LARGE PRIME NUMBERS

1. Fermat Pseudoprimes

Fermat's Little Theorem states that for any positive integer n,

$$\textit{if } n \textit{ is prime then } b^n \,\%\, n = b \textit{ for } b = 1, \dots, n-1.$$

In the other direction, all we can say is that

$$\textit{if } b^n \,\%\, n = b \textit{ for } b = 1, \dots, n-1 \textit{ then } n \textit{ might be prime.}$$

If $b^n \,\%\, n = b$ where $b \in \{1, \dots, n-1\}$ then n is called a **Fermat pseudoprime base** b.

There are 669 primes under 5000, but only five values of n (561, 1105, 1729, 2465, and 2821) that are Fermat pseudoprimes base b for $b = 2, 3, 5$ without being prime. This is a false positive rate of less than 1%. The false positive rate under 500,000 just for $b = 2, 3$ is 0.118%.

On the other hand, the bad news is that checking more bases b doesn't reduce the false positive rate much further. There are infinitely many **Carmichael numbers**, numbers n that are Fermat pseudoprimes base b for all $b \in \{1, \dots, n-1\}$ but are not prime.

In sum, Fermat pseudoprimes are reasonable candidates to be prime.

2. Strong Pseudoprimes

The **Miller–Rabin test** on a positive integer n and a positive test base b in $\{1, \dots, n-1\}$ proceeds as follows.

- Factor $n-1$ as $2^s m$ where m is odd.
- Replace b by $b^m \,\%\, n$.
- If $b = 1$ then return the result that n could be prime, and terminate.
- Do the following s times: If $b = n-1$ then return the result that n could be prime, and terminate; otherwise replace b by $b^2 \,\%\, n$.
- If the algorithm has not yet terminated then return the result that n is composite, and terminate.

(Slight speedups here: (1) If the same n is to be tested with various bases b then there is no need to factor $n-1 = 2^s m$ each time; (2) there is no need to compute $b^2 \,\%\, n$ on the sth time through the step in the fourth bullet.)

A positive integer n that passes the Miller–Rabin test for some b is a **strong pseudoprime base** b.

For any n, at least $3/4$ of the b-values in $\{1, \dots, n-1\}$ have the property that if n is a strong pseudoprime base b then n is really prime. But according to the theory, up to $1/4$ of the b-values have the property that n could be a strong pseudoprime base b but not be prime. In practice, the percentage of such b's is much lower. For n up to 500,000, if n is a strong pseudoprime base 2 and base 3 then n is prime.

3. Generating Candidate Large Primes

Given n, a simple approach to finding a candidate prime above $2n$ is as follows. Take the first of $N = 2n+1$, $N = 2n+3$, $N = 2n+5$, ... to pass the following test.

(1) Try trial division for a few small primes. If N passes, continue.
(2) Check whether N is a Fermat pseudoprime base 2. If N passes, continue.
(3) Check whether N is a strong pseudoprime base b as b runs through the first 20 primes.

Any N that passes the test is extremely likely to be prime. And such an N should appear quickly. Indeed, using only the first *three* primes in step (3) of the previous test finds the following correct candidate primes:

The first candidate prime after	10^{50}	is	$10^{50} + 151$.
The first candidate prime after	10^{100}	is	$10^{100} + 267$.
The first candidate prime after	10^{200}	is	$10^{200} + 357$.
The first candidate prime after	10^{300}	is	$10^{300} + 331$.
The first candidate prime after	10^{1000}	is	$10^{1000} + 453$.

4. Certifiable Large Primes

The **Lucas–Pocklington–Lehmer Criterion** is as follows. *Suppose that* $N = p \cdot U + 1$ *where* p *is prime and* $p > U$. *Suppose also that there is a base* b *such that* $b^{N-1} \,\%\, N = 1$ *but* $\gcd(b^U - 1, N) = 1$. *Then* N *is prime.*

The proof will be given in the next section. It is just Fermat's Little Theorem and some other basic number theory.

As an example of using the result, start with

$$p = 1000003.$$

This is small enough that its primality is easily verified by trial division. A candidate prime above $1000 \cdot p$ of the form $p \cdot U + 1$ is

$$N = 1032 \cdot p + 1 = 1032003097.$$

And $2^{N-1} \,\%\, N = 1$ and $\gcd(2^{1032} - 1, N) = 1$, so the LPL Criterion is satisfied, and N is prime. Rename it p.

