MATHEMATICS OF COMPUTATION
VOLUME 37, NUMBER 156
OCTOBER 1981

On the Distribution of Pseudoprimes

By Carl Pomerance*

Abstract. Let $\mathcal{P}(x)$ denote the pseudoprime counting function. With

$$L(x) = \exp\{\log x \log\log\log x/\log\log x\},$$

we prove $\mathcal{P}(x) < x \cdot L(x)^{-1/2}$ for large x, an improvement on the 1956 work of Erdös. We conjecture that $\mathcal{P}(x) = x \cdot L(x)^{-1+o(1)}$.

1. Introduction. A composite natural number n for which $2^{n-1} \equiv 1 \pmod{n}$ is called a pseudoprime. The least such n is 341. Let $\mathcal{P}(x)$ denote the number of pseudoprimes not exceeding x. It is known that, for some $c > 0$ and all large x,

$$\exp\{(\log x)^{5/14}\} \leqslant \mathcal{P}(x) \leqslant x \cdot \exp\{-c(\log x \log_2 x)^{1/2}\},$$

where $\log_k$ denotes the k-fold iteration of the natural logarithm. The lower bound was accomplished recently in [6], while the upper bound is due to Erdös [3].

The main purpose of this paper is to present an improvement of Erdös's upper bound for $\mathcal{P}(x)$. We show that, for all large x,

$$\mathcal{P}(x) \leqslant x \cdot L(x)^{-1/2}, \tag{1}$$

where

$$L(x) = \exp\{\log x \log_3 x/\log_2 x\}.$$

We believe that (1) is near to best possible for $\mathcal{P}(x)$. In fact we conjecture that

$$\mathcal{P}(x) = x \cdot L(x)^{-1+o(1)}. \tag{2}$$

We briefly discuss how the method of proof of (1) can be applied to two other problems: the distribution of Carmichael numbers and the study of the number of solutions m of $\varphi(m) = n$ where φ is Euler's function.

We remark that our main result (1) can easily be generalized to pseudoprimes to an arbitrary base $b \geqslant 2$. (A composite natural number n is called a pseudoprime to base b if $b^{n-1} \equiv 1 \pmod{n}$.) Thus, if $\mathcal{P}_b(x)$ denotes the counting function for the base b pseudoprimes, we have $\mathcal{P}_b(x) \leqslant x \cdot L(x)^{-1/2}$ for all $x \geqslant x_0(b)$.

Throughout the paper the letter p will always denote a prime. If S is a set, then $\#S$ will denote the cardinality of S.

2. The Exponent to Which 2 Belongs Modulo n. If n is odd, let $l_2(n)$ denote the exponent to which 2 belongs modulo n. Thus a composite natural number n is a pseudoprime if and only if $l_2(n) \mid n - 1$. To achieve our main result about pseudoprimes, we shall first prove a theorem on the number of solutions m of the equation $l_2(m) = n$.

Received September 12, 1980; revised January 19, 1981.
1980 *Mathematics Subject Classification.* Primary 10A15.
Key words and phrases. Pseudoprime, Carmichael number, Euler's function.
* Research supported in part by a grant from the National Science Foundation.

0025-5718/81/0000-0174/$02.75

THEOREM 1. *There is an x_0 such that if n is a natural number and $x \geqslant x_0$, then*

$$\#\{m \leqslant x: l_2(m) = n\} \leqslant x \cdot \exp\left(-\log x \cdot \frac{3 + \log_3 x}{2 \log_2 x}\right).$$

Proof. We may assume $x > n$ for otherwise there are no $m \leqslant x$ with $l_2(m) = n$. If $c > 0$, then

$$\sum_{\substack{m \leqslant x \\ l_2(m) = n}} 1 \leqslant x^c \sum_{l_2(m) = n} m^{-c} \leqslant x^c \sum_{p \mid m \Rightarrow l_2(p) \mid n} m^{-c} = x^c \prod_{l_2(p) \mid n} (1 - p^{-c})^{-1} = x^c A,$$

say. We shall choose $c = 1 - (4 + \log_3 x)/(2 \log_2 x)$, where x is large enough so that $c \geqslant 7/8$.

