Pseudoprime Statistics to 10^{19}

Jens Kruse Andersen and Harvey Dubner

CONTENTS

1. Introduction
2. Background Information
3. Results
References

A base-b pseudoprime (psp) is a composite N satisfying $b^{N-1} \equiv 1 \pmod{N}$. We use computer searches to count odd base-3 psp near 10^n for n up to 19. The counts indicate that a good approximation to the probability of a random odd number near z being a psp is $P(z) = z^{-0.59}$. Integrating P yields a psp-counting function, $Q(x) = (x^{0.41})/0.82$, which gives estimated counts close to known actual counts up to 10^{19}, although these estimates are probably not valid for all x.

A table comparing pseudoprime counts up to 10^{11} for bases 2, 3, 5, 7, 11, 13, 17, is included.

1. INTRODUCTION

Fermat's "little theorem" states that if b is an integer prime to N, and if N is prime, then

$$b^{N-1} \equiv 1 \pmod{N}.$$

Unfortunately, this is not a sufficient condition for a number to be prime, since there are composite numbers that satisfy this congruence. Such a number is called a base-b pseudoprime, or psp(b). In fact, there are composite numbers called Carmichael numbers that satisfy this congruence for all values of b that are relatively prime to N.

Because of the existence of pseudoprimes, a successful Fermat test can indicate only that N is a probable prime, and further analysis is required to certify true primality. For almost all primes, certification takes more than 1000 times longer than a Fermat test. However, there are statistically oriented projects in number theory that require only a high probability of primality rather than absolute certainty.

The book [Ribenboim 04] has an excellent discussion of pseudoprimes, with many references. Included is a table of counts of base-2 pseudoprimes up to 10^{13}. Also included are several asymptotic formulas for upper and lower bounds for psp(b) for large values of N such as 10^{100} and 10^{1000}, but these formulas are far too inaccurate to be useful for $N < 10^{20}$.

2000 AMS Subject Classification: Primary 11A99

Keywords: Psp, pseudoprime

1058-6458/2007 $0.50 per page

The main objective of this article is to heuristically extend the pseudoprime statistics to 10^{19}, which seems to be as high as currently can be accomplished in a reasonable amount of time.

2. BACKGROUND INFORMATION

2.1 Past Results

Let $P\pi_b(x)$ denote the number of pseudoprimes to the base b less than or equal to x. Table 1 includes data for $P\pi_2(x)$ up to 10^{13} from [Ribenboim 04, p. 219]. The counts for 10^{12} and 10^{13} are from [Pinch 00]. In column 3 each count is converted to the probability that a random odd number below x is a base-2 pseudoprime.

Pomerance [Pomerance 81] has shown that for all large x,

$$P\pi_b(x) \leq \frac{x}{L(x)^{1/2}},$$

where

$$L(x) = e^{\log x \log\log\log x/\log\log x}.$$

This formula gives upper bounds that may be useful for large values of x, but unfortunately seems of little practical value for relatively small x. For example, it predicts that $P\pi_2(10^{13}) < 45{,}740{,}536{,}510$ compared to the actual value of 264,239.

2.2 Pseudoprime Program

The goal of our program is to estimate the probability that a random odd integer near a given number is a base-b pseudoprime. This is done by finding all odd psp(b) in chosen intervals large enough to give a reasonable statistical base. The probability is simply the ratio of found pseudoprimes to odd numbers.

The input to the program is the base b and the interval to be tested. Only composites can be pseudoprimes, so the primes are eliminated by sieving to the square root of the interval's end. For efficiency, a Fermat test is not performed on all odd composites.

Let $l_b(p)$ be the least positive integer h for which $b^h \equiv 1 \pmod{p}$. Proposition 3 in [Pomerance et al. 80] states that if the prime p divides the psp(b) n, then $n \equiv 1 \pmod{l_b(p)}$, and hence $n \equiv p \pmod{p \cdot l_b(p)}$.

We precomputed $l_b(p)$ for all primes p below a chosen limit. During the sieving phase, each of those p is processed by all odd numbers that are divisible by p but not congruent to $p \pmod{p \cdot l_b(p)}$ being marked "known not to be psp(b)." In total, around 85% of odd composites are eliminated as psp(b) candidates. A base-b Fermat test is performed only on the remaining odd numbers. This produces all psp(b) in the chosen interval.

Avoiding Fermat tests on most numbers means that much of the time is spent on sieving. To optimize this, the program uses as much RAM as the user allows, and a precomputed compressed file of all 32-bit primes. Only 32-bit and a little 64-bit arithmetic is used for sieving. This limits the possible size to 2^{64}, which is near 1.84×10^{19}. In any case, finding enough pseudoprimes above 10^{19} for reasonable statistics would be very time-consuming.

