Sneaky Composites

Number Theory

Matthew Junge

March 12th, 2010

Abstract

We define n to be a pseuodoprime base a if $a^{n-1} \equiv 1 \mod n$ and n is composite. I was intrigued by a claim I read in the book "Number Theory in Science and Communication" by Manfred Schroeder. The author stated in passing that if $\psi_2(n)$ denotes the number of pseudoprimes less than or equal to n, then it holds that

$$\psi_2(n) \sim k \cdot \pi(\sqrt{n}) \tag{1}$$

with $k = (1 + \sqrt{5})/2$, the golden ratio, and $\pi(n)$ is the prime counting function.[1]

I have a softspot for the golden ratio and I thought a nice paper topic would be to run some Sage computations and show that indeed these two functions are asymptotic.

However, after researching more thoroughly and running some calculations I found that (1) is false! Thus, the underlying motivation for this paper is to give evidence that the statement is untrue, discuss why such a proposition may have been mad. From here I would like to explore more thoroughly what we do know about this function, and, in generality, what we know and would like to know about the pseudoprimes.

The paper will begin by proving there are infinite pseudoprimes to any base. From there, we will go on to describe useful aspects of the pseudoprimes and look into important theorems and conjectures pertaining to these sneaky composites.

Contents

1 Introduction

Fermat's Little Theorem states that for all primes it holds that for any $a \in \mathbb{Z}$

$$a^p \equiv a \mod p$$

Since modular exponentiation can be computed very quickly on a computer, Fermat's theorem provides an effective probabilistic algorithm for computing primality. Of course knowing with what degree of certainty that a conjectured prime actually is prime is critical to the security of many cryptosystems that protect our important information.

A special case converse of Fermat's Little Theorem was conjectured over twenty five centuries ago by the Chinese. Namely that,

Theorem 1.1. *If n divides $2^{n-1} - 1$ then n is prime.*

It was not until 1819 that the counterexample of 341 was offered to this theorem.[2] Subsequently, it was discovered that there are many composite numbers, n, with the property that for some fixed a it holds that

$$a^{n-1} \equiv 1 \mod n$$

For example, we can factor $341 = 11 \cdot 33$ and $91 = 7 \cdot 13$, yet it holds that $2^{340} \equiv 1 \mod 341$ and $3^{91} \equiv 3 \mod 91$. I call these composites sneaky because they have the potential to give a false positive for a hastily applied primality test with only a few different bases considered. To give more clarity to the idea of a sneaky composite number let me give a precise definition:

Definition 1.1. *An integer n is a pseudoprime base a if (1) n is composite and (2) $a^{n-1} \equiv 1 \mod n$*

The presence of such pseudoprimes and their potential to influence false positives in primality tests motivates further inquiry into the density of such numbers. Since this paper will pay the most attention to pseudoprimes base 2, the reader should assume that unless explicitly stated a pseudoprime refers to a pseudoprime base 2. Some questions that could come to the mind of an earnest prime-smith may be:

1. Given an integer, a, how many pseudoprimes to the base a are there?
2. Are pseudoprimes useful?
3. What can be said about the density of pseudoprimes?
4. Is there an analogue of the prime counting function, $\pi(n)$, for pseudoprimes base a?

This paper will seek to answer these questions by pulling from several articles and books on the matter as well as through direct computation in Sage.

2 Number of Pseudoprimes

2.1 Proving There are Infinite Pseudoprimes to any Base

Since life with prime numbers is rarely easy, one might correctly guess that there are infinitely many pseudoprimes to any base. We begin our discussion by proving that there are infinite pseudoprimes base a for all $a \in \mathbb{Z}$.

Theorem 2.1. *For each integer $a \geq 2$ there are infinitely many pseudoprimes base a.*[3]

Proof. We shall show that if p is any odd prime not dividing $a^2 - 1$, then $n = \frac{a^{2p}-1}{a^2-1}$ is a pseudoprime base a.

