

On pseudoprimes and Carmichael numbers.

Dedicated to the memory of my friend Tibor Szele.

By P. ERDŐS in Haifa.

A number n is said to be a pseudoprime if

$$(1) \quad 2^n \equiv 2 \pmod{n}.$$

It is said to be an absolute pseudoprime or a Carmichael number if for every $(a, n) = 1$

$$(2) \quad a^n \equiv a \pmod{n}.$$

Denote by $P(x)$ the number of pseudoprimes and by $C(x)$ the number of Carmichael numbers not exceeding x . It is known that¹⁾

$$(3) \quad c_1 \log x < P(x) < x \exp(-c_2(\log x)^{\frac{1}{4}}).$$

KNÖDEL²⁾ recently proved that

$$(4) \quad C(x) < x \exp(-c_3(\log x \log \log x)^{\frac{1}{2}}),$$

it is not yet known whether $C(x) \rightarrow \infty$ as $x \rightarrow \infty$ i. e. it is not known if there are infinitely many Carmichael numbers. In the present paper I prove by KNÖDEL's method that

$$(5) \quad P(x) < x \exp(-c_4(\log x \log \log x)^{\frac{1}{2}})$$

and

$$(6) \quad C(x) < x \exp(-c_5 \log x \log \log x / \log \log x).$$

KNÖDEL conjectured that $C(x) < x^{1-\delta}$ for a suitable positive δ . I would rather conjecture that $C(x) > x^{1-\varepsilon}$ for every $\varepsilon > 0$ and $x > x(\varepsilon)$, in fact I believe that (6) can not be very much improved. I shall give some heuristic reasons for this guess. Finally I shall state some theorems without proof.

As far as I know D. H. LEHMER¹⁾ was the first to discover that there are even numbers n which satisfy (1), and BEEGER²⁾ proved that there are

¹⁾ P. ERDŐS, *Amer. Math. Monthly* 57 (1950), 404—407.

²⁾ *Archiv der Math.* 4 (1953), 282—284.

³⁾ *Amer. Math. Monthly* 58 (1951), 553—555.

infinitely many such integers. I do not know if there are any composite numbers n for which $a^n \equiv a \pmod{n}$ for every integer a (i. e. not only for the $(a, n) = 1$).*)

Throughout this paper $c_1, c_2, \dots$ will denote positive absolute constants, p_i and P_k will denote primes, $\log_k x$ will denote the k times iterated logarithm. First we prove the following

Lemma 1. Denote by $N(p_1, p_2, \dots, p_k; x)$ the number of integers not exceeding x composed of $p_1, p_2, \dots, p_k$. Put $k^u = x$. Then for $u < \log x / \log_2 x$ (i. e. $k > \log x$)

$$N(p_1, p_2, \dots, p_k; x) < x \exp(-c_6 u \log u).$$

Clearly $N(p_1, p_2, \dots, p_k; x) \leq N(2, 3, \dots, P_k; x)$ where P_k denotes the k -th prime. Now $\pi(k^2) > k$. Thus by a theorem of DE BRUIJN⁴⁾

$$N(p_1, p_2, \dots, p_k; x) \leq N(2, 3, \dots, P_k; x) < \psi(x, k^2) < x \exp(-c_6 u \log u),$$

where $\psi(x, y)$ denotes the number of integers $\leq x$ all whose prime factors are $\leq y$. Thus our Lemma is proved.

Now we prove (5). Denote by $l_2(p)$ the smallest exponent satisfying $2^{l_2(p)} \equiv 1 \pmod{p}$. We split the pseudoprimes not exceeding x into two classes. In the first class are the pseudoprimes n every prime factor of which satisfies

$$l_2(p) < \exp((\log x \cdot \log_2 x)^{\frac{1}{2}}).$$

Clearly all the pseudoprimes of the first class are composed of the prime factors of

$$(7) \quad 2^t - 1, \quad 1 < t < \exp((\log x \cdot \log_2 x)^{\frac{1}{2}}).$$

The number of prime factors of $2^t - 1$ is clearly less than t , thus the number k of prime factors of all the numbers (7) clearly satisfies

$$k < t^2 < \exp(2(\log x \log_2 x)^{\frac{1}{2}}).$$

Thus by Lemma 1 the number of pseudoprimes of the first class is less than

$$(8) \quad x \exp(-c_7(\log x \log_2 x)^{\frac{1}{2}}).$$

The u of Lemma 1 here equals $c_8(\log x / \log_2 x)^{\frac{1}{2}}$.

