristically, that the probability of this happenning is around $1/\prod \log(ga_i)$, which is $\geq 1/\log^k(x^{1/k})$, since the geometric mean of the ga_i is $\leq x^{1/k}$. Note that $\log^k(x^{1/k}) \leq \max\left(e^{\log x/\log\log x}, (\log\log x)^{2k}\right)$ for $3 \leq k \leq \log x/\log\log x$.

Combining these remarks leads to the following "educated guess":

Conjecture 4. *If k is an integer in the range $3 \leq k \leq y := \log x/\log\log x$, then $C_k(x) = N_k(x)e^{O(y+k\log\log\log x)}$ uniformly.*

We will show that this guess, or conjecture, implies Conjecture 3, and thus both Conjecture 1 and all parts of Conjecture 2, by estimating $N_k(x)$ as follows:

Theorem 5. *For $3 \leq k \leq \log\log x$ we have*

$$N_k(x) = \frac{1}{k!}x^{1/k}e^{O\left((\log x)^{0.7}\right)}.$$

For $\log\log x \leq k \leq y$, we have

$$N_k(x) = \frac{1}{k!}x^{\frac{\log k - \log\log(2y/k))}{\log y}}e^{O(y+k\log\log\log x)}.$$

We prove most of Theorem 5 by estimating $N_k(x)$ in terms of

$$L_k(x) := \sum_{\substack{a_1<\cdots<a_k \\ a_1\dots a_k \leq x}} \frac{1}{\operatorname{lcm}[a_1, \dots, a_k]},$$

a function which may be of independent interest. The key observation to link these two functions is

Proposition 1. *For integer k, $3 \leq k \leq y$, we have*

$$x^{1/k}L_k(x) \geq N_k(x) \geq (x/k!)^{1/k}/[1, 2, \dots, k].$$

For $\log\log x \leq k \leq y$, we have $N_k(x) = L_k(x)e^{O(y)}$.

Note that if $3 \leq k \leq \log\log x$, then $(x/k!)^{1/k}/[1, 2, \dots, k] = x^{1/k}e^{O(k)}$, by the prime number theorem, giving the lower bound in the first part of Theorem 5.

Proof. To get the first lower bound, note that $N_k(x)$ is at least the size of the one term in the sum defining $N_k(x)$ which has $a_i = i$ for each i.

Now, $1 \leq (x/a_1 \dots, a_k)^{1/k} \leq x^{1/k}$ for all choices of the a_i's in the sum defining $N_k(x)$, so that

$$\sum_{\substack{a_1<a_2<\cdots<a_k \\ a_1a_2\dots a_k \leq x \\ \gcd(a_1,a_2,\dots,a_k)=1}} \frac{1}{\operatorname{lcm}[a_1, \dots, a_k]} \leq N_k(x) \leq x^{1/k}L_k(x).$$

The final upper bound in our result follows since $x^{1/k} \leq e^y$ when $k \geq \log\log x$.

The difference between the sum in the display above and the sum defining $L_k(x)$ is that in the sum defining $L_k(x)$ we allow the a_i's to have a common factor $g > 1$. This increases the value of the sum by a factor $\leq \sum_{g \leq x} 1/g \ll \log x$, so that our inequality above becomes $N_k(x) \gg L_k(x)/\log x$ for all $k \geq 3$. □

Theorem 5A. *For $3 \le k \le \log\log x$ we have*

$$L_k(x) = \frac{1}{k!} e^{O\left((\log x)^{0.7}\right)}.$$

Proof. For $3 \le k \le \log\log x$ we have

$$k! L_k(x) \le \sum_{n \le x} \frac{1}{n} \left(\sum_{a \mid n} 1 \right)^k = \sum_{n \le x} \frac{\tau(n)^k}{n} \le \prod_{p \le x} \left(1 + \frac{2^k}{p} + \frac{3^k}{p^2} + \dots \right)$$

$$\ll \prod_{p \le x} \left(1 + \frac{1}{p} \right)^{2^k} < (c \log x)^{2^k} = e^{O\left((\log x)^{0.7}\right)},$$

where, when $p < (2-\epsilon)^k$, we have used the inequality

$$\sum (n+1)^k / p^n < k!/(1-1/p)^k.$$

□

Theorem 5B. *For $\log\log x \le k \le y := \log x / \log\log x$, we have*

$$L_k(x) = \frac{1}{k!} x^{\frac{\log k - \log\log(2y/k)}{\log y}} e^{O(y + k \log\log\log x)}.$$

