A 60,000 DIGIT PRIME NUMBER OF THE FORM $x^2 + x + 41$

1. Introduction

1.1. **Euler's polynomial.** Euler observed that $f(x) = x^2 + x + 41$ takes on prime values for $0 \leq x \leq 39$.

Even after this point $f(x)$ takes on a high frequency of primes.

For instance among $f(1), f(2), \ldots f(10^6)$, 261,080 are prime compared to the sequence $1, 2, \ldots 10^6$ where there are 78,498 primes.

1.1.1. ***Stark-Heegner Theorem.*** Let $d > 0$ be a square-free integer then $\mathbb{Q}(\sqrt{-d})$ has class number $1 \iff d \in \{1, 2, 3, 7, 11, 19, 43, 67, 163\}$.

Originally conjectued by Gauss, this was essentially proved by Karl Heegner albeit with gaps in 1952. It was completely proved by Harold Stark in 1967. The numbers are known as Heegner numbers.

In 1913 Rabinowitz proved that $n^2 + n + A$ gives primes for $0 \leq n \leq A - 2$ if and only if its discriminant $1 - 4A$ is minus a Heegner number.

$A - 1$ yields A^2 so $A - 2$ is maximal.

$1, 2, 3$ are not of the required form since $4A - 1 = h \Rightarrow A = \dfrac{h+1}{4}$. So the Heegner numbers that work 7, 11, 19, 43, 67, 163, yield prime generating functions corresponding to $A = 2, 3, 5, 11, 17, 41$.

These are called the Lucky Numbers of Euler.

1.2. **Hardy and Littlewood Conjecture 1.** *Hardy and Littlewood had a number of precise conjectures about prime distributions.*

The first conjecture: the prime k-tuples conjecture (prime k-tuple: $(p, p + a_1, \ldots p + a_k)$ is a sequence of primes such that a_k is as "least" as possible.) implies that for any positive integer m, $\exists A$ such that $x^2 + x + A$ is prime for $0 \leq x \leq m$.

Note that we aren't saying that $m = A - 2$ so this does not contradict Rabinowitz. So with a large enough choice of A, Euler's polynomial can be beaten.

1.3. **Hardy and Littlewood's Conjecture F.** Consider $f(x) = ax^2 + bx + c$. If $p \mid f(x)$ for some $x \in \mathbb{Z}$, then $\Delta = b^2 - 4ac$, the discriminant of $f(x)$ must be a square modulo p.
Thus if Δ is not a square modulo p for many primes p, we expect $f(x)$ to take on many prime values asymptotically.
Hardy & Littlewood formalized this in conjecture F.

1.3.1. ***Conjecture (F).*** Let $a > 0$, b, c be integers such that $\gcd(a, b, c) = 1$, $\Delta = b^2 - 4ac$ is not a square and $a + b$, c are not both even. Then there are infinitely many primes of the form $f(x)$, and

$$\pi_f(n) \sim \epsilon C_f Li(n),$$

where

$$Li(n) = \int_2^n \frac{dx}{\log x}$$

,

$$\epsilon = \begin{cases} \frac{1}{2} & \text{when} 2 \nmid a + b, \\ 1 & \text{otherwise,} \end{cases}$$

and

$$C_f = \prod_{\substack{p>2 \\ p|(a,b)}} \frac{p}{p-1} \prod_{\substack{p>2 \\ p \nmid a}} \left(1 - \frac{\left(\frac{\Delta}{p}\right)}{p-1} \right)$$

The products in the expression for C_f are taken over the primes only, and $\left(\frac{\Delta}{p}\right)$ denotes the Legendre symbol. Note here that ϵC_f is what really determines the density of prime values assumed by f, since $Li(n)$ is a function of n only. The larger ϵC_f is, the higher the asymptotic density of prime values for any quadratic polynomial of discriminant Δ.

For polynomials of the form $f_A(x) = x^2 + x + A$, $\pi_{f_A}(n) \sim C(\Delta)L_A(n)$, where

$$C(\Delta) = \prod_{p \geq 3} \left(1 - \frac{\left(\frac{\Delta}{p}\right)}{p-1} \right)$$

Here $\Delta = 1 - 4A$.

No one has found a polynomial of the form $f_A(x)$ that represents distinct primes for more than the first 40 values of x.

So Euler's polynomial $f_{41}(x) = x^2 + x + 41$ holds the record. For Euler's polynomial $C(-163) = 3.3197732$.

H.C. Williams, M. Jacobson, G.W. Fung *have looked at finding quadratic polynomials which have a high density of prime values.*

For instance $x^2 + x + 3399714628553118047$ has $C(\Delta) = 5.3670819$.

So we expect this polynomial to assume more primes asymptotically than f_{41}.

It starts off slowly, only 24 primes for $x \leq 100$ compared to 87, but for $x \leq 10^7$ it assumes 2517022 prime values compared to only 2208197 by Euler's polynomial.

1.4. **Primality Testing.**

- *The best general primality proving method not based on factorizations is the* Elliptic Curve Primality Proving method (ECPP). The largest prime proved by ECPP has 26,642 digits by Francois Moran in 2011.
- The Lucas-Lehmer-Riesel test is a primality test for numbers of the form $N = k2^n - 1$, with $2^n > k$. It is the fastest deterministic algorithm known for numbers of that form.
- The Brillhart-Lehmer-Selfridge test is the fastest deterministic algorithm for numbers of the form $N = k2^n + 1$. The largest known prime of the form $x^2 + 1$ is $75898^{524288} + 1$, with 2,558,647 digits.