Every pseudoprime of the second class has a prime factor p satisfying $l_2(p) \geq \exp((\log x \log_2 x)^{\frac{1}{2}})$. Since n is a pseudoprime we must have

$$(9) \quad n \equiv 0 \pmod{p}, \quad n \equiv 1 \pmod{l_2(p)}, \quad n > p,$$

* Added in proof: In a recent letter Dr. KNÖDEL proved that every Carmichael number has the above property.

⁴⁾ Indag. Math. 13 (1951), 50–60

(for if $n = p$, n would not be a pseudoprime). Thus $n > p \cdot l_2(p)$. Let now $p_1, p_2, \dots, p_r$ be the primes not exceeding x for which $l_2(p) \cong \exp((\log x \log_2 x)^{\frac{1}{2}})$. We have by (9) that the number of pseudoprimes of the second class does not exceed

$$(10) \quad x \sum_{i=1}^r \frac{1}{p_i l_2(p_i)} < x \exp(-(\log x \cdot \log_2 x)^{\frac{1}{2}}) \sum_{p < x} \frac{1}{p} < \\ < x \exp\left(-\frac{1}{2}(\log x \cdot \log_2 x)^{\frac{1}{2}}\right).$$

(8) and (10) clearly imply (5).

Now we prove (6). Let k be an integer and denote by $f(k)$ the least common multiple of $p_j - 1$, $j = 1, 2, \dots$ where p_j runs through all the prime factors of k . First we state

Lemma 2. *The number of solutions of $f(k) = t$, $k \leq y$ does not exceed*

$$y \exp(-c_0 \log y \cdot \log_3 y / \log_2 y)$$

(independently of t !)

Let us assume that Lemma 2 has already been proved. Then the proof of (6) proceeds as follows: It is well known (and obvious) that n is a Carmichael number if and only if it is a composite, squarefree number such that for every prime factor q of n , $q - 1$ divides $n - 1$. We split the Carmichael numbers not exceeding n into two classes. In the first class are the Carmichael numbers whose greatest prime factor is greater than $x^{\frac{1}{6}}$. Let n be a Carmichael number of the first class and p its largest prime factor. We evidently have

$$n \equiv 0 \pmod{p}, \quad n \equiv 1 \pmod{p-1}, \quad n > p.$$

Thus as in (10) we have that the number of Carmichael numbers of the first class is less than

$$(11) \quad x \sum_{\substack{p < x^{\frac{1}{6}}}} \frac{1}{p(p-1)} < x^{\frac{5}{6}}.$$

Let now n be a Carmichael number of the second class. Write

$$n = p_1 p_2 \dots p_k, \quad x^{\frac{1}{6}} \cong p_1 > p_2 > \dots > p_k.$$

Assume $n > x^{\frac{2}{3}}$. Define

$$p_1, p_2, \dots, p_{i-1} \leq x^{\frac{1}{2}} < p_1 p_2 \dots p_i \leq x^{\frac{2}{3}} < n.$$

Now

$$(12) \quad n \equiv 0 \pmod{p_1 \dots p_i}, \quad n \equiv 1 \pmod{f(p_1 \dots p_i)}, \quad n > p_1 \dots p_i \\ \text{(i. e. } n \equiv 1 \pmod{p_j - 1}), \quad 1 \leq j \leq i).$$

Let k be any integer satisfying $x^{\frac{1}{2}} < k \leq x^{\frac{2}{3}}$. We have from (12) (as in (10)) that the number of Carmichael numbers of the second class is less than

$$(13) \quad x^{\frac{2}{3}} + x \sum' \frac{1}{kf(k)},$$

where the dash indicates that $x^{\frac{1}{2}} < k \leq x^{\frac{2}{3}}$. Now we have to estimate

$$(14) \quad \sum' \frac{1}{kf(k)} = \sum_1' + \sum_2',$$

where in $\sum_1'$ $f(k) > \exp(c_{10} \log x \log_3 x / \log_2 x)$. Clearly

$$(15) \quad \sum_1' < \exp(-c_{10} \log x \log_3 x / \log_2 x) \sum_{k < x} \frac{1}{k} < \exp(-c_{10}/2 \cdot \log x \log_3 x / \log_2 x).$$