Proof. For the lower bound, let M denote the least common multiple of the integers up to y, so that $M = e^{y+o(y)}$ by the prime number theorem, and therefore

$$L_k(x) \ge \frac{1}{M} \sum_{\substack{a_1 < \dots < a_k \le x^{1/k} \\ a_i \mid M}} 1 \ge \frac{1}{M} \binom{\psi_0(x^{1/k}, y)}{k} = e^{O(y)} \frac{\psi_0(x^{1/k}, y)^k}{k!}$$

provided that $\psi_0(x^{1/k}, y) \ge 2k$, where $\psi_0(z, y)$ is the number of squarefree y-smooth integers $\le z$.

Let $u = \log z / \log y$. We have, uniformly for all y, z with $1 < y < z$, that

$$\psi_0(z, y) \ge z \exp(-u(\log u + \log\log(2u) + O(1))). \tag{13}$$

This inequality is known for the larger function $\psi(z, y)$, the number of all y-smooth integers up to z, see Canfield, Erdős, and Pomerance [5]. Thus (13) follows from Ivić and Tenenbaum [13] in the range $y > (\log z)^3$, since they showed that in this range $\psi_0(z, y) \gg \psi(z, y)$. For the range $\log z \le y \le (\log z)^3$, the inequality (13) follows by estimating the number of $[u]$-element subsets of the set of primes up to y since each such subset corresponds to an integer counted by $\psi_0(z, y)$. The range $y < \log z$ is trivial for (13), since then $z \le \exp(u(\log u + \log\log(2u) + O(1)))$.

Now let $u = \log(x^{1/k})/\log y$, which is $\sim y/k$ in our range, and so

$$ku(\log u + \log\log(2u) + O(1)) = \log x \left(1 - \frac{\log k - \log\log(2y/k)}{\log y} \right) + O(y).$$

Combining this with (13) (with $z = x^{1/k}$) yields

$$N_k(x) \ge \frac{e^{O(y)}}{k!} \psi_0(x^{1/k}, y)^k = \frac{1}{k!} x^{(\log k - \log\log(2y/k))/\log y} e^{O(y)},$$

which is slightly stronger the lower bound in the theorem.

We use Rankin's moment method to find an upper bound on $L_k(x)$. Let

$$\nu = \frac{\log k - \log\log(20y/k) + 1}{\log y}, \quad \text{so that } \frac{1}{\log y} \le \nu < 1 - \frac{0.09}{\log y}$$

in our range, once x is sufficiently large. Let $\tau_k(n)$ denote the number of ordered factorizations of n into k positive factors, so that $\sum_{j\ge 0}\tau_k(p^j)z^j = (1-z)^{-k}$ for any prime p, and $|z|<1$. Recall that $\gamma(n)$ denotes the largest squarefree divisor of n. Since $\nu > 0$ we have

$$L_k(x) \le \frac{1}{k!}\sum_{n\le x}\frac{\tau_k(n)}{\gamma(n)} \le \frac{1}{k!}x^\nu \sum_{n\le x}\frac{\tau_k(n)}{n^\nu\gamma(n)}$$

$$\le \frac{1}{k!}x^\nu \prod_{p\le x}\left(1+\frac{1}{p}\sum_{j=1}^{\infty}\tau_k(p^j)p^{-\nu j}\right) < \frac{1}{k!}x^\nu\prod_{p\le x}\left(1+\frac{1}{p}\left(1-p^{-\nu}\right)^{-k}\right).$$

Now, if p is a prime in the range $y\le p\le x$, then

$$\left(1-p^{-\nu}\right)^{-k} \le \left(1-y^{-\nu}\right)^{-k} = \left(1-\frac{\log(20y/k)}{ek}\right)^{-k}$$

$$\le \exp\left(\frac{2}{e}\log(20y/k)\right) < 20y/k,$$

so that

$$\prod_{y\le p\le x}\left(1+\frac{1}{p}\left(1-p^{-\nu}\right)^{-k}\right) \le \prod_{y\le p\le x}\left(1+\frac{1}{p}\right)^{20y/k} < (\log x)^{20y/k} < e^{20y}.$$