The latter two methods require knowledge of partial prime factorizations of $N + 1$ and $N - 1$ respectively

1.5. **Brillhart-Lehmer-Selfridge Theorem.** Suppose that $N > 1$ is odd and write $N - 1 = FR$ where F is even and the prime factorization of F is known. Suppose also that

(1) $F > \left(\frac{N}{2}\right)^{1/3}$

(2) For each prime p, dividing F, there is an integer a_i, so that $a_i^{N-1} \equiv 1 \pmod{N}$ and $\gcd(a_i^{\frac{N-1}{p_i}} - 1, N) = 1$,

(3) If we write $R = 2Fq + r$, where $1 \leq r < 2F$, then either $q = 0$ or $r^2 - 8q$ is not a perfect square, then N is prime.

1.6. **Approach.** To find large primes of the form $f(x) = x^2 + x + 41$, find polynomials $g(x)$ so that $f(g(x)) - 1$ is reducible.

Doing a computer search reveals the choice

$g(x) = 40x^3 + 41x^2 + 42x + 1$, for which

$$f(g(x)) - 1 = (40x^2 + x + 1)(40x^4 + 81x^3 + 123x^2 + 84x + 42)$$

Let $h(x) = 40x^2 + x + 1$ and $i(x) = 40x^4 + 81x^3 + 123x^2 + 84x + 42$. Thus for any given choice of x, we have $N = f(g(x))$ with $2h(x) = F$ and $\frac{i(x)}{2} = R$.

The goal: Find a choice of x (with known prime factorization), for which $40x^2 + x + 1$ is prime (use B-L-S theorem) and for which $f(g(x))$ is prime (again use B-L-S theorem). *Thus there is a simultaneous primality requirement which increases the number of candidate values of x we must search.*

2. Strategy

2.1. **How Many Numbers Do We Need to Check?** We are looking for a 20,000 digit prime (corresponding to $h(x)$) and a 60,000 digit prime corresponding to $f(g(x))$. By The Prime Number Theorem, the density of primes close to an integer N is approximately equal to $\frac{1}{\ln(N)}$.

If we assume the same density of primes within the values of Euler's Polynomial as within the set of all integers, we multiply the probabilities of finding a 20,000 digit prime and a

60,000 digit prime.

$$\left(\frac{1}{\ln(10^{20000})}\right)\left(\frac{1}{\ln(10^{60000})}\right) \approx \frac{1}{6,342,314,060}$$

If we were to test M numbers, the chance of finding at least one prime pair would be $1-((N-1)/N)^M$ where $N = 6,342,314,060$. If we set $M = 3N$, then

$$1-((N-1)/N)^{3N} \approx 1-\left(\frac{1}{e}\right)^3 \approx 95.0\%$$

2.2. **Pre-Sieving with Primorials.** The primorial $n\#$ is the product of primes less than or equal to n.

$$f(g(x)) = 1600x^6 + 3280x^5 + 5041x^4 + 3564x^3 + 1887x^2 + 126x + 43$$

. By choosing $x = k.\frac{n\#}{43}$, $h(x)$ and $f(g(x))$ will not be divisible by any prime less than n. The density of numbers n divisible by p is $\frac{1}{p}$.
thus for each potential prime divisor eliminated, the number of potential primes decreases by $\frac{1}{p}$ *where* p *is the divisor eliminated.*
Choose $\frac{23,143\#}{43}$ as a factor of x. Thus the number of numbers to check should be

$$\left(\prod_{\text{p prime} < 23{,}143} 1-\frac{1}{p}\right)^2 (3 \cdot 6,342,314,060)$$

Merten's theorem states that

$$\prod_{\substack{p \text{ prime} \\ p \leq x}} \left(1-\frac{1}{p}\right) \sim \frac{1}{e^\gamma \ln(x)}$$

With this approximation there are 59,481,223 numbers to check after pre-sieve.

2.3. **Sieving.** In $f(g(x))$ we had $x = k\left(\frac{23,143\#}{43}\right)$, thus sieving was designed to eliminate values of k for which $f(g(x))$ was a multiple of a prime p.

To do this, look for roots of the polynomials, $h(x)$ *and* $f(g(x))$*, in* $\mathbb{F}_p$.

Factor $h(x) = 40x^2 + x + 1 \bmod p$ and compute its roots in $\mathbb{F}_p$. Then, eliminate the choices of k for which the corresponding x value is a root. Repeat this process on $f(g(x))$. The number of numbers left after sieving up to $5 \cdot 10^{10}$ is approximately

$$(59, 481, 223) \cdot \prod_{\substack{p \text{ prime} \\ p \leq x}} \left(1 - \frac{1}{p}\right)^2$$

By Merten's Theorem approximately, 9,914,204 numbers would remain after sieving.

2.4. **Computations.** The computations were done on a cluster of 1200 nodes, each running tests on groups of 900 numbers.

- Run Fermat pseudo primality tests to find probable primes base 3 on OpenPFGW. ($3^{x-1} \equiv 1 \bmod x$). This generates a list of 20,000 digit pseudo primes.
- Periodically check the corresponding 60,000 digit numbers for pseudo primality base 3.
- Test the 60,000 digit pseudo prime for primality.

The total amount of CPU time used was:

- 5 days for sieving
- 535 days for 20,000 digit pseudo-primality tests
- 4 days 60,000 digit pseudo-primality tests.

3. Result

3.1. **Theorem(Justin Debenedetto and Jeremy Rouse).** Let $f(x) = x^2+x+41$ and $g(x) = 40x^3 + 41x^2 + 42x + 1$. If we set

$$x = \frac{310927391 \cdot 23143\#}{43}$$

then $f(g(x))$ is a 60,000 digit prime number.