2.3 Choosing a Base

Is there an optimal base that should be used for Fermat testing? Such a base would produce a minimal number of pseudoprimes. Since we do not know of any analytic approach to finding such a base, we tested many bases by counting the actual number of pseudoprimes up to 10^9. The results for a variety of bases are shown in Table 2.

x	$P\pi_2(x)$	Probability
10^3	3	6.00×10^{-3}
10^4	22	4.40×10^{-3}
10^5	78	1.56×10^{-3}
10^6	245	4.90×10^{-4}
10^7	750	1.50×10^{-4}
10^8	2057	4.11×10^{-5}
10^9	5597	1.12×10^{-5}
10^{10}	14884	2.98×10^{-6}
10^{11}	38975	7.80×10^{-7}
10^{12}	101629	2.03×10^{-7}
10^{13}	264239	5.28×10^{-8}

TABLE 1. Psp(2) counts for 10^3 to 10^{13}.

Base	# Psp	Base	# Psp
2	5597	17	5346
3	5767	19	6529
4	10173	23	7048
5	5146	29	5513
6	6204	32	17925
7	4923	64	26682
8	14629	$3^3 = 27$	12823
9	8670	$3^4 = 81$	12526
10	5599	$3^5 = 243$	16944
11	5020	$3^6 = 729$	19412
12	7781	$3^7 = 2187$	17018
13	5082	101	7464
14	5848	103	7300
15	4665	105	5584
16	13422	$8! = 40320$	7115
		prime 10000019	8562

TABLE 2. Psp counts for odd numbers up to 10^9 for various bases.

x	$P\pi_2(x)$	$P\pi_3(x)$	$P\pi_5(x)$	$P\pi_7(x)$	$P\pi_{11}(x)$	$P\pi_{13}(x)$	$P\pi_{17}(x)$
10^3	3	5	3	5	8	7	6
10^4	22	22	17	15	23	21	23
10^5	78	76	66	69	79	81	75
10^6	245	243	238	229	236	257	243
10^7	750	749	726	651	672	719	741
10^8	2057	2131	1910	1782	1891	1929	2025
10^9	5597	5767	5146	4923	5020	5082	5346
10^{10}	14884	15404	13634	13019	12953	13373	14030
10^{11}	38975	39751	35635	33898	33561	34112	36010

TABLE 3. Comparing bases 2, 3, 5, 7, 11, 13, 17 psp counts for odd numbers.

From these data it seems that choosing any small prime as the base for the Fermat test is reasonable. However, care must be taken when specific forms of numbers are being tested rather than random numbers. For example, Mersenne numbers, 2^p-1 with p prime, all satisfy the Fermat test with base 2, although only a small number of them are truly prime. In fact, generalized repunits, $(b^p-1)/(b-1)$ with p prime, almost always satisfy the Fermat test with base b, even though only a few are truly prime.

Counts of pseudoprimes for the first seven prime bases are compared in Table 3. Once again we do not have a reason to prefer one base over another.

We will arbitrarily use base 3 for the remainder of this article because it is often used by prime searchers.

2.4 Poisson Distribution

Although the exact distribution of counts of pseudoprimes is not known, it seems reasonable to assume that such counts might be approximated by a Poisson distribution, since this is true of many distributions of rare phenomena. For the Poisson distribution the standard deviation is the square root of the psp count. If the count of pseudoprimes can be approximated by a Poisson distribution, then this can be used to estimate the accuracy of the probabilities we will measure. Note that the Poisson conjecture is not used to obtain pseudoprime counts but is used only to estimate the accuracy of such counts.

We tested the Poisson hypothesis by collecting distribution data for various sizes of pseudoprimes. Some representative results are shown in Table 4. The probability that a particular observed distribution is from a Poisson distribution is shown in column 3. These probabilities were computed using the one-sample Kolmogorov–Smernov goodness-of-fit test, and are high enough to support this Poisson conjecture for our purposes. The actual standard deviations (SD) and the computed Poisson SDs are also shown, and they are close enough to help support the conjecture.

These Poisson probabilities are encouraging enough that we are planning a future project to see how the Poisson conjecture holds for other rare phenomena such as twin primes, Sophie Germain primes, and large prime gaps.

Size	Average	K-S Prob.	Actual SD	Poisson SD
10^{10}	104.1	0.906	10.6	10.2
10^{11}	33.1	0.891	6.5	5.7
10^{12}	9.0	0.998	3.5	3.0
10^{13}	20.8	0.939	3.4	4.6

TABLE 4. Poisson distribution tests with 20 samples.

3. RESULTS

3.1 Measuring Pseudoprime Probability

We used seven PC computers averaging 2.0 GHz to collect data for estimating the probability of an odd number being a pseudoprime. Results are shown in Table 5. It took about 40 computer-days for testing numbers of size 10^{18} and 67 computer-days for 10^{19}.