For all $p\le y$ we have $1+\frac{1}{p}(1-p^{-\nu})^{-k} < (1-p^{-\nu})^{-k}$, so that

$$\prod_{p\le y}\left(1+\frac{1}{p}\left(1-p^{-\nu}\right)^{-k}\right) \le \prod_{p\le y}\left(1-p^{-\nu}\right)^{-k}.$$

If $e^{1/2\nu}\le p\le y$ then $p^{-\nu}\le e^{-1/2}$, so that $(1-p^{-\nu})^{-k}\le \exp(O(kp^{-\nu}))$. Therefore

$$\log\prod_{e^{1/2\nu}\le p\le y}\left(1-p^{-\nu}\right)^{-k} \ll \sum_{p\le y}\frac{k}{p^\nu} \ll k\sum_{p\le e^{1/(1-\nu)}}\frac{1}{p} + k\sum_{e^{1/(1-\nu)}<p\le y}\frac{1}{p^\nu}$$

$$\ll k|\log(1-\nu)| + \frac{ky^{1-\nu}}{(1-\nu)\log y} \ll k\log\log y + y$$

by the prime number theorem, since $p^{1-\nu}\le e$ when $p\le e^{1/(1-\nu)}$.

If there are any primes $p\le e^{1/2\nu}$, then $\nu\le 1/\log 4$, so that $k\le(\log x)^{3/4}$. Moreover, $1-p^{-\nu}\le 1-2^{-\nu}<\nu\log 2$, so that

$$\log\prod_{p\le e^{1/2\nu}}\left(1-p^{-\nu}\right)^{-k} \ll k\pi(e^{1/2\nu})\log(1/\nu) \ll ke^{1/2\nu}\nu\log(1/\nu)$$

$$\ll ke^{1/2\nu} \ll (\log x)^{3/4} \ll y,$$

since if $k\ge(\log x)^{1/4}$ then $\nu\asymp 1$, and if $k\le(\log x)^{1/4}$ we use the fact that $e^{1/\nu}\le y$.

Combining these last four displayed inequalities gives

$$\prod_{p\le x}\left(1+\frac{1}{p}\left(1-p^{-\nu}\right)^{-k}\right) \le e^{O(y+k\log\log\log x)},$$

and the upper bound follows. □

We have the immediate conditional corollary formed by combining Conjecture 4 with Theorem 5.

Corollary 5. *Assume Conjecture 4, and let* $y = \log x/\log\log x$. *We have uniformly, for* $3 \le k \le \log\log x$, *that*

$$C_k(x) = \frac{1}{k!}x^{1/k}e^{O(y+k\log\log\log x)}.$$

For $\log\log x \le k \le y$, *we have uniformly that*

$$C_k(x) = \frac{1}{k!}x^{\frac{\log k-\log\log(2y/k))}{\log y}}e^{O(y+k\log\log\log x)}.$$

One can easily deduce Conjecture 3 from Corollary 5, and thus Conjectures 1 and 2.

The error factor $e^{O(k\log\log\log x)}$ in Corollary 5 is swamped by $e^{O(y)}$, the other error factor, in almost the entire range for k. It is only when $k > y/\log\log\log x$ that the more complicated error factor takes over. In fact if $k \ge \epsilon y$, then Corollary 5 implies the conditional result that $C_k(x) = (x/k!)e^{O_\epsilon(y\log\log\log x)}$. However the results in [14] give a rigorous upper bound for $C_k(x)$ which is stronger than our conditional result:

