FERMAT PSEUDOPRIMES

SHUGUANG LI AND CARL POMERANCE

Abstract. We give an upper bound for the distribution of base-a pseudoprimes that is uniform in the base and does not require coprimality to the base. In addition we show that there are infinitely many "near Carmichael numbers" meaning that they are pseudoprimes for a positive proportion of bases, but not all bases.

1. Introduction

Fermat's "little theorem" asserts that

$$a^n \equiv a \pmod{n} \tag{1}$$

whenever n is prime and a is an integer. If (1) holds with n composite, then we say that n is a *base-a Fermat pseudoprime*. Since it is computationally easy to compute $a^n \pmod{n}$ via a powermod algorithm, one often checks (1) with a large number n which is not known to be prime or composite. If the congruence holds one can suspect that n is prime, and if it does not hold, one knows that n is composite. This test is not useful when $a = 0$ or 1, since every n satisfies the congruence, nor is it interesting in the case $a = -1$, since every odd n passes. In this paper we shall assume that $a \geq 2$.

Historically the special case $a = 2$ was tacitly assumed, and base-2 Fermat pseudoprimes were simply referred to as pseudoprimes. We have long known that there are infinitely many, in fact, if $p > 3$ is prime, then $(4^p - 1)/3$ is a pseudoprime. Another classical proof of their infinitude is based on the fact that if n is a pseudoprime, so is $2^n - 1$. There are numbers that are Fermat pseudoprimes to every base, these are the Carmichael numbers. The best result currently known is that there are at least $x^{0.3389}$ Carmichael numbers up to x once x is sufficiently large, see [4]. So in fact there are many more pseudoprimes than suggested by the historical arguments.

The existence of pseudoprimes and Carmichael numbers seems to invalidate using (1) for distinguishing between primes and composites.

Date: April 10, 2025.

2010 *Mathematics Subject Classification.* 11A51, 11N25, 11N32, 11Y11.

Key words and phrases. pseudoprime, Carmichael number, multiplicative order.

But the ease of checking (1) suggests we should not give up on it. In fact, Erdős showed that though there are infinitely many pseudoprimes, there are far fewer of them up to a large number x than there are primes, suggesting that if a random n satisfies (1) it is highly likely to be prime, see [5]. Currently the best upper bound known for the number of odd pseudoprimes $\le x$ is

$$x^{1-\frac{1}{2}\log\log\log x/\log\log x}$$

for x sufficiently large, see [7]. It is conjectured that the "correct" bound is the same expression but with $\frac{1}{2}$ replaced with $1+o(1)$. This has been proved for Carmichael numbers and a heuristic argument suggests that this estimate is tight (as well as for Fermat pseudoprimes to any fixed base).

Though it is surely not difficult in determining whether an even number is prime or composite, we point out that the above upper bound was proved only for odd pseudoprimes. Further, one may ask about the more general problem of the distribution of base-a Fermat pseudoprimes. Let $P_a(x)$ denote the number of base-a Fermat pseudoprimes $\le x$. Our first result shows that in a wide range for a there is a universal upper bound of the same quality as previously shown for the case of odd pseudoprimes.

Theorem 1. *There is a number x_0 such that*

$$P_a(x) \le x^{1-\frac{1}{2}\log\log\log x/\log\log x}$$

for all $x \ge x_0$ and $2 \le a \le x$.

We remark that it follows from an argument of Beeger [1] that if there is one Fermat pseudoprime n base a with $\gcd(n,a) = d$, then there are infinitely many such numbers n. Here is a proof. For positive, coprime integers a, n let $\ell_a(n)$ denote the exponent that a belongs to modulo n. Suppose that n satisfies the above conditions. We know from Bang's theorem that there is some prime p with $\ell_a(p) = n-1$. Since $\gcd(p,a) = 1$ we have $\gcd(pn,a) = d$. Also, $pn \equiv 1 \pmod{n-1}$ so that $a^{pn} \equiv a \pmod{p}$. From $a^n \equiv a \pmod{n}$, we have $a^k \equiv a \pmod{n}$ for any $k \equiv 1 \pmod{n-1}$. Thus, $a^{pn} \equiv a \pmod{n}$, and we conclude that $a^{pn} \equiv a \pmod{pn}$ as claimed. See [8] for a discussion on what the true magnitude of the count of these pseudoprimes may be.

Let $\mathcal{F}(n) = \{a \pmod n : a^{n-1} \equiv 1 \pmod n\}$. It is easy to see that $\mathcal{F}(n)$ is a subgroup of $(\mathbb{Z}/n\mathbb{Z})^*$. After Monier and Baillie–Wagstaff, we know that $F(n) := \#\mathcal{F}(n) = \prod_{p|n}(p-1, n-1)$. Now let

$$\mathcal{F}^*(n) = \{a \pmod n : a^n \equiv a \pmod n\}, \quad F^*(n) = \#\mathcal{F}^*(n).$$

As noted in [8],

$$F^*(n) = \prod_{p|n}(1 + (p-1, n-1)).$$

We have $F^*(n) = n$ if and only if n is prime, n is a Carmichael number, or $n = 1$. Further, if $F^*(n) < n$, then $F^*(n) \le \frac{3}{5}n$ except that $F^*(6) = 4$, see [2, Exercise 3.14]. It is asked there if $F^*(n) = \frac{3}{5}n$ infinitely often, and if there is a positive number ϵ with $1 > F^*(n)/n > \epsilon$ infinitely often. We prove this latter assertion.