Next we estimate $\sum_2'$. We have by Lemma 2 that the number of $k < y$ ($y > x^{\frac{1}{2}}$) satisfying $f(k) < \exp(c_{10} \log x \log_3 x / \log_2 x)$ is for sufficiently small c_{10} less than

$$(16) \quad y \cdot \exp(-c_9 \log y \log_3 y / \log_2 y) \cdot \exp(c_{10} \log x \log_3 x / \log_2 x) < \\ < y \cdot \exp(-c_{11} \log x \log_3 x / \log_2 x).$$

Thus from (16) we have

$$(17) \quad \sum_2' \frac{1}{kf(k)} < \sum_2' \frac{1}{k} < \exp(-c_{11} \log x \log_3 x / \log_2 x) \sum_{k < x} \frac{1}{k} < \\ \exp\left(-\frac{c_{11}}{2} \log x \log_3 x / \log_2 x\right).$$

From (13), (14), (15) and (17) we have that the number of Carmichael numbers of the second class is less than

$$(18) \quad x \cdot \exp(-c_{12} \log x \log_3 x / \log_2 x).$$

(11) and (18) prove (6)

Thus we only have to prove Lemma 2. $f(k) = t$ implies that all prime factors p of k satisfy $(p-1) | t$. Denote by $p_1, p_2, \dots$ all primes for which $(p-1) | t$. Then the number of solutions of $f(k) = t$, $k \leq y$ is clearly not greater than the number of integers $\leq y$ composed of the primes $p_1, p_2, \dots$. Denote by $q_1, q_2, \dots$ the primes among the p_j 's not exceeding $\exp((\log_2 y)^2 / \log_3 y)$ and by $r_1 < r_2 < \dots$ the primes among the p_j 's greater than $\exp((\log_2 y)^2 / \log_3 y)$. Put $k = QR$ where Q is composed entirely of the q 's and R is composed entirely of the r 's. Consider

$$f(k) = t, \quad y^{\frac{1}{2}} < k \leq y.$$

Then clearly either $Q > y^{\frac{1}{4}}$ or $R > y^{\frac{1}{4}}$. Thus the number of solutions of

$f(k) = t$ is clearly not greater than (the dashes indicate that $y^{\frac{1}{4}} < Q < y$, $y^{\frac{1}{4}} < R < y$)

$$(19) \quad y^{\frac{1}{2}} + y \left(\sum' \frac{1}{Q} + \sum' \frac{1}{R} \right)$$

From Lemma 1 we have that the number of integers $Q < z$, $y^{\frac{1}{4}} < z < y$ is less than

$$(20) \quad z \cdot \exp(-c_{13} \log y \log_3 y / \log_2 y).$$

(The u of Lemma 1 here equals $c_{14} \log y \log_3 y / (\log_2 y)^2$). Hence by (20)

$$(21) \quad \sum' \frac{1}{Q} < \exp(-c_{14} \log y \log_3 y / \log_2 y) \sum_{l=y}^{\frac{1}{l}} < \exp(-c_{14}/2 \cdot \log y \log_3 y / \log_2 y).$$

Now we estimate $\sum' \frac{1}{R}$. The r 's are the primes $\leq y$ greater than $\exp \cdot ((\log_2 y)^2 / \log_3 y)$ which satisfy $(r_i - 1) | t$. First we show that (P_i is the i -th prime)

$$(22) \quad r_i > (2i \log i)^{1+\alpha} > P_i^{1+\alpha}, \quad \alpha = c_{15} \log_3 y / \log_2 y.$$

Let $s_1, s_2, \dots, s_j$ be all the prime factors of t . Then the number of r 's not exceeding r_i (which of course equals i) is not greater than $N(s_1, s_2, \dots, s_j; r_i)$. Now⁵⁾ $t < k \leq y$ thus $j < \log y$. Put $r_i = (\log y)^{u_i}$. Then we have as in the proof of Lemma 1 by the Theorem of DE BRUIJN⁴⁾ ($\pi((\log y)^2) > \log y$)

$$(23) \quad i \leq N(s_1, s_2, \dots, s_j; (\log y)^{u_i}) < \psi((\log y)^{u_i}, (\log y)^2) < < (\log y)^{u_i} \exp(-c_{16} u_i \log u_i)$$

($r_i \leq y$, thus Lemma 1 applies).