Assuming that the Poisson distribution can approximate the actual psp distribution, then the standard deviation of a psp count is equal to the square root of the

Size	# Odd Numbers	# Psp	Psp(3) Probability
10^9	0.2×10^9	910	4.55×10^{-6}
10^{10}	1×10^9	1218	1.22×10^{-6}
10^{11}	5×10^9	1517	3.03×10^{-7}
10^{12}	10×10^9	863	8.63×10^{-8}
10^{13}	32×10^9	730	2.28×10^{-8}
10^{14}	192×10^9	1025	5.34×10^{-9}
10^{15}	272×10^9	362	1.33×10^{-9}
10^{16}	756×10^9	300	3.97×10^{-10}
10^{17}	663×10^9	57	8.60×10^{-11}
10^{18}	1500×10^9	47	3.13×10^{-11}
10^{19}	2000×10^9	17	8.50×10^{-12}

TABLE 5. Psp probabilities, base 3.

count. About 68% of the time the "correct" count will differ from the indicated count by less than 1 SD, and 99.7% of the time by less than 3 SD. This means that for most practical purposes the psp probabilities shown in Table 5 can be accepted as reasonably accurate.

3.2 Predicting Pseudoprime Probability

Note that the pseudoprime probability decreases by about a factor of 4 for each increase of a factor of 10 in size. Let $P(z)$ be the probability of an odd number of about the size of z being a pseudoprime. Assume that this can be approximated by $P(z) = z^{-k}$. This means that $P(10^n) = 10^{-kn}$ and $P(z)/P(10z) = 10^k$.

We computed the value of k corresponding to each measured psp(3) probability from 10^9 to 10^{19}. We were surprised to find that k varies only a little over this range, less than we expected. These results are shown in Table 6. The average k is near 0.59, so we assume $P(z) = z^{-.59}$. The tabulated prediction ratio is this $P(z)$ divided by the measured probability. This ratio is close to unity for n up to 17 and has a large deviation from unity only where few psp have been found, which is statistically reasonable. For example, at 10^{18}, if 8 fewer psp were found (about 1 SD), the prediction ratio goes from 0.77 to 0.95.

Assuming that k is constant and equal to 0.59, by integrating $P(z)$ over odd z we can derive a psp-counting function, $Q(x)$:

$$Q(x) = \frac{1}{2}\int_1^x z^{-k}dz = \frac{x^{1-k}}{2(1-k)} = \frac{x^{0.41}}{0.82}, \qquad (3\text{–}1)$$

where the factor $\frac{1}{2}$ is required to eliminate the even values of z, and the lower limit of the integral is neglected. A comparison of the predicted psp count to the actual count is shown in Table 7.

Size	Measured Probability	Computed k	Prediction Ratio
10^9	4.55×10^{-6}	0.5936	1.08
10^{10}	1.22×10^{-6}	0.5914	1.03
10^{11}	3.03×10^{-7}	0.5925	1.07
10^{12}	8.63×10^{-8}	0.5887	0.96
10^{13}	2.28×10^{-8}	0.5878	0.94
10^{14}	5.34×10^{-9}	0.5909	1.03
10^{15}	1.33×10^{-9}	0.5917	1.06
10^{16}	3.97×10^{-10}	0.5876	0.91
10^{17}	8.60×10^{-11}	0.5921	1.09
10^{18}	3.13×10^{-11}	0.5836	0.77
10^{19}	8.50×10^{-12}	0.5827	0.73

TABLE 6. Predicting psp probability. The average value of k is 0.5893.

Size	Predicted	Actual	
10^9	5973	5767	
10^{10}	15353	15404	
10^{11}	39463	39751	
10^{12}	101435	101629	(base 2)
10^{13}	260727	264239	(base 2)

TABLE 7. A comparison of the predicted psp count to the actual count.

The last two "actual" entries are taken from the base-2 data in Table 1, since such base-3 data is not available. Table 3 shows that pseudoprime counts for bases 2 and 3 are quite similar.

To summarize, we have found that the probability of a random odd number near 10^n being a base-3 pseudoprime is

$$P(10^n) = 10^{-kn} = 10^{-0.59n}.$$

The data show that this equation is quite accurate up to $n = 19$. It follows from this that we can derive the pseudoprime-counting function shown in equation (3–1). We also think that by adjusting k these equations will apply to many other bases.

We believe that these last two equations will apply to some values of $n > 19$, but we have no idea for a limit on n. Actually, based on conjectures of Erdős and Shanks concerning the count of Carmichael numbers for large x, it is highly likely that the number of base-3 pseudoprimes eventually exceeds $x^{0.5}$. See [Granville and Pomerance 01] for a full analysis and discussion of this topic.