Theorem 6. [14] *If* $3 \le k \le y$, *we have*

$$C_k(x) \le \frac{1}{k!}xe^{O(y)}.$$

8. Carmichael numbers with three prime factors, a more precise conjecture

Conjecture 1 stems from the belief that the vast majority of Carmichael numbers with exactly three prime factors should come from long sequences from families of "prime triplets". That is, most three prime factor Carmichael numbers should be imprimitive; in other words, $C_3^0(x) \sim C_3(x)$. We have seen that to have such a Carmichael number we must have $(ag+1),(bg+1),(cg+1)$ all prime with $g = g_0(a,b,c) + mabc$ and $a < b < c$ pairwise coprime. If the corresponding Carmichael number is $\le x$, then $m \le (x/(abc)^4)^{1/3}$. If $abc < x^{o(1)}$, then the expected number of such triplets is a constant, depending on the arithmetic properties of a,b,c, times $(x/(abc)^4)^{1/3}/\log^3(x^{1/3})$. Note that this constant is precisely predicted by the Hardy-Littlewood Conjecture. Summing these quantities up, and writing $n = abc$, we are led to the more precise conjecture that

$$C_3(x) \sim \kappa_3\lambda\frac{x^{1/3}}{\log^3 x},$$

where

$$\lambda := \frac{243}{2}\prod_{p>3}\left(\frac{1-3/p}{(1-1/p)^3}\right) \approx 77.1727\ldots$$

and

$$\kappa_3 = \sum_{n\ge1}\frac{(n,6)}{n^{4/3}}\prod_{\substack{p\mid n\\ p>3}}\frac{p}{p-3}\sum_{\substack{a<b<c,\ n=abc\\ a,b,c \text{ pairwise coprime}}}\delta_3(a,b,c)\prod_{\substack{p\nmid n\\ p>3}}\frac{p-\omega_{a,b,c}(p)}{p-3},$$

with $\delta_3(a,b,c) = 2$ if $a \equiv b \equiv c \not\equiv 0 \pmod 3$ and 1 otherwise, and $\omega_{a,b,c}(p)$ is the number of distinct residues modulo p represented by a,b,c.

Writing the summand in the sum for κ_3 in the form $f(n)/n^{4/3}$, we evidently see that $f(n) = n^{o(1)}$, so that the sum is convergent. However the sum converges so

slowly that we have found it difficult to determine an accurate estimate for κ_3, so we now discuss a heuristic argument to "guesstimate" κ_3. Note that $f(n)$ is not far from being $3^{\omega(n)}$ times some other factors which should be, on average, constant. Since the average order of $3^{\omega(n)}$ is $\log^2 n$, it therefore seems reasonable that there is a constant α such that

$$\kappa_3 = \kappa_3(N) + (\alpha + o(1)) \int_N^\infty \frac{\log^2 t}{t^{4/3}}\, dt,$$

where $\kappa_3(N)$ is the partial sum over the integers $n \le N$. Thus, with two approximations $\kappa_3(N)$, and assuming the "$o(1)$" is negligible in the above expression, one might infer an extrapolated value for κ_3. Doing this with the approximations

$$\kappa_3(10^7) \approx 24.7875, \quad \kappa_3(2.2 \times 10^7) \approx 25.1801,$$
$$\kappa_3(6 \times 10^7) \approx 25.5882, \quad \kappa_3(8.5 \times 10^7) \approx 25.7092,$$

kindly computed for us by John Chick and Gordon Davies, we conjecture that 27 is a fairly good approximation for κ_3. (With the first two values of $\kappa_3(N)$ above we get 27.125, with the first and third we get 27.113, with the first and last we get 27.109, with the second and third we get 27.106, with the second and last we get 27.103, and with the last two we get 27.095.) A rigorous numerical determination of the value of κ_3 seems quite difficult. But using $\kappa_3 \approx 27$, we would have $\tau_3 := \kappa_3 \lambda \approx 2100$.

To try to numerically verify this conjecture, one should bear in mind that the more precise expression predicted by Hardy and Littlewood for a specific triple a, b, c involves $\int_2^{x^{1/3}/(abc)^{4/3}} dt/\log(a(g_0+abct))\log(b(g_0+abct))\log(c(g_0+abct))$ instead of $27x^{1/3}/((abc)^{4/3}\log^3 x)$, though the two are asymptotically equal for fixed a, b, c. For numerical comparisons we use both $x^{1/3}/\log^3 x$ and $(1/27)\int_2^{x^{1/3}} dt/\log^3 t$. Again, the two expressions are asymptotically equal, but at finite values can be considerably different. Thus we predict that

$$C_3(x) \sim \tau_3 \frac{x^{1/3}}{\log^3 x} \sim \frac{\tau_3}{27} \int_2^{x^{1/3}} \frac{dt}{\log^3 t},$$