The theorem will follow if we show

$$\log A = o(\log x / \log_2 x). \tag{3}$$

Note that, since $c \geqslant 7/8$,

$$\log A = \sum_{l_2(p) \mid n} p^{-c} + O(1) = \sum_{d \mid n} \sum_{l_2(p) = d} p^{-c} + O(1).$$

The primes p with $l_2(p) = d$ all divide $2^d - 1$, so there are less than d such primes. Say they are $q_1, q_2, \ldots, q_t$ where $0 \leqslant t < d$. Each $q_i \equiv 1 \pmod{d}$, so that

$$\sum_{l_2(p) = d} p^{-c} = \sum_{i=1}^{t} q_i^{-c} \leqslant \sum_{i=1}^{t} (di + 1)^{-c} < d^{-c} \sum_{i=1}^{d-1} i^{-c} < (1 - c)^{-1} d^{1-2c}.$$

Thus,

$$\log A \leqslant (1 - c)^{-1} \sum_{d \mid n} d^{1-2c} + O(1) \leqslant (1 - c)^{-1} \prod_{p \mid n} (1 - p^{1-2c})^{-1} + O(1). \tag{4}$$

Now, since $1 - 2c \leqslant -3/4$,

$$\log \prod_{p \mid n} (1 - p^{1-2c})^{-1} = \sum_{p \mid n} p^{1-2c} + O(1) \leqslant \sum_{p < 2 \log x} p^{1-2c} + O(1), \tag{5}$$

where x is sufficiently large, so that $\prod_{p \leqslant 2 \log x} p \geqslant x$. Using (5), partial summation, and the prime number theorem, it is seen that

$$\log \prod_{p \mid n} (1 - p^{1-2c})^{-1} \ll \frac{(\log x)^{2-2c}}{(2 - 2c) \log_2 x} \ll \frac{\log_2 x}{\log_3 x}.$$

Thus, if x is sufficiently large, we have

$$\prod_{p \mid n} (1 - p^{1-2c})^{-1} \leqslant (\log x)^{1/2},$$

so that from (4) we have

$$\log A \leqslant \frac{2 \log_2 x}{4 + \log_3 x} (\log x)^{1/2} + O(1),$$

which proves (3) and the theorem.

3. Pseudoprimes.

THEOREM 2. *For all sufficiently large x, we have*

$$\mathcal{P}(x) \leqslant x \cdot \exp\left(-\frac{\log x \log_3 x}{2 \log_2 x}\right).$$

Proof. Recalling the definition of $L(x)$ as $\exp(\log x \log_3 x/\log_2 x)$, we divide the pseudoprimes $n \leqslant x$ into four possibly overlapping classes:

(i) $n \leqslant x \cdot L(x)^{-1}$,

(ii) there is a prime $p \mid n$ with $l_2(p) \leqslant L(x), p > L(x)^3$,

(iii) there is a prime $p \mid n$ with $l_2(p) > L(x)$,

(iv) $n > x \cdot L(x)^{-1}$ and every prime $p \mid n$ is at most $L(x)^3$.

The number of n in class (i) is obviously at most

$$x \cdot L(x)^{-1}. \tag{6}$$

The number of primes p with $l_2(p) \leqslant L(x)$ is exactly

$$\sum_{m \leqslant L(x)} \sum_{l_2(p)=m} 1 < \sum_{m \leqslant L(x)} m < L(x)^2.$$

Thus the number of n in class (ii) is at most

$$\sum_{\substack{p > L(x)^3 \\ l_2(p) \leqslant L(x)}} x/p < x \cdot L(x)^{-3} \sum_{l_2(p) \leqslant L(x)} 1 < x \cdot L(x)^{-1}. \tag{7}$$

If n is a pseudoprime and $d \mid n$, then

$$n \equiv 0 \pmod{d}, \qquad n \equiv 1 \pmod{l_2(d)}, \quad (d, l_2(d)) = 1. \tag{8}$$

Thus the number of pseudoprimes $n \leqslant x$ with $d \mid n$ is at most $1 + x/(dl_2(d))$. If $d = p$, a prime, then we throw out the solution $n = p$ of (8), so that in this case there are at most $x/(pl_2(p))$ pseudoprimes $n \leqslant x$ with $p \mid n$. Thus the number of n in class (iii) is at most

$$\sum_{\substack{2 < p \leqslant x \\ l_2(p) > L(x)}} \frac{x}{pl_2(p)} < \frac{x}{L(x)} \sum_{p \leqslant x} \frac{1}{p} \sim \frac{x \log_2 x}{L(x)}. \tag{9}$$