A quick aside, this statement is equivalent to the theorem since we know that given any a there are infinitely many primes larger than $a^2 - 1$ and hence infinitely many n such that

$$n = \frac{a^{2p} - 1}{a^2 - 1} = \frac{a^p - 1}{a - 1} \cdot \frac{a^p + 1}{a + 1}$$

For example if $a = 2$ and $p = 5$, we know that $5 \nmid 2^2 - 1$ thus our formula gives

$$n = \frac{2^{2\cdot 5} - 1}{2^2 - 1} = \frac{2009}{3} = 341$$

Back to the proof of our theorem we see by squaring both sides in Fermat's Little Theorem we obtain

$$a^{2p} \equiv a^2 \mod p$$

So, p divides $a^{2p} - a^2$. Since p does not divide $a^2 - 1$ by hypothesis, and since

$$n - 1 = \frac{a^{2p} - a^2}{a^2 - 1}$$

we conclude that p divides $n - 1$. We now note that the identity

$$\sum_{i=1}^{n-1} a^{2(p-i)} = a^{2(p-1)} + a^{2(p-2)} + \ldots + a^2 \equiv n - 1 \mod p$$

implies that $n-1$ is the sum of an even number of terms of the same parity, so $n-1$ must be even. This means that both 2 and p divide $n - 1$, hence by an elementary theorem we have $2p \mid n-1$. Putting this together gives $a^{2p} - 1 \mid a^{n-1} - 1$. However, $a^{2p} - 1 = n \cdot (a^2 - 1)$, and thus, is by construction a multiple of n. This implies that

$$a^{2p} - 1 \equiv 0 \mod n$$

which gives the statement

$$a^{n-1} \equiv 1 \mod n$$

This means that the set of pseudoprimes base a is injective into the set of primes that do not divide $a^2 - 1$. And since for all a there are infinite primes greater than $a^2 - 1$, there must be infinite pseudoprimes to any base.

□

2.2 Computing Pseudoprimes in Sage

We have now have an explicit formula to generate pseudoprimes. We can define some functions in Sage and take a look at how quickly these numbers grow. To do this I defined the following function:

sage: def q(a,p):
sage: return (aˆ(2*p)-1)/(aˆ2-1)

Listed below are the values of $q(2, p)$ and $q(3, p)$ for all $p < 50$:

Table 1: A Pseudoprime Generating Function Defined on the Primes

p	**q(2,p)**	**q(3,p)**
2	5	10
3	21	91
5	341	7381
7	5461	597871
11	1398101	3922632451
13	22369621	317733228541
17	5726623061	2084647712458321
19	91625968981	168856464709124011
23	23456248059221	1107867264956562636991
29	96076792050570581	588766087155780604365200461
31	1537228672809129301	476900530596182228953581237351
37	6296488643826193618261	25344449488056571213320166359119221
41	1611901092819505566274901	166284933091139163730593611482181209801
43	257904174851120890603984 21	1346907958038227226217808253005667799 3891
47	660234687618869479946199 5861	883706311268880883121503994797018643179 19671

It is evident that this pseudoprime function grows rather quickly and for very large primes becomes difficult to compute. Though there is little optimization in my function, Sage will only compute up to $p = \text{next\_prime}(10^8)$.

2.3 Enumerating The Pseudoprimes Base 2

Now that we have proven there are infinite pseudoprimes by means of a function defined on the primes, one might wonder if we have found a way to describe all the pseudoprimes of a given base. Unfortunately, the function $q : \{a\} \times P \to P_a$ with $P = \{p : p \text{ is prime, } p \nmid a^2 - 1\}$ and $P_a = \{\text{pseudoprimes base a}\}$ is not a surjection. This is quickly seen by the fact that $2^{560} \equiv 1 \mod 561$ with $561 = 3 \cdot 11 \cdot 17$, but $561 \notin P$. So, this means that there is more to be said about the density of pseudoprimes.

3 Primality Testing and Pseudoprimes

We will begin to investigate the density of the pseudoprimes, but first we should motivate why pseudoprimes deserve this attention.

3.1 A Very Fast Primality Test...

Large primes are the name of the game in much of cryptography and cryptographers certainly want to reduce the risk as much as possible of letting a sneaky composite pass as a prime in a cryptosystem. For this reason alone it would be of great interest to know which composites share common structure with primes. However, the application goes deeper.

