ON THE NUMBER OF FALSE WITNESSES FOR A COMPOSITE NUMBER[1]

Paul Erdős
Mathematical Institute of the Hungarian Academy of Sciences
Budapest, Hungary

Carl Pomerance[2]
University of Georgia
Athens, Georgia 30602/USA

When presented with a large number n which one would like to test for primality, one usually begins with a modicum of trial division. If n is not revealed as composite, the next step is often to perform the simple and cheap test of computing $a^{n-1} \bmod n$ for some pre-chosen number $a > 1$ with $(a,n) = 1$. If this residue is not 1, then n is definitely composite (by Fermat's little theorem) and we say a is a *witness* for n. If the residue is 1, then n is probably prime, but there are exceptions. If we are in this exceptional case where

$$a^{n-1} \equiv 1 \bmod n \text{ and } n \text{ is composite}$$

then we say a is a *false witness* for n, or equally, that n is a *pseudoprime* to the base a.

The problem of distinguishing between pseudoprimes and primes has been the subject of much recent work. For example, see [4].

Let

$$\mathcal{F}(n) = \{a \bmod n : a^{n-1} \equiv 1 \bmod n\}, \quad F(n) = \#\mathcal{F}(n) .$$

Thus, if n is composite, then $\mathcal{F}(n)$ is the set (in fact, group) of residues mod n that are false witnesses for n and $F(n)$ is the number of such residues. If n is prime, then $F(n) = n - 1$ and $\mathcal{F}(n)$ is the entire group of reduced residues mod n. For any n, Lagrange's theorem

---

[1]Extended abstract, details to appear elsewhere.

[2]Research supported in part by an NSF grant.

gives

$$F(n) \mid \phi(n)$$

where ϕ is Euler's function.

There are composite numbers n for which $F(n) = \phi(n)$, such as $n = 561$. Such numbers are called Carmichael numbers and probably there are infinitely many of them, but this has never been proved. It is known that Carmichael numbers are much rarer than primes.

At the other extreme, there are infinitely many numbers n for which $F(n) = 1$. For example, any number of the form $2p$ will do, where p is prime. It is possible to show that while these numbers n with $F(n) = 1$ have asymptotic density 0 , they are much more common than primes.

So what is the normal and/or average behavior of the function $F(n)$? It is to these questions that this paper is addressed. We show (where $\sum'$ denotes a sum over composite numbers)

$$\text{(1)} \qquad \frac{1}{x} \sum_{n \le x}' F(n) > x^{15/23}$$

for x large and

$$\text{(2)} \qquad \frac{1}{x} \sum_{n \le x}' F(n) \le x \exp\{-(1+o(1)) \log x \text{ logloglog } x/\text{loglog } x\}$$

as $x \to \infty$. We conjecture that equality holds in (2). Our proof of the lower bound (1) uses recent work of Balog [2] on the distribution of primes p such that all primes in p-1 are small. With continued improvements expected on this kind of result, the exponent 15/23 will probably "creep up" towards 1.

Let $L(x) = \exp(\log x \text{ logloglog } x/\text{loglog } x)$. Let $P_a(x)$ denote the number of $n \le x$ such that n is a pseudoprime to the base a . Thus $P_a(x)$ is the number of composite $n \le x$ with $a \bmod n \in \mathcal{F}(n)$. For a fixed value of a , the sharpest results known on $P_a(x)$ are that

$$\exp\{(\log x)^{5/14}\} < P_a(x) < x\,L(x)^{-1/2} \tag{3}$$

for all $x \geq x_0(a)$ - see [5], [6]. (Using Balog's result, we may replace the "5/14" in the lower bound with 15/38.) We trivially have

$$\sum_{a \leq x} P_a(x) \geq \sum_{n \leq x}{}' F(n) .$$

On the other hand

$$\sum_{a \leq x} P_a(x) \leq \sum_{n \leq x}{}' \sum_{\substack{a \leq x \\ a^{n-1} \equiv 1 \bmod n}} 1$$

$$\leq \sum_{n \leq x}{}' F(n)\left(\frac{x}{n} + 1\right) .$$

Thus, by using partial summation and (1), (2) we can obtain a result that is, *on average*, much better than (3):

$$x^{15/23} < \frac{1}{x} \sum_{a \leq x} P_a(x) \leq x\,L(x)^{-1+o(1)}$$

for x large.

We can compute the geometric mean value of $F(n)$ with more precision: there are positive constants c_1, c_2 such that

$$\left(\prod_{n \leq x} F(n)\right)^{1/x} = c_2(\log x)^{c_1} + o(1)$$

as $x \to \infty$. If the geometric mean is taken just for composite numbers, then the result is the same except that c_2 is replaced by c_2/e.

Concerning the normal value of $F(n)$, we show that $\log F(n)/\log\log n$ has a distribution function $D(u)$. That is, $D(u)$ is the asymptotic density of the integers n for which

$$F(n) \leq (\log n)^u .$$

The function $D(u)$ is continuous, strictly increasing, and singular on $[0,\infty)$. Moreover, $D(0) = 0$ and $D(+\infty) = 1$. Thus, for example, the set of n with $F(n) = 1$ has density 0.

The starting point for our results is the elegant and simple formula of Monier [3] and Baillie-Wagstaff [1]:

$$F(n) = \prod_{p|n} (p-1,\ n-1) \tag{4}$$

where p denotes a prime. For example, (4) immediately implies $F(2p) = 1$.

We are also able to prove analogous results for certain pseudoprime tests more stringent than the Fermat congruence, namely the Euler test and the strong pseudoprime test. It is to be expected that there will be similar results for all Fermat-type tests; for example, the Lucas tests. Such an undertaking might gain useful insights into the nature of these tests.

Finally we address some further questions including the maximal order of $F(n)$ for n composite, the nature of the range of F, the normal number of prime factors of $F(n)$, and the universal exponent for the group $\mathcal{F}(n)$.

Acknowledgement. The second author gratefully acknowledges the hospitality of the Discrete Mathematics Department at Bell Communications Research, Inc., where much of the work for this paper was done.

References

[1] R. Baillie and S.S. Wagstaff, Jr., Lucas pseudoprimes, Math. Comp. 35 (1980), 1391-1417.

[2] A. Balog, $p + a$ without large prime factors, Séminaire de Théorie des Nombres de Bordeaux (1983-84), no. 31.

[3] L. Monier, Evaluation and comparison of two efficient probabilistic primality testing algorithms, Theoretical Comp. Sci. 12 (1980), 97-108.

[4] C. Pomerance, Recent developments in primality testing, Math. Intelligencer 3 (1981), 97-105.

[5] C. Pomerance, On the distribution of pseudoprimes, Math. Comp. 37 (1981), 587-593.

[6] C. Pomerance, A new lower bound for the pseudoprime counting function, Illinois J. Math. 26 (1982), 4-9.