If n is in class (iv), then n must have a divisor d with

$$x \cdot L(x)^{-4} < d \leqslant x \cdot L(x)^{-1}. \tag{10}$$

Thus, by the comment following (8), the number of n in class (iv) is at most

$$\begin{aligned} \sum{}' \left(1 + \frac{x}{dl_2(d)}\right) &\leqslant x \cdot L(x)^{-1} + x \sum{}' \frac{1}{dl_2(d)} \\ &= x \cdot L(x)^{-1} + x \sum_{m \leqslant x} \frac{1}{m} \sum_{l_2(d)=m}{}' \frac{1}{d}, \end{aligned}$$

where $\sum'$ denotes the sum over odd d satisfying (10). Using Theorem 1 and partial summation, the inner sum is, for large x, at most

$$\exp\left(-\log x \cdot \frac{2 + \log_3 x}{2 \log_2 x}\right).$$

Thus, for large x, the number of n in class (iv) is at most

$$x \cdot \exp\left(-\log x \cdot \frac{1 + \log_3 x}{2 \log_2 x}\right). \tag{11}$$

Hence, using the estimates (6), (7), (9), (11) for the number of pseudoprimes $n \leqslant x$ in each of the four classes, we have our theorem.

Remark. Concerning our conjecture (2), an improvement would be attainable in Theorem 2 if we could improve Theorem 1. In the proof of Theorem 1, if a less crude estimate for $\sum_{l_2(p)=d} p^{-c}$ could be obtained, possibly only on average, then Theorem 1 could be strengthened. (The methods of Erdös [4] may be helpful for this.) We conjecture that uniformly for all n

$$\#\{m \leqslant x: l_2(m) = n\} \leqslant x \cdot L(x)^{-1+\theta(x)},$$

where $\theta(x) \to 0$ as $x \to \infty$. From such a result, the method of proof of Theorem 2 gives

$$\mathcal{P}(x) \leqslant x \cdot L(x)^{-1+o(1)}.$$

This would be half of the battle for (2). For the other half, in [7], elaborating on an argument given by Erdös [3], a heuristic argument is presented that

$$C(x) \geqslant x \cdot L(x)^{-2+o(1)},$$

where $C(x)$ is the number of Carmichael numbers not exceeding x. (We say n is a Carmichael number if it is a pseudoprime to every base b with $(b, n) = 1$.) It is easy to show that every Carmichael number is a pseudoprime, so $\mathcal{P}(x) \geqslant C(x)$. Moreover, in the next section we shall show how the heuristic argument in [7], just alluded to, can be improved to

$$C(x) \geqslant x \cdot L(x)^{-1+o(1)}.$$

This argument then supports the other half of our conjecture (2).

4. Applications of the Method. The method of proof of Theorem 1 has been used profitably by Rankin [8] and de Bruijn [1] in the study of the distribution of integers with no large prime factors. The method has recently been used in the proofs of Theorems 5.1 and 6.1 of [2]. (These theorems deal with the number of factorizations of an integer.) In this section we give two further applications of this method.

The Distribution of Carmichael Numbers. It is easily seen that a composite natural number n is a Carmichael number if and only if $\lambda(n) \mid (n-1)$, where $\lambda(n)$ is the universal exponent modulo n. It is well known that $\lambda(n)$ is the least common multiple of the numbers $p^{a-1}(p-1)$ for prime powers $p^a \mid n$, except when $8 \mid n$, and then $\lambda(n)$ is the least common multiple of 2^{a-2} and $\lambda(n/2^a)$ where $2^a \parallel n$.