Theorem 2. *There are infinitely many numbers n with*

$$1 > F^*(n)/n > 1/2451.$$

2. A preliminary result

By way of notation, let $\log_k$ denote the k-fold iterated natural logarithm. We reserve the letter p for prime numbers.

We prove an analogue of [7, Theorem 1].

Theorem 3. *For all sufficiently large numbers x we have*

$$\#\{m \le x : (a, m) = 1,\, \ell_a(m) = n\} \le x^{1-(3+\log_3 x)/\log_2 x}$$

for all $n \ge 1$ and $2 \le a \le x$.

Proof. Since $\ell_a(m) \le m \le x$, we may assume that $n \le x$. For a number c with $0 < c < 1$, we have

$$\sum_{\substack{m \le x \\ \ell_a(m)=n}} 1 \le x^c \sum_{\ell_a(m)=n} m^{-c} \le x^c \sum_{p|m \Longrightarrow \ell_a(p)|n} m^{-c} = x^c \prod_{\ell_a(p)|n} (1 - p^{-c})^{-1}.$$

Denote this last product by A and choose $c = 1 - (4 + \log_3 x)/(2\log_2 x)$. So, it suffices to show that A is fairly small. In fact, we will show that $\log A < (\log x)^{3/4}$, which is sufficient for the theorem.

Assume that x is large enough that $c > 7/8$. Then

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

The different primes $q_1, \dots, q_t$ with $\ell_a(q_j) = d$ are all divisors of $a^d - 1$ and are all $\equiv 1 \pmod d$. A crude upper bound for t is $d \log a \le d \log x$. Thus,

$$\sum_{\ell_a(p)=d} p^{-c} = \sum_{j \le t} q_j^{-c} \le \sum_{j \le t} (dj + 1)^{-c} < d^{-c} \sum_{j \le t} j^{-c} < d^{-c}(1 - c)^{-1} t^{1-c}.$$

Putting in our bound for t, we find that

$$\log A < (1-c)^{-1}(\log x)^{1-c}\sum_{d\mid n} d^{1-2c} + O(1).$$

Using that $\prod_{p\le 2\log x} p > x$, the sum over $d \mid n$ is less than

$$\prod_{p\mid n}(1-p^{1-2c})^{-1} \le \prod_{p\le 2\log x}(1-p^{1-2c})^{-1} = e^{O(\log_2 x/\log_3 x)}.$$

Thus, when x is large enough,

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

and so

$$\log A < (1-c)^{-1}(\log x)^{1-c+1/2} + O(1) \le (\log x)^{3/4}.$$

This completes the proof. □

3. An upper bound

In this section we prove Theorem 1.

As in [7] we consider numbers n with various, perhaps overlapping properties. Let

$$L(x) = x^{\log_3 x/\log_2 x};$$

our goal is to prove that $P_a(x) \le x/L(x)^{1/2}$. Suppose that (1) holds for a, n, where $a, n \le x$ and n is composite. Write $n = uv$ where $u = (a, n)$. Since (1) holds, we have $(a, v) = 1$. At least one of the following conditions holds:

(i) $v \le x/L(x)^2$,
(ii) there is a prime $p \mid v$ with $p > L(x)^3$ and $\ell_a(p) \le L(x)$,
(iii) there is a prime $p \mid v$ with $\ell_a(p) > L(x)$,
(iv) v has a divisor d with $x/L(x)^5 < d \le x/L(x)^2$.

Indeed, if (ii) and (iii) both fail, then every prime factor p of v has $p \le L(x)^3$. Thus, if (i) also fails, then (iv) must hold.

Suppose that (i) holds. There are at most $x/L(x)$ numbers $n \le x/L(x)$. Assume that $n > x/L(x)$. Then $u > L(x)$. The number of $n \le x$ divisible by a divisor u of a with $u > L(x)$ is at most $\tau(a)x/L(x)$, where $\tau(a)$ denotes the total number of divisors of a. We know after Wigert that $\tau(a) \le x^{(\log 2+o(1))/\log_2 x}$, so that $\tau(a) \le L(x)^{o(1)}$. Thus, the number of n satisfying (i) is at most $x/L(x)^{1+o(1)}$.