Let us imagine that through painstaking computation we have computed a database of all pseudoprimes base 2 up to 2^M for some integer M. With our list in hand, we then encounter a large integer, n, for which we would like to know whether or not n is prime. Let us say that n can be encoded with no more than $M - 1$ bits. If this is the situation we offer the following primality test.

Step 1: Does n appear in our list of pseudoprimes base 2. If yes, output "False" and terminate. If no, go on to Step 2.

Step 2: Compute $2^{n-1} \mod n$. If this is equal to 1, output "True" and terminate. If this is not equal to 1 output "False" and terminate.

Since our list of pseudoprimes base 2 is a list of all composites, m, such that $2^{m-1} \equiv 1 \mod m$, it follows immediately that any numbers not in our list with this property must be prime. Moreover, because modular exponentiation base two is very fast, we have devised an extremely efficient primality testing algorithm.

3.2 How difficult is it to enumerate the pseudoprimes?

The above section demonstrates that if we had a comprehensive list of the pseudoprimes we could devise a fast primality test, but how realistic is it to obtain a database of the pseudoprimes?

3.2.1 A Pseudoprime Counting Function in Sage

I attempted to answer this question by doing my own computations in Sage. I wished to construct a pseudoprime counting function $\psi_2(n)$. To do this I defined the function:

```
sage: def psi(n):
sage:     return sum(2.powermod(i-1,i)==1 for i in range(3,n+1,2)) - prime_pi(n)+1
```

Basically, ψ_2 cycles through all of the integers $i \leq n$ and sums each result when $2^{i-1} \equiv 1 \mod i$. Then it subtracts off the number of primes less than n. This function is able to count the pseudoprimes up to $n = 10^7$ fairly quickly. But after that it is no longer reasonable to count using this method. For instance, it takes 77 seconds in sage to compute $\psi_2(10^7)$. Below is a table of the function values:

Table 2: A Sage Pseudoprime Counting Function

n	10^1	10^2	10^3	10^4	10^5	10^6	10^7
$\psi_\mathbf{2}(\mathbf{n})$	0	0	3	22	78	245	750

I will return to a discussion of the pseudoprime counting function later in the paper. Similar to primes, it is more desirable to have an explicit list of the pseudoprimes. We now look at some current research into enumerating all of the pseudoprimes base 2.

3.2.2 Current Research in Psuedoprimes

Early number theorists recognized the importance of the pseudoprimes and by 1926 Poulet had enumerated all of the pseudoprimes base 2 up to 50 million ($\approx 2^{26}$) [4]. Eighty years later[5], Feitsma has produced a complete list up to 10^{17} ($\approx 2^{56}$).

This means that we can very quickly test the the primality of a 56 bit (24 digit) number. Moreover, Feitsma is close to computing up to 2^{64}. Though she is yet to publish the algorithm she uses for pseudoprimality discovery, she has posted a table of her results. Moreover, her website contains an explicit database of all pseudoprimes less than 10^{17}. If we let $\psi_a(n)$ denote the number of pseudoprimes base a less than n then the counts as of September 2009 are given in the table below:

Table 3: Number of Pseudoprimes Less Than or Equal to n

$\mathbf{n}$	$\psi_{\mathbf{2}}(\mathbf{n})$
10^3	3
10^4	22
10^5	78
10^6	245
10^7	750
10^8	2057
10^9	5597
10^{10}	14884
10^{11}	38975
10^{12}	101629
10^{13}	264239
10^{14}	687007
10^{15}	1801533
10^{16}	4744920
10^{17}	12604009
10^{18}	33763684*
10^{19}	91210364*
2^{64}	118968378*

*These figures are not yet confirmed and are still being computed.

The amount of time put in by Feitsma indicates that determining pseudoprimes becomes difficult for large n, however soon we will be able to quickly determine the primality of any 64 bit number. So, this means that with our current techniques, the outlook for the primality testing algorithm described in 3.2.1 to test say 128 bit integers or 256 bit integers is a ways off. However, a more realistic solution is to turn our attention back to ψ_2, the pseudoprime counting function. If we can compute this accurately, then at least we will be able to make probabilistic arguments concerning the likeliness a large integer is prime.