In Erdös [3] it is shown that

$$C(x) \ll x \cdot L(x)^{-a} \tag{12}$$

for some positive constant a, where $C(x)$ is the Carmichael number counting function. In [7] (Theorem 6) the estimates in Erdös's argument are sharpened to show that in (12) we may choose a as any number less than 1. By the methods of this paper we can achieve now the sharper result

$$C(x) \leqslant x \cdot \exp\left\{-\frac{\log x}{\log_2 x}\left(\log_3 x + \log_4 x + \frac{\log_4 x - 1}{\log_3 x} + O\left(\left(\frac{\log_4 x}{\log_3 x}\right)^2\right)\right)\right\}. \tag{13}$$

In [7] we hold out the possibility that $C(x) \sim F(x)$ where

$$F(x) = x \cdot \exp\{-\log x(1 + \log_3 x)/\log_2 x\},$$

since $F(x)$ does a good job of approximating $C(x)$ for $x \leqslant 25 \cdot 10^9$. However, (13) now shows that $C(x)$ and $F(x)$ are not asymptotic: in fact we have $C(x) = o(F(x))$.

We say a word on how (13) is established. We first need to prove an analogue for Theorem 1 where $\lambda(n)$ replaces $l_2(n)$. This is in fact easy since the sum $\Sigma_{\lambda(p)=d} p^{-c}$ can now be trivially replaced by the larger quantity d^{-c}. By choosing

$$c = 1 - \frac{1}{\log_2 x}\left(\log_3 x + \log_4 x + \frac{\log_4 x - 1}{\log_3 x} - 2\left(\frac{\log_4 x}{\log_3 x}\right)^2\right), \tag{14}$$

as in the proof of Theorem 5.1 of [2], we obtain

$$\sum_{\substack{m \leqslant x \\ \lambda(m)=n}} 1 \leqslant x \cdot \exp\left\{-\frac{\log x}{\log_2 x}\left(\log_3 x + \log_4 x + \frac{\log_4 x - 1}{\log_3 x} + O\left(\left(\frac{\log_4 x}{\log_3 x}\right)^2\right)\right)\right\}. \tag{15}$$

In the proof of Theorem 6 in [7] we now use (15) instead of Lemma 2, choosing $\delta = 1 - c$ where c is given by (14) above. We thus obtain (13).

One might well wonder why we obtain such accuracy in the upper bound estimate for $C(x)$ when so little is known about lower bounds: we cannot even disprove $C(x) = O(1)$! However, in [7], as we mentioned above, a heuristic argument is presented for a lower bound for $C(x)$. This argument draws an analogy between the functions

$$\Psi(x, y) = \#\{n \leqslant x : P(n) \leqslant y\},$$

$$\Psi'(x, y) = \#\{\text{primes } p \leqslant x : p - 1 \text{ square-free}, P(p-1) \leqslant y\},$$

where $P(n)$ denotes the greatest prime factor of n. In particular, part of the heuristic argument is the conjecture

$$\frac{1}{x}\Psi(x, y) \asymp \frac{1}{\pi(x)}\Psi'(x, y)$$

for y in the vicinity of $\exp\{(\log x)^{1/2}\}$. Using what was known about $\Psi(x, y)$, the heuristic argument in [7] gave $C(x) \geqslant x \cdot L(x)^{-2+o(1)}$. However, a recent development is that in [2] (Section 3), a much sharper estimate for $\Psi(x, y)$ is obtained. Using this new theorem together with the same heuristic argument in [7], we now have

$$C(x) \geqslant x \cdot \exp\left\{-\frac{\log x}{\log_2 x}\left(\log_3 x + \log_4 x + \frac{\log_4 x - 1}{\log_3 x} + O\left(\left(\frac{\log_4 x}{\log_3 x}\right)^2\right)\right)\right\}.$$

(We change the argument in [7] at only one place: we now let A denote the product of the primes up to $\log x/(\log_2 x)^2$.) That is, the heuristic argument of [7] now implies the conjecture that equality holds in (13).

Popular Values of Euler's Function. In [5] we studied the function $N(n) = \#\{m: \varphi(m) = n\}$, where φ denotes Euler's function. We gave a proof (that depended on Lemma 2 of [7]) that $N(n) \leqslant n \cdot L(n)^{-1+o(1)}$. Moreover, we gave a heuristic argument that equality holds for an infinite set of n. We were able only to prove that $N(n) > n^{5/9}$ for infinitely many n.