Now $r_i > \exp((\log_2 y)^2 / \log_3 y)$, thus $u_i > \log_2 y / \log_3 y$, hence (22) follows from (23) by a simple computation.

Now $R_1, R_2, \dots$ are the integers composed of the r 's. We have by (22) $R_i > i^{1+\alpha}$. Thus

$$(24) \quad \sum' \frac{1}{R} = \sum_{R_i > y^{\frac{1}{4}}} \frac{1}{R_i} = \Sigma_1 + \Sigma_2$$

where in Σ_1 $R_i > y^{\frac{1}{4}}$, $i \leq y^{\frac{1}{8}}$ and in Σ_2 $i > y^{\frac{1}{8}}$ (Σ_1 may be empty.) We evidently have

$$(25) \quad \Sigma_1 < y^{\frac{1}{8}} \cdot \frac{1}{y^{\frac{1}{4}}} = \frac{1}{y^{\frac{1}{8}}}, \quad \Sigma_2 < \sum_{i > y^{\frac{1}{8}}} \frac{1}{i^{1+\alpha}} < c_{16} (\alpha y^{\frac{1}{8}})^{-1} < < \exp(-c \log y \cdot \log_3 y / \log_2 y).$$

⁵⁾ Namely $t = f(k) \leq \varphi(k) < k$.

From (24) and (25)

$$(26) \quad \sum' \frac{1}{R} < \exp(-c_{17} \log y \cdot \log_3 y / \log_2 y).$$

(19), (21) and (26) complete the proof of Lemma 2.

Let now $A = p_1 p_2 \dots p_k$ be the product of the consecutive primes less than $\varepsilon \log x$, and denote by $r_1, r_2, \dots$ the primes for which $(r_i - 1) | A$. It is easy to see that $A < x^{2\varepsilon}$, for sufficiently large x . It is reasonable to expect that for $u < (\log x)^{c_{18}}$ there are more than $c_{19} \pi(u)$ ($c_{19} = c_{19}(c_{18})$) r 's not exceeding u (though this will probably be very hard to prove). If the preceding statement is true, then a simple computation shows that there are more than $x^{1-\varepsilon}$ composite, squarefree integers $n \leq x$ composed entirely of the r 's. Again it is reasonable to assume that these numbers are roughly equidistributed (mod A), and thus one can assume that there are more than $x^{1-4\varepsilon}$ composite squarefree integers less than $x \equiv 1 \pmod{A}$ which are all composed of the r 's.

Let n be such a number, then n is clearly a pseudoprime, since all prime factors of n are r 's, $n \equiv 1 \pmod{A}$, $(r_i - 1) | A$. Thus, if all the above conjectures are true, $\log(C(x))/\log x \rightarrow 1$.

In a previous paper⁶⁾ I proved that for a suitable infinite sequence x_i the number of solutions of $\varphi(n) = x_i$ is greater than $x_i^{c_{20}}$, where $\varphi(n)$ stands for the Euler-function. The above arguments (using only the first conjecture) would imply that c_{20} can be taken as close to 1 as we please. By arguments similar to those used in proving (6) I can show that the number of solutions of $\varphi(n) = x$ is less than

$$x \exp(-c_{21} \log x \cdot \log_3 x / \log_2 x).$$

I can further show that for any ε, l and $x > x_0(\varepsilon, k)$

$$\frac{x^2}{\log x} (\log_2 x)^l < \sum_{k=1}^x f(k) < \frac{x^2}{\log x} (\log x)^\varepsilon;$$

and if one neglects a set of integers n_i of density 0 then for every $\varepsilon > 0$

$$\log n - (1 + \varepsilon) \log_2 n \log_3 n < \log(f(n)) < \log n - (1 - \varepsilon) \log_2 n \log_3 n.$$

Thus, in particular, for almost all n and every c

$$f(n) = o\left(\frac{n}{(\log n)^c}\right).$$

(Received September 5, 1955.)

⁶⁾ *Quarterly J. Oxford Ser. 6* (1935), 211–213.