3.3 Pseudoprime Samples

The following is a list of the 17 pseudoprimes and their factors that we found near 10^{19}:

1. $10110000003796004401 = 17 \cdot 31 \cdot 101 \cdot 181 \cdot 331 \cdot 661 \cdot 1453 \cdot 3301$
2. $10110000012794832601 = 11 \cdot 11 \cdot 19 \cdot 61 \cdot 151 \cdot 601 \cdot 4561 \cdot 174169$
3. $10220000019848498881 = 19 \cdot 271 \cdot 313 \cdot 1093 \cdot 20593 \cdot 281737$
4. $10250000001010674901 = 891772741 \cdot 11493959761$
5. $10350000050142179417 = 2863893649 \cdot 3613961033$
6. $10360000008716580301 = 1216552507 \cdot 8515867543$
7. $10360000081881824917 = 1762952077 \cdot 5876506921$
8. $10370000069286181633 = 73 \cdot 83 \cdot 353 \cdot 617 \cdot 739 \cdot 1933 \cdot 5501$
9. $10370000088337339801 = 13 \cdot 31 \cdot 379 \cdot 541 \cdot 1009 \cdot 5701 \cdot 21817$
10. $10410000086147738203 = 2281446919 \cdot 4562893837$
11. $10430000093919155905 = 5 \cdot 109 \cdot 113 \cdot 199 \cdot 463 \cdot 769 \cdot 883 \cdot 2707$
12. $10470000037434900481 = 149 \cdot 257 \cdot 307 \cdot 1871 \cdot 476008961$
13. $10510000004524618081 = 11 \cdot 19 \cdot 29 \cdot 31 \cdot 43 \cdot 73 \cdot 97 \cdot 2017 \cdot 91081$

14. $10510000087397028751 = 13 \cdot 31 \cdot 61 \cdot 139 \cdot 1093 \cdot 2269 \cdot 1240219$
15. $10530000017578826101 = 936749701 \cdot 11240996401$
16. $10530000143485615621 = 2294558797 \cdot 4589117593$
17. $10560000152053505917 = 1624807693 \cdot 6499230769$

An integer N is a Carmichael number if and only if N is odd, has at least three prime factors, all factors are distinct, and $(f_j - 1)$ divides $(N - 1)$ for each factor f_j. Psp 1 in the above list is a Carmichael number, the only one of the 17.

Psp 3, 8, 9, 13 are almost Carmichael, since only one factor does not satisfy the above rule.

Psp 6, 10, 15, 16, 17 all have exactly two prime factors, and $(f_1 - 1)$ divides $(f_2 - 1)$. If $r = (f_2 - 1)/(f_1 - 1)$, then N can be written $N = (m + 1)(rm + 1)$.

As explained in [Ribenboim 04, p. 94], in general the number of bases less than N for which N is a pseudoprime is given by

$$B_{\text{psp}}(N) = \prod_{f|N} \gcd(N - 1, f - 1),$$

where f runs over the prime factors of N. Thus, in our special case in which $N = (m + 1)(rm + 1)$, the number of bases for which N is a pseudoprime is m^2, which when divided by N is very close to $1/r$.

For psp 10 and 16, $r = 2$, so it appears to be a good guess that they each would have a 50% chance of being psp(3) (which of course they are).

It seems that a closer examination of such pseudoprimes might lead to a better understanding of them and assist in predicting pseudoprime probabilities for larger sizes.

REFERENCES

[Granville and Pomerance 01] A. Granville and C. Pomerance. "Two Contradictory Conjectures Concerning Carmichael Numbers." *Math. Comp.* 71 (2001), 883–908.

[Pinch 00] R. G. E. Pinch. "The Pseudoprimes up to 10^{13}." In *Proc. Fourth Int. Symp. on Algorithmic Number Theory*, edited by W. Bosma, pp. 459–474, Lecture Notes in Computer Science 1838. New York: Springer-Verlag, 2000.

[Pomerance 81] Carl Pomerance. "On the Distribution of Pseudoprimes." *Math. Comp.* 37 (1981), 587–593.

[Pomerance et al. 80] C. Pomerance, J. L. Selfridge, and S. S. Wagstaff, Jr. "The Pseudoprimes to $25 \cdot 10^9$." *Math. Comp.* 29 (1980), 1003–1026.

[Ribenboim 04] P. Ribenboim. *The Little Book of Bigger Primes*, second edition. New York: Springer-Verlag, 2004.

Jens Kruse Andersen, Kirkeskov Alle 50, 1mf, 3050 Humlebaek, Denmark (jens.k.a@get2net.dk)

Harvey Dubner 85 Long Hill Road, Oakland, New Jersey 07436 (harvey@dubner.com)

Received February 22, 2006; accepted in revised form October 31, 2006.