The functions $\lambda(m)$ and $\varphi(m)$ are so similar that virtually the same proof that gives (15) also gives

$$(16)\quad N(n) \leqslant n \cdot \exp\left\{ -\frac{\log n}{\log_2 n}\left(\log_3 n + \log_4 n + \frac{\log_4 n - 1}{\log_3 n} + O\left(\left(\frac{\log_4 n}{\log_3 n}\right)^2\right)\right)\right\}.$$

Moreover, the new sharp results in [2] on $\Psi(x, y)$, when combined with the heuristic argument in [5], imply the conjecture that equality in (16) holds for infinitely many n.

5. Numerical Evidence. In the Table we have presented values of the functions $k(x), j(x), k_2(x), j_2(x)$ (defined below) for selected values of x.

As in [7], $k(x)$ is defined by the equation

$$C(x) = x \cdot \exp\{-k(x)\log x \log_3 x/\log_2 x\}.$$

In the Table we have reproduced the data in Table 3 of [7] on $k(x)$. We conjecture $\lim k(x) = 1$. Note that (13) implies $k(x) > 1$ for all large x. The conjecture that we have equality in (13) implies that if we define $j(x)$ by the equation

$$C(x) = x \cdot \exp\left\{ -\frac{\log x}{\log_2 x}(\log_3 x + \log_4 x + j(x))\right\},$$

then $j(x)$ tends to 0 slowly as $x \to \infty$.

We have also checked our conjecture (2) numerically. If we define $k_2(x)$ by the equation

$$\mathcal{P}(x) = x \cdot \exp\{-k_2(x)\log x \log_3 x/\log_2 x\},$$

then (2) is equivalent to the assertion $\lim k_2(x) = 1$. Of course (1) implies $k_2(x) \geqslant 1/2$ for all large x. (The values for $\mathcal{P}(x)$ in the Table were computed from one of the print-outs associated with [7].) Say we now make an even stronger assertion than (2), namely that if $j_2(x)$ is defined so that

$$\mathcal{P}(x) = x \cdot \exp\left\{ -\frac{\log x}{\log_2 x}(\log_3 x + \log_4 x + j_2(x))\right\},$$

then $\lim j_2(x) = 0$. In the Table we have presented values of $j_2(x)$ for selected values of x.

TABLE

$x/10^9$	$C(x)$	$\mathcal{P}(x)$	$k(x)$	$j(x)$	$k_2(x)$	$j_2(x)$
1	646	5597	1.8799	0.8723	1.5951	0.5565
5	1184	11108	1.8722	0.8633	1.5974	0.5519
10	1547	14884	1.8687	0.8593	1.5989	0.5508
15	1782	17658	1.8686	0.8591	1.5998	0.5504
20	1983	19865	1.8678	0.8582	1.6009	0.5506
25	2163	21853	1.8668	0.8570	1.6013	0.5503

Department of Mathematics
University of Georgia
Athens, Georgia 30602

1. N. G. DE BRUIJN, "On the number of positive integers $\leq x$ and free of prime factors $>y$. II," *Nederl. Akad. Wetensch. Proc. Ser. A*, v. 69, 1966, pp. 239–247.

2. E. R. CANFIELD, P. ERDÖS & C. POMERANCE, "On a problem of Oppenheim concerning "Factorisatio Numerorum"," *J. Number Theory*. (To appear.)

3. P. ERDÖS, "On pseudoprimes and Carmichael numbers," *Publ. Math. Debrecen*, v. 4, 1956, pp. 201–206.

4. P. ERDÖS, "On the sum $\Sigma_{d|2^n-1} d^{-1}$," *Israel J. Math.*, v. 9, 1971, pp. 43–48.

5. C. POMERANCE, "Popular values of Euler's function," *Mathematika*, v. 27, 1980, pp. 84–89.

6. C. POMERANCE, "A new lower bound for the pseudoprime counting function," *Illinois J. Math.* (To appear.)

7. C. POMERANCE, J. L. SELFRIDGE & S. S. WAGSTAFF, JR., "The pseudoprimes to $25 \cdot 10^9$," *Math. Comp.*, v. 35, 1980, pp. 1003–1026.

8. R. A. RANKIN, "The difference between consecutive prime numbers," *J. London Math. Soc.*, v. 13, 1938, pp. 242–247.