4 The Density of the Pseudoprimes

The main interest of this paper is the pseudoprime counting function:

$$\psi_2(n) = \#\{i \leq n : 2^{i-1} \equiv 1 \mod i,\ i \text{ is not prime}\}$$

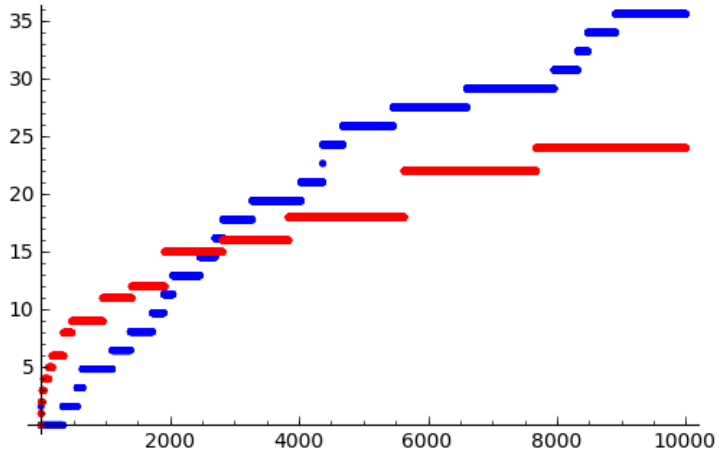
Namely, we will first give strong computational evidence concerning the falsity of the claim:

$$\psi_2(n) \sim \frac{1+\sqrt{5}}{2} \cdot \pi(\sqrt{n}) \tag{2}$$

4.1 Computational Evidence Against Schroeder's Claim

For simplicity, we denote $\aleph(n) = k \cdot \pi(\sqrt{n})$, with k the golden ratio and $\pi(n)$ the prime counting function. We begin our analysis by constructing some charts.

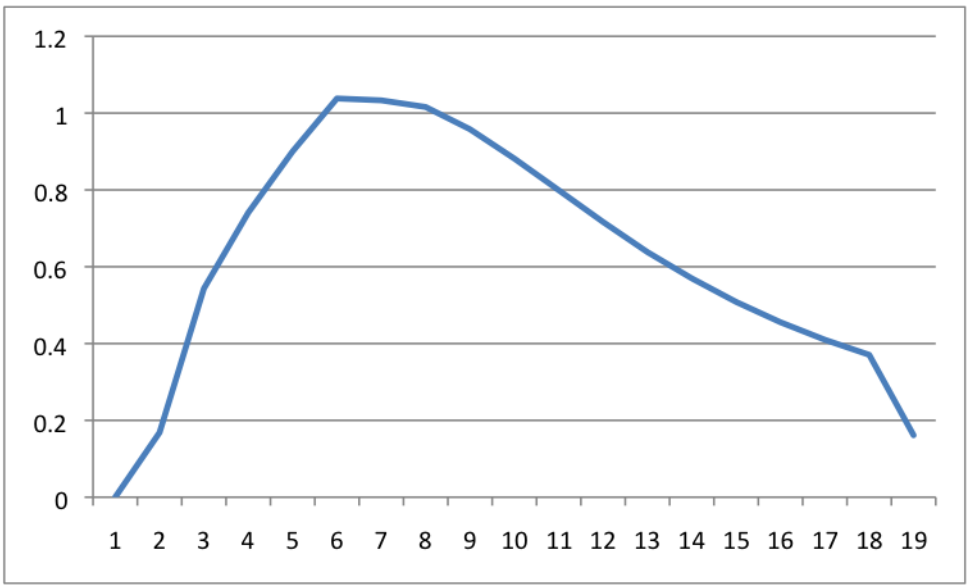
Figure 1: A plot of $\psi_2(n)$ (red) and $\aleph(n)$ (blue)



Pictured above is a listplot made in sage from $[0, 10^5]$. In this domain it seems plausible that the two functions are asymptotic. And, at the time the claim was published (1984), the number of pseudoprimes base 2 was known only up to 10^9.