where $\tau_3 \approx 2100$. Due to the above considerations, we also predict that the convergence of $C_3(x)/(x^{1/3}/\log^3 x)$ to τ_3 should be eventually from above, while the convergence of $C_3(x)/\int_2^{x^{1/3}} dt/\log^3 t$ to τ_3 should be eventually from below. The data of Pinch, Chick, Davies, and Williams give the following:

x	10^3	10^4	10^5	10^6	10^7	10^8	10^9	10^{10}	10^{11}
$C_3(x)$	1	7	12	23	47	84	172	335	590
β	32.96	253.9	394.5	606.5	913.5	1131	1531	1898	2065
γ	9.092	53.13	78.07	128.1	220.2	321.9	519.1	761.3	961.5

x	10^{12}	10^{13}	10^{14}	10^{15}	10^{16}	10^{17}	10^{18}	10^{19}	10^{20}
$C_3(x)$	1000	1858	3284	6083	10816	19539	35586	65409	120459
β	2110	2313	2370	2506	2510	2525	2534	2538	2535
γ	1113	1349	1496	1680	1763	1839	1899	1947	1982

$$C_3(x) = \beta x^{1/3}/\log^3 x = (\gamma/27)\int_2^{x^{1/3}} dt/\log^3 t$$

Although these data may not be too persuasive that the numbers β and γ are tending to a common limit which is about 2100, they at least suggest that our heuristic is not too wildly wrong.

We can compare our conjecture with that made by Galway [10] for the number of 2-pseudoprimes $\le x$ with exactly two prime factors. Note that $n = (ag+1)(bg+1)$ is a 2-pseudoprime, where each of the two factors are primes, and $(a,b) = 1$, if and only if $2^g \equiv 1 \pmod n$, if and only if 2 is an ath power $\pmod p$ and is a bth power $\pmod q$. By the Chebotarev density theorem we know that 2 is an mth power $\pmod p$ for a proportion $1/m$ of the primes $p \equiv 1 \pmod m$ except when $4|m$ in which case the proportion increases to $2/m$. Assuming the independence of these proportions when taking $m = a$ and $m = b$, Galway conjectures that the number of such 2-pseudoprimes is

$$\sim 8\kappa_2 \prod_{p>2} \left(\frac{1-2/p}{(1-1/p)^2} \right) \frac{x^{1/2}}{\log^2 x},$$

where

$$\kappa_2 = \sum_{n\ge 1} \frac{\delta_n}{n^{3/2}} \sum_{\substack{a<b,\ n=ab \\ a,b \text{ pairwise coprime}}} \prod_{\substack{p|n(b-a) \\ p>2}} \frac{p-1}{p-2},$$

where $\delta_n = 2$ if $4|n$, and $\delta_n = 0$ otherwise. Galway also compares this persuasively with Pinch's data [16]. This may evidently be compared to our conjecture for $C_3(x)$.

We might also suppose that for any given integer $k \ge 3$, we have $C_k(x) \sim \tau_k x^{1/k}/\log^k x$ for some constant $\tau_k > 0$.

9. An upper bound for the number of Carmichael numbers with $k \ge 3$ prime factors

Theorem 7. *We have $C_k(x) \ll x^{2/3}(\log x)^{(2^{k-2}-1)/3}$ holding uniformly for each integer $k \ge 3$.*

Proof. Let $A = (\log x)^{(2^{k-2}-1)/3}$. If $A \ge x^{1/3}$ the theorem is trivially true, so assume that $A < x^{1/3}$. Suppose that n is a Carmichael number. If prime p divides n then, by Korselt's criterion, $n \equiv 1 \equiv p \pmod{p-1}$, so that $n \equiv p \pmod{p^2-p}$. However $n > p$, so the number of such Carmichael numbers up to x is $\le x/(p^2-p)$. Thus the total number of Carmichael numbers up to x with a prime factor $\ge x^{1/3}/A$ is $\le \sum_{p>x^{1/3}/A} x/(p^2-p) \ll x^{2/3}A$.