A candidate prime above $10^9 \cdot p$ of the form $p \cdot U + 1$ is

$$N = p \cdot (10^9 + 146) + 1 = 1032003247672452163.$$

Again $b = 2$ works in the LPL Criterion, so N is prime. Again rename it p.

A candidate prime above $10^{17} \cdot p$ of the form $p \cdot U + 1$ is

$$N = p \cdot (10^{17} + 24) + 1 = 1032003247672452410680779441388519 13.$$

Again $b = 2$ works in the LPL Criterion, so N is prime. Again rename it p.

A candidate prime above $10^{34} \cdot p$ of the form $p \cdot U + 1$ is

$$N = p \cdot (10^{34} + 224) + 1 = 103200324767245241068077944138854224687274786293399924945948710282851 3.$$

Again $b = 2$ works in the LPL Criterion, so N is prime. Again rename it p.

A candidate prime above $10^{60} \cdot p$ of the form $p \cdot U + 1$ is

$$\begin{aligned} N = p \cdot (10^{60} + 1362) + 1 = &103200324767245241068077944138854224 \\ &6872747862933999249460892691251842 8 \\ &8018334722159917119454024068258931 61 \\ &06977763821434052434707. \end{aligned}$$

Again $b = 2$ works in the LPL Criterion, so N is prime. Again rename it p.

A candidate prime above $10^{120} \cdot p$ of the form $p \cdot U + 1$ is

$$\begin{aligned} N = p \cdot (10^{120} + 796) + 1 = &1032003247672452410680779441388542 2 \\ &46872747862933999249460892691251842 8 \\ &80183347221599171194540240682589316 1 \\ &06977763822255527019854272118901900 4 \\ &35345279628510707298895463402570870 5 \\ &82236466932625944388392940270854031 5 \\ &83341095621154300001861505738026773. \end{aligned}$$

Again $b = 2$ works in the LPL Criterion, so N is prime.

5. Proof of the Lucas–Pocklington–Lehmer Criterion

Recall the Lucas–Pocklington–Lehmer Criterion: *Suppose that $N = p \cdot U + 1$ where p is prime and $p > U$. Suppose also that there is a base b such that $b^{N-1} \,\%\, N = 1$ but $\gcd(b^U - 1, N) = 1$. Then N is prime.*

The proof begins with an observation that goes back to Fermat and Euler:

Fermat–Euler Criterion. *Let p be prime. Let N be an integer such that*

$$N \,\%\, p = 1.$$

If there is an integer b such that

$$b^{N-1} \,\%\, N = 1 \quad \textit{and} \quad \gcd(b^{(N-1)/p} - 1, N) = 1$$

then

$$q \,\%\, p = 1 \quad \textit{for each prime divisor } q \textit{ of } N.$$

To prove the Fermat–Euler criterion, let q be any prime divisor of N. Since $b^{N-1} \,\%\, N = 1$, it follows that

$$b^{N-1} \,\%\, q = 1.$$

Let t be the smallest positive integer such that $b^t \,\%\, q = 1$. Thus $t \mid N - 1$, and also $t \mid q - 1$ by Fermat's Little Theorem. On the other hand, we claim that

$$b^{(N-1)/p} \,\%\, q \neq 1,$$

so that $t \nmid (N - 1)/p$. Indeed, if equality were to hold in the previous display, then we would have $b^{(N-1)/p} - 1 = kq$, violating the condition $\gcd(b^{(N-1)/p} - 1, N) = 1$. Now we have,

$$t \mid N - 1, \quad t \nmid (N - 1)/p$$

so that $p \mid t$, and in fact

$$p \mid t, \quad t \mid q - 1.$$

It follows that $p \mid q - 1$, i.e., $q \,\%\, p = 1$ as desired.

Returning to the Lucas–Pocklington–Lehmer Criterion, recall that we have $N = p \cdot U + 1$ where $p > U$. The properties of the base b show that all prime divisors q of N satisfy $q \,\%\, p = 1$. If N were to be composite then it would have a prime divisor $q \leq \sqrt{N}$. But this forces $q < p$, and hence $q \,\%\, p \neq 1$, contradiction. Therefore N is prime.

6. Discussion of the Miller–Rabin Test

Given a positive integer n and a base b, reason as follows.