As in the proof of Theorem 3, the number of primes p with $\ell_a(p) = k$ is at most $k \log x$. The number of $n \le x$ in case (ii) is thus at most

$$\sum_{\substack{p > L(x)^3 \\ \ell_a(p) < L(x)}} \frac{x}{p} < \frac{x}{L(x)^3} \sum_{k < L(x)} k \log x < \frac{x \log x}{L(x)}.$$

For n satisfying (1) and $p \mid v$, we have

$$n \equiv 0 \pmod{p} \quad \text{and} \quad n \equiv 1 \pmod{\ell_a(p)},$$

so that $n \equiv p \pmod{p\ell_a(p)}$. Since n is composite, we have $n > p$, so the number of n is $< x/p\ell_a(p)$. Thus, the number of n in case (iii) is less than

$$\sum_{\substack{p \le x \\ \ell_a(p) > L(x)}} \frac{x}{p\ell_a(p)} < \frac{x}{L(x)} \sum_{p \le x} \frac{1}{p} \ll \frac{x \log\log x}{L(x)}.$$

Let $I = (x/L(x)^5, x/L(x)^2]$. The number of n in case (iv) is at most

$$\sum_{\substack{d \in I \\ n \equiv 0 \pmod{d} \\ n \equiv 1 \pmod{\ell_a(d)}}} 1 \le \sum_{d \in I} \left(1 + \frac{x}{d\ell_a(d)}\right) \le \frac{x}{L(x)^2} + x \sum_{d \in I} \frac{1}{d\ell_a(d)}$$

$$= \frac{x}{L(x)^2} + x \sum_{m \le x} \frac{1}{m} \sum_{\substack{d \in I \\ \ell_a(d) = m}} \frac{1}{d}.$$

Let $C(t) = t^{-(3+\log_3 t)/2\log_2 t}$. Note that $C(t)$ is decreasing for large t. We use Theorem 3 and partial summation on the inner sum, getting

$$\sum_{\substack{d \in I \\ \ell_a(d) = m}} \frac{1}{d} \ll C(x/L(x)^5) \log x < x^{-(2+\log_3 x)/2\log_2 x}$$

for x sufficiently large. Thus, from the above, the count in case (iv) is at most $x^{1-(2+\log_3 x)/2\log_2 x} \log x < x^{1-(1+\log_3 x)/2\log_2 x}$. Collecting the estimates in the various cases completes the proof of Theorem 1.

4. Proof of Theorem 2

We will show there are two different positive integers $a, b \le 50$ with $ak+1, bk+1$ simultaneously prime infinitely often. Theorem 2 follows, for if p, q are the two primes, then

$$\begin{aligned} F^*(pq) &= (1 + (p-1, pq-1))(1 + (q-1, pq-1)) \\ &= (1 + (p-1, q-1))^2 \ge (1+k)^2 \end{aligned}$$

and $\lim_{k\to\infty}(1+k)^2/(ak+1)(bk+1) = 1/ab \geq 1/49 \cdot 1/50 = 1/2450$. It remains to note that pq is not a Carmichael number since every Carmichael number has at least 3 prime factors.

To show that $ak+1, bk+1$ are both prime infinitely often we use a generalization of Zhang's theorem, as improved by the Polymath project, on small gaps between primes: For any distinct positive integers $a_1, \dots, a_{50}$, there are two of them a_i, a_j with a_ik+1, a_jk+1 both prime infinitely often, see Granville [3, p. 175] and Maynard [6]. (Note that the linear forms a_ik+1 form an "admissible" set since their product at $k=0$ is 1.)

Acknowledgments

We are grateful to Andrew Granville, Jared Duker Lichtman, and James Maynard for their help.

References

[1] N. G. W. H. Beeger, On even numbers m dividing 2^m-2, Amer. Math. Monthly **58** (1951), 553–555.
[2] R. E. Crandall and C. Pomerance, Prime numbers: a computational perspective, second ed., Springer, New York, 2005.
[3] A. Granville, Primes in intervals of bounded length, *Bull. Amer. Math. Soc.* **52** (2015), 171–222.
[4] J. D. Lichtman, Primes in arithmetic progressions to large moduli, and shifted primes without large prime factors, arXiv:2211.09641 [math.NT].
[5] J. D. Lichtman and C. Pomerance, Improved error bounds for the Fermat primality test on random inputs, Math. Comp. **87** (2018), 2871–2890.
[6] J. Maynard, Gaps between primes, 2018 ICM Proceedings, 345–361. https://doi.org/10.1142/9789813272880_0057
[7] C. Pomerance, On the distribution of pseudoprimes, *Math. Comp.* **37** (1981), 587–593.
[8] C. Pomerance and S. S. Wagstaff, Jr., Some thoughts on pseudoprimes, *Bulletin, Classe des Sciences Mathématiques et Naturelles, Sciences mathématiques* **46** (2021), 53–72.

Department of Mathematics, Natural Sciences Division, University of Hawaii-Hilo, 200 W. Kawili Street, Hilo, HI 96720-4091, USA
E-mail address: shuguang@hawaii.edu

Mathematics Department, Dartmouth College, Hanover, NH 03755, USA
E-mail address: carlp@math.dartmouth.edu