Now consider Carmichael numbers $n = p_1p_2\dots p_k \in (x/2, x]$ whose prime factors satisfy $p_1 < p_2 < \cdots < p_k < x^{1/3}/A$. Select l minimally so that $m = p_1p_2\dots p_l \ge \frac{1}{2}x^{1/3}A^2$; then $m < \frac{1}{2}x^{2/3}A$ and so $l \le k-2$. If $n = mr$ then, by Korselt's criterion, r belongs to some residue class $\pmod{\lambda(m)}$. Therefore the number of such r is $\le 1 + x/m\lambda(m) \ll x^{2/3}A^{-2}/\lambda(m)$, and so the total number of such Carmichael numbers for a given value of $l \le k-2$ is

$$\ll x^{2/3}A^{-2} \sum_{\substack{m<\frac{1}{2}x^{2/3}A \\ \omega(m)=l}} \frac{1}{\lambda(m)} \le x^{2/3}A^{-2} \sum_{\substack{m<x \\ \omega(m)=l}} \frac{1}{\lambda(m)}.$$

To determine $\lambda(m)$ we will need to understand the common factors of the $p_i - 1$ in detail. Recall that $g(n)$ denotes the gcd of the numbers $p-1$ where p runs over

the prime factors of n. For each divisor $d > 1$ of m, let

$$g_d := \prod_{j|m/d} g(jd)^{\mu(j)}.$$

The numbers g_d have the following properties:

$$\prod_{j|m/d} g_{jd} = g(d) \quad \text{for each } d|m, d > 1,$$

$$\prod_{d|m,d>1} g_d = \lambda(m),$$

$$\gcd(g_{d_1}, g_{d_2}) = 1 \text{ for all } d_1|m, d_2|m \text{ with } d_1 \nmid d_2, d_2 \nmid d_1.$$

In particular, if $d_1|m, d_2|m$ with $\omega(d_1) = \omega(d_2)$ and $d_1 \neq d_2$, then g_{d_1} and g_{d_2} are coprime. Write $b_j = \prod_{\omega(d)=j} g_d$, and note that the number of ways a given number b_j can arise is no larger than the number of ways one can write b_j as the product of $\binom{l}{j}$ pairwise coprime integers, which is $\leq \binom{l}{j}^{\omega(b_j)}$. We note that $\prod_j b_j^j = \phi(m) < m$ and $\lambda(m) = \prod_j b_j$. Therefore we have

$$\sum_{\substack{m<x \\ \omega(m)=l}} \frac{1}{\lambda(m)} \leq \sum_{b_1 b_2^2 \ldots b_l^l \leq x} \prod_j \frac{\binom{l}{j}^{\omega(b_j)}}{b_j} \leq \prod_j \left\{ \sum_{b \leq x^{1/j}} \frac{\binom{l}{j}^{\omega(b)}}{b} \right\}.$$

Now

$$\sum_{b \leq B} t^{\omega(b)}/b \leq \prod_{p \leq B} (1 + t/(p-1)) \leq \prod_{p \leq B} (1 + 1/(p-1))^t \leq (3 \log B)^t,$$

uniformly for all $t \geq 1, B \geq 2$. Thus the last product over j is $\ll (\log x)^{2^l - 1}$, uniformly for all $l \geq 1$. Summing over all $l \leq k-2$, we get that

$$\sum_{\substack{m<x \\ \omega(m) \leq k-2}} \frac{1}{\lambda(m)} \ll (\log x)^{2^{k-2}-1} = A^3,$$

uniformly in k, which implies our result. □

Remark. If we could find an m dividing n that we can guarantee is closer (logarithmically) to $x^{1/2}$, then we could use the proof above to improve the bound in Theorem 7. It seems that the integers n that cause us to have so poor an estimate as in Theorem 7 are those that have three prime factors which are each $\gg x^{1/3}$, and the rest bounded: If we could show that there are few such Carmichael numbers then perhaps we could improve our estimate above.

Acknowledgments. We gratefully acknowledge Richard Pinch for sharing various of his unpublished computations of Carmichael numbers and 2-pseudoprimes, and for determining, at our request, the counts for imprimitive Carmichael numbers up to 10^{16}. We are indebted to John Chick and Gordon Davies for sharing their Carmichael computations, especially their work on the computation of the constant κ_3 in Section 8, and to Matthew Williams for his calculations of $C_3(10^n)$ reported in Sections 1 and 8. We also thank Will Galway for sharing his ideas from [10].