- Factor $n - 1 = 2^s \cdot m$ where m is odd.
- If n is prime then $b^{n-1} = 1$ (here and throughout this discussion, all arithmetic is being carried out modulo n). So by contraposition, if $b^{n-1} \neq 1$ then n is composite.
- Hence we continue reasoning only if $b^{n-1} = 1$. In this case we know a square root of 1: it is $b^{(n-1)/2}$.
- If $b^{(n-1)/2} \neq \pm 1$ then too many square roots of 1 exist mod n for n to be prime, and so n is composite.
- If $b^{(n-1)/2} = -1$ then we have no evidence that n is composite, nor can we proceed, since we have no new square roots of 1 to study. The algorithm terminates, reporting that n could be prime.
- But if $b^{(n-1)/2} = 1$ then we do have a new square root of 1 at hand: it is $b^{(n-1)/4}$.
- This process can continue until $b^{2m} = 1$, so that b^m is a square root of 1. If $b^m \neq \pm 1$ then n is composite. Otherwise, n could be prime.

To encode the algorithm efficiently, the only wrinkle is to compute the powers of b from low to high, even though the analysis here considered them from high to low. Inspecting the highest power b^{n-1} turns out to be redundant.

Another way to think about the Miller–Rabin test is as follows. Again let $n-1 = 2^s \cdot m$. Then

$$\begin{aligned}
X^{2^s m - 1} - 1 &= (X^{2^{s-1}m} + 1)(X^{2^{s-1}m} - 1) \\
&= (X^{2^{s-1}m} + 1)(X^{2^{s-2}m} + 1)(X^{2^{s-2}m} - 1) \\
&= (X^{2^{s-1}m} + 1)(X^{2^{s-2}m} + 1)(X^{2^{s-3}m} + 1)(X^{2^{s-3}m} - 1) \\
&\;\;\vdots \\
&= (X^{2^{s-1}m} + 1)(X^{2^{s-2}m} + 1)(X^{2^{s-3}m} + 1) \cdots (X^m + 1)(X^m - 1).
\end{aligned}$$

That is, rewriting the left side and reversing the order of the factors of the right side,

$$X^{n-1} - 1 = (X^m - 1) \cdot \prod_{r=0}^{s-1} (X^{2^r m} + 1).$$

It follows that

$$b^{n-1} - 1 = (b^m - 1) \cdot \prod_{r=0}^{s-1} (X^{2^r m} + 1) \bmod n, \quad \text{for } b = 1, \ldots, n-1.$$

If n is prime then $b^{n-1} - 1 = 0 \bmod n$ for $b = 1, \ldots, n$, and also $\mathbf{Z}/n\mathbf{Z}$ is a field, so that necessarily one of the factors on the right side vanishes modulo n as well.

That is, given any base $b \in \{1, \dots, n-1\}$, if n is prime then at least one of the factors

$$b^m - 1, \quad \{b^{2^r m} + 1 : 0 \le r \le s-1\}$$

vanishes modulo n. So conversely, given any base $b \in \{1, \dots, n-1\}$, if none of the factors vanishes modulo n then n is composite. This analysis shows that the Miller–Rabin test can be phrased as earlier in this writeup.

(Beginning of analysis of false positives.)

Lemma. *Let p be an odd prime. Let n be a positive integer divisible by p^2. Let x, y be integers such that $x = y \bmod p$ and $x^{n-1} = y^{n-1} = 1 \bmod n$. Then $x = y \bmod p^2$.*

First we note that $x^p = y^p \bmod p^2$. This follows quickly from the relation

$$x^p - y^p = (x-y)(x^{p-1} + x^{p-2}y + \cdots + xy^{p-2} + y^{p-1}),$$

because the condition $x = y \bmod p$ makes each of the multiplicands on the right side a multiple of p. Second, raise both sides of the relation $x^p = y^p \bmod p^2$ to the power n/p to get $x^n = y^n \bmod p^2$. But since $x^n = x \bmod n$, certainly $x^n = x \bmod p^2$, and similarly for y. The result follows.

Proposition. *Let p be an odd prime. Let n be a positive integer divisible by p^2. Let B denote the set of bases b between 1 and $n-1$ such that n is a Fermat pseudoprime base b, i.e.,*

$$B = \{b : 1 \le b \le n-1 \text{ and } b^{n-1} \,\%\, n = 1\}.$$

Then

$$|B| \le \frac{p-1}{p^2} n \le \frac{1}{4}(n-1).$$

To see this, decompose B according to the values of its elements modulo p,

$$B = \bigcup_{d=1}^{p-1} B_d$$

where

$$B_d = \{b \in B : b \,\%\, p = d\}, \quad 1 \le d \le p-1.$$

For any d such that $1 \le d \le p-1$, if $b_1, b_2 \in S_d$ then we know that $b_1 = b_2 \bmod p^2$. It follows that $|S_d| \le n/p^2$, and the result follows.