However, if we go out further we see convincing evidence that the two functions are not asymptotic. Fortunately, the values of $\pi(n)$ are well known up to $n = 10^{10}$ and if we also utilize the table given by Feitsma we can make a simple plot that indicates whether the functions are asymptotic. This is best seen by dividing $\psi_2(n)$ by $\aleph(n)$:

Figure 2: A plot of $\frac{\psi_2(n)}{\aleph(n)}$ with domain $[1, 10^{19}]$



The plot above indicates that the functions appear plausibly asymptotic up to $n = 10^8$. But, for larger values it becomes clear that $\aleph$ is growing much more quickly. Since this is literally the cutting edge of the data we have on pseudoprimes, this indicates that the conjecture is most likely not true.

4.2 A Heuristic for the Falsity of the Claim

The evidence given above raises the question, "Why make such a claim?" A likely explanation is that at the time the data suggested the truth of the claim. And it is important to note that the claim is made in passing and no further justification is given. So, Schroeder has very little stake in its correctness and it is most likely an unchecked error in the text.

Another bit of evidence against the claim is that after researching this topic extensively I could find no other reference to such a result. It seems likely that if such a nice statement were true, there would be mention of it somewhere else.

However, my research did yield some results concerning $\psi_2(n)$. Namely, an upper and lower bound for the function.

4.3 Bounding $\psi_2(n)$

In Paul Erdös's 1950 article, "On Almost Primes", he proves the existence of an upper bound for $\psi_2(n)$.

Theorem 4.1. *(Erdös)*[6] *If $\psi_2(x)$ denotes the pseudoprime counting function then for sufficiently large x it holds that:*

$$\psi_2(x) < xe^{\left(-\frac{1}{3}(\log x)^{\frac{1}{4}}\right)}$$

A very bare bones sketch of his four page proof goes as follows. Let P denote the set of all pseudoprimes less than x. We partition P by considering the function $g : \{1, 2, \ldots, x\} \to \mathbb{N}$ defined as

$$g(n) = \min\{r : 2^r \equiv 1 \mod n\}$$

We then form the partition

$$C_1 = \{n \in P : g(n) \leq \exp((\log x)^{\frac{1}{2}})\}$$

$$C_2 = P - C_1$$

Through partitioning both C_1 and C_2 into smaller subclasses, Erdös finds bounds for both partitions, which he sums to arrive at his upper bound. However, if we call the upper bound $M(n)$, then the table below reveals that the bound proven by Erdös is very generous.

Table 4: An Upper Bound on $\psi_2(n)$

n	$\psi_2(\mathbf{n})$	**M(n)**
10^3	3	583
10^4	22	5596
10^5	78	54118
10^6	245	525901
10^7	750	5127877
10^8	2057	50129228
10^9	5597	491053335
10^{10}	14884	4818200394

Nonetheless, this result is important since it establishes that there is a definite ceiling on the number of pseudoprimes base 2. In another article by Erdös, "On the Converse of Fermat's Theorem", he claims that he can prove the following inequality:

Conjecture 4.1. *(Erdös)*[7] *For fixed c_1, c_2 and for every $k \geq 1$ there exists x sufficiently large such that*

$$c_1 \cdot \log x < \psi_2(x) < c_2 \cdot \frac{x}{(\log x)^k}$$

I was unable to locate Erdös's proof of this claim, however if his conjecture holds this immediately confirms the empirical data which suggests that $\psi(n) < \pi(n)$. This follows since the lower bound for $\pi(n)$ is $\frac{x}{\log x}$. Since Erdös's work in the mid 19th century, little has been discovered to directly bound ψ_2. However, there are alternate ways to approach this problem.