Added after posting

Replace the text on page 906, on the fourth to last line of Section 8, immediately following the last displayed equation, with this correction:
where $\delta_n = 2$ if $4|n$, and $\delta_n = 1$ otherwise.

References

[1] W. R. Alford and J. Grantham, *Carmichael numbers with exactly k prime factors* (to appear).

[2] W. R. Alford, A. Granville and C. Pomerance, *There are infinitely many Carmichael numbers*, Ann. Math. **140** (1994), 703–722. MR **95k:**11114

[3] W. R. Alford, A. Granville and C. Pomerance, *On the difficulty of finding reliable witnesses*, Lecture Notes in Computer Sci. **877** (1995), 1–16. MR **96d:**11136

[4] R. Balasubramanian and S. V. Nagaraj, *Density of Carmichael numbers with three prime factors*, Math. Comp. **66** (1997), 1705–1708. MR **98d:**11110

[5] E. R. Canfield, P. Erdős and C. Pomerance, *On a problem of Oppenheim concerning "Factorisatio Numerorum"*, J. Number Theory **17** (1983), 1–28. MR **85j:**11012

[6] I. Damgård, P. Landrock and C. Pomerance, *Average case error estimates for the strong probable prime test*, Math. Comp. **61** (1993), 177–194. MR **94b:**11124

[8] P. Erdős, *On pseudoprimes and Carmichael numbers*, Publ. Math. Debrecen **4** (1956), 201–206. MR **18:**18e

[9] P. Erdős and M. Kac, *The Gaussian law of errors in the theory of additive number theoretic functions*, Amer. J. Math **62** (1940), 738–742. MR **2:**42c

[10] W. Galway, *The density of pseudoprimes with two prime factors* (to appear).

[11] A. Granville, *Primality testing and Carmichael numbers*, Notices Amer. Math. Soc. **39** (1992), 696–700.

[12] G. H. Hardy and J. E. Littlewood, *Some problems on partitio numerorum III. On the expression of a number as a sum of primes*, Acta Math. **44** (1923), 1–70.

[13] A. Ivić and G. Tenenbaum, *Local densities over integers free of large prime factors*, Quart. J. Math. Oxford Ser. (2) **37** (1986), 401–417. MR **88a:**11092

[14] R. G. E. Pinch, *The Carmichael numbers up to* 10^{15}, Math. Comp. **61** (1993), 381–391. MR **92m:**11137

[15] R. G. E. Pinch, *The Carmichael numbers up to* 10^{16} (to appear).

[16] R. G. E. Pinch, *The pseudoprimes up to* 10^{12} (to appear).

[17] C. Pomerance, *On the distribution of pseudoprimes*, Math. Comp. **37** (1981), 587–593. MR **83k:**10009

[19] C. Pomerance, *Two methods in elementary analytic number theory*, Number Theory and Applications (Banff, 1988; R. A. Mollin, ed.), NATO Adv. Sci. Inst. Ser. C: Math. Phys. Sci., vol. 265, Reidel, Dordrecht, 1989, pp. 135–161.

[20] C. Pomerance, *Carmichael numbers*, Nieuw Arch. Wisk. **11** (1993), 199–209. MR **94h:**11085

[21] C. Pomerance, J. Selfridge and S. S. Wagstaff Jr., *The pseudoprimes to* $25 \cdot 10^9$, Math. Comp. **35** (1980), 1003–1026. MR **82g:**10030

[22] A. Schinzel and W. Sierpiński, *Sur certaines hypothèses concernant les nombres premiers*, Acta Arith. **4** (1958), 185–208. MR **21:**4936

[23] ———, *Sur certaines hypothèses concernant les nombres premiers.* Erratum, Acta Arith. **5** (1959), 259. MR **21:**493b

[24] D. Shanks, *Solved and unsolved problems in number theory*, 3rd ed., Chelsea, New York, 1985. MR **86j:**11001

Department of Mathematics, University of Georgia, Athens, Georgia 30602
E-mail address: andrew@math.uga.edu

Fundamental Mathematics Research, Bell Laboratories, 600 Mountain Ave., Murray Hill, New Jersey 07974
E-mail address: carlp@research.bell-labs.com