5 Pseudoprimes of k Factors

Since directly attacking the problem of the density of pseudoprimes seems a difficult route, one might start considering ways to reduce the problem. Unlike the primes, pseudoprimes have the additional structure of being composite. Thus, we can ask questions regarding the amount of pseudoprimes with k factors. A priori it is possible that given an integer, k, there are none or perhaps only finitely many pseudoprimes with k factors. Or, one might start studying the density of pseudoprimes with k factors. We will see below that some headway has been made in both of these pursuits.

5.1 There Exist Infinite Pseudoprimes with k Factors

In 1949 Erdös proved the following theorem.

Theorem 5.1. *(Erdös)*[8] *For every k there exist infinitely many squarefree pseudoprimes, n, with $v(n) = k$.*

This result tells us the somewhat discouraging result that even for k equal to some astronomically large integer, there are infinite pseudoprimes with $v(n) = k$. However, the result does not suggest that the density of such factors is uniform. In fact, it seems reasonable to conjecture that the a large amount of pseudoprimes up to some bound, T, will have only two factors. William Galway has done some research into this matter.

5.2 The Density of Pseudoprimes with 2 Factors

In a 2001 lecture entitled "The Density of Pseudoprimes with Two Prime Factors" William Galway studies the density of pseudoprimes with two factors. For p, q primes Galway defines the function

$$P_2(x) = \#\{n \leq x : n = pq,\ p < q,\ 2^{n-1} \equiv 1 \mod 2\}$$

P_2 is a counting function for pseudoprimes with two factors. He then makes the following conjecture:

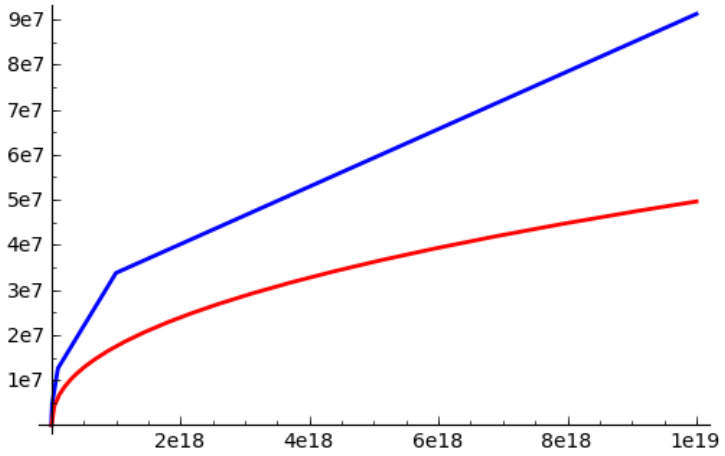
Conjecture 5.1. *(Galway)*[9]

$$P_2(x) \sim C\sqrt{x}/\log^2(x)$$

with $C \approx 30.03$.

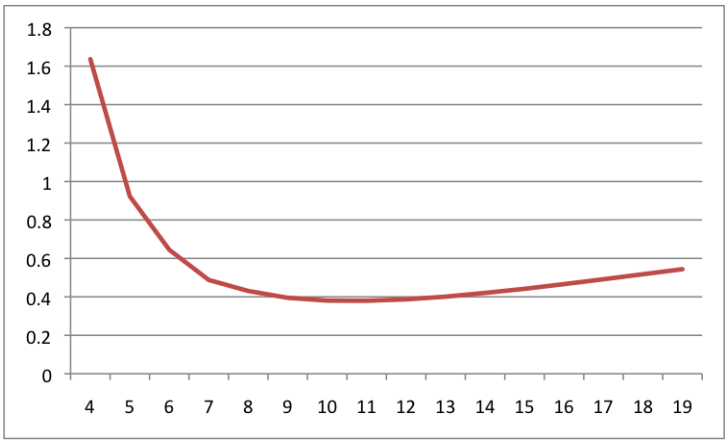
Galway makes this conjecture by looking at pairs of primes that fall on lines determined by prime factors of pseudoprimes. If we look at his conjecture it reveals information regarding $\psi_2(n)$. Let $C(n)$ denote Galway's conjectured function that asymptotic to $P_2(n)$. Plotted below is $\psi_2(n)$ and $C(n)$:

Figure 3: $\psi_2(n)$ (blue) and $C(n)$ (red) on the domain $[10^3, 10^{19}]$



We see that this function seems to be a much better estimate than Erdös's upper bound and, at the very least, it confirms the fact that $P_2 < \psi_2$ for all sufficiently large n. We also can look at $\frac{\psi_2}{C}$:

Figure 4: $\frac{\psi_2(n)}{C(n)}$on the domain $[10^4, 10^{19}]$



Although Galway's conjecture has not been proven, it gives a nice lower bound on ψ_2. Moreover, if we let $Q(n) = \frac{\psi_2}{C}$ on the interval $[10^{11}, 10^{19}]$ it appears that $Q(n)$ is increasing. It seems reasonable to conjecture that $Q(n) < 1$ for sufficiently large n. Thus, $C(n)$ may come close to approximating $\psi_2(n)$. Still, we need much more data to even begin to conjecture about the relation between C and ψ_2. Nonetheless, by providing a lower bound, $C(n)$ is helpful in determining the likelihood that an integer that passes a base two pseudoprime test is indeed a prime.

6 Conclusion and Suggested Topics for Further Inquiry

This paper was motivated by a beautiful claim that turned out to be false. However, the claim inspired research into an overlooked subject that is both deep and potentially useful. Though a comprehensive list of pseudoprimes base 2 up to some large bound is far in the future, I hope that I have shown that sneaky composites are not necessarily the enemy of a prime-smith. In fact, by studying pseudoprimes and their properties we can learn much more about the structure of the primes and integers in general.

If the reader enjoyed this topic I suggest looking into other formulations of pseudoprimes. Particularly fascinating are composites known as Carmichael Numbers. These perhaps are the sneakiest of all composites for they are a pseudoprime in all bases! That is if n is a Carmichael Number then for all a, $a^{n-1} \equiv 1 \mod n$.

In addition to Carmichael Numbers, other types of pseduoprimes can be defined. Some examples are, Euler Pseudoprimes, Absolute Pseudoprimes and Strong Pseudoprimes. A good introduction to these numbers is contained in "Number Theory: An Introduction Via the Distribution of Primes" by Fine and Rosenberger, and also in "Prime Numbers" by Crandall and Pomerance.

References

[1] Crandall, Richard; Pomerance, Carl, *Prime Numbrs.* Springer-Verlag. New York. 2001. [PN]

[2] Erdös, Paul, *On The Converse of Fermat's Last Theorem.* The American Mathematical Monthly, vol 56 , No. 9. Nov., 1949. pp. 623-624[Converse]

[3] Erdös, Paul, *On Almost Primes.* The American Mathematical Monthly, vol 57 , No. 6. July, 1950. pp. 404-407[Almost]

[4] Feitsma, Jan, *Pseudoprimes.* Rijksuniversiteit Groningen. July 21st, 2009. <http://www.janfeitsma.nl/math/psp2/index>. [Pseudoprimes]

[5] Fine, Benjamin; Rosenberger, Gerhard, *Number Theory.* Birkhauser. Boston. 2007. [NT]

[6] Galway, William, *The Density of Pseudoprimes with Two Prime Factors.* University of Illinois. 2001. <http://www.cecm.sfu.ca/ wfgalway/SlidesETC/Draft-DissectedSieve-Slides.pdf/>

[7] Lehmer, Derrick, *On The Converse of Fermat's Last Theorem I.* The American Mathematical Monthly, vol 43 , No. 6. July, 1936. pp. 347-354[Converse I]

[8] Lehmer, Derrick, *On The Converse of Fermat's Last Theorem II.* The American Mathematical Monthly, vol 56 , No. 6. May, 1949. pp. 300-309[Converse II]

[9] Schroeder, Manfred, *Number Theory in Science and Communication.* Springer-Verlag. New York. 1984. [NT in SC]

Notes

[1]Schroeder p 206
[2]Converse II p 347
[3]Primes p 121
[4]Converse II p 301
[5]Pseudoprimes
[6]On Almost Primes
[7]On The Converse of Fermat's Last Theorem 3
[8]On the Converse of Fermat's Last Theorem
[9]Galway