Quad Sieve for twin and Sophie Germain primes

Yves Gallot

March 4, 2022

If we search for twin and Sophie Germain prime numbers, we first test the primality of $k\,2^n - 1$. If it is prime, then we check the primality of $k\,2^n + 1$, $k\,2^{n-1} - 1$ and $k\,2^{n+1} - 1$. A quad sieve ensures that the four forms don't have any small prime divisor.

Based on Bateman and Horn conjecture, the expected numbers of remaining candidates, primes of the form $k\,2^n - 1$, twin primes and of Sophie Germain primes in a fixed sieved range are calculated in this note.

Conjecture (Bateman and Horn).
Let $f_1, f_2, \cdots, f_m$ be a set of m distinct irreducible polynomials with integral coefficients and a positive leading coefficient. Let f be their product and suppose that f does not vanish identically modulo any prime. Let $w(p)$ be the number of solutions to the congruence $f(x) \equiv 0 \pmod{p}$. An integer n is prime-generating if every polynomial $f_i(n)$ produces a prime number. π_f is the number of prime-generating integers for $n_1 \leq n \leq n_2$. If $n_2 - n_1 \to \infty$ then

$$\begin{aligned}
\pi_f &\sim C_f \sum_{n=n_1}^{n_2} \frac{1}{\log f_1(n) \cdot \log f_2(n) \cdots \log f_m(n)} \\
&\sim \frac{C_f}{D_f} \int_{n_1}^{n_2} \frac{dt}{(\log t)^m} \\
&\sim \frac{C_f}{D_f} \left(\frac{n_2}{(\log n_2)^m} - \frac{n_1}{(\log n_1)^m} \right),
\end{aligned}$$

where D_f is the product of the degrees of the polynomials $f_1, f_2, \cdots, f_m$ and

$$C_f = \prod_{p \text{ prime}} \frac{1 - w(p)/p}{(1 - 1/p)^m}.$$

C_f is a sieving constant. Let S be a set containing n m-tuples of random integers. S is sieved for $p \leq p_{max}$: an element is removed from the set if at least one integer of the m-tuple is divisible by p. The expected number of remaining elements is $\prod_{p \leq p_{max}} (1 - 1/p)^m \, n$. If the m-tuples of the set are generated with the polynomials $f_i(n)$ then the expected number of remaining elements is $\mathrm{cand}_{p_{max}}(n) = \prod_{p \leq p_{max}} (1 - w(p)/p) \, n$.

$k\,2^n - 1$ is odd and $C_{k\,2^n - 1} = 2$.

For twin primes, we have $f_1(x) = 2x - 1$ and $f_2(x) = 2x + 1$. $w(2) = 0$ and $w(p) = 2$ for $p > 2$. $C_{(k\,2^n - 1,\, k\,2^n + 1)} = 4\,C_2$, where C_2 is the twin prime constant.
$C_{(k\,2^n - 1,\, k\,2^n + 1)} \approx 2.6406473$.

$f_1(x) = 2x - 1$ and $f_2(x) = 4x - 1$ for Sophie Germain primes. We still have $w(2) = 0$ and $w(p) = 2$ for $p > 2$ then $C_{(k\,2^n-1,\,k\,2^{n+1}-1)} = C_{(k\,2^n-1,\,k\,2^n+1)}$.

The functions of the quadruplet are $f_1(x) = 2x - 1$, $f_2(x) = 4x - 1$, $f_3(x) = 8x - 1$ and $f_4(x) = 4x + 1$. $w(2) = 0$, $w(3) = 2$ and $w(p) = 4$ for $p > 3$.
$C_{(k\,2^{n-1}-1,\,k\,2^n-1,\,k\,2^{n+1}-1,\,k\,2^n+1)} \approx 8.3023617$.

In the following table, the formula with the summation is applied.

Number of prime patterns: k is odd, $1 \leq k \leq 10^8$

	Form	Actual	Expected.
single	$8k-1$	5143331	5143482.8
twin	$8k-1, 8k+1$	350779	350497.4
Sophie Germain I	$4k-1, 8k-1$	363795	363558.9
Sophie Germain II	$8k-1, 16k-1$	338858	338346.4
quadruplet	$4k-1, 8k-1, 8k+1, 16k-1$	3042	2978.2

Mertens' third theorem is $\lim_{n\to\infty} \log n \prod_{p\leq n}(1-1/p) = e^{-\gamma}$, where $\gamma \approx 0.57721566$ is the Euler's constant. We have $\text{cand}_{p_{max}}(n) = \prod_{p\leq p_{max}}(1 - w(p)/p)\, n \approx C_f \prod_{p\leq p_{max}}(1-1/p)^m\, n$, hence

$$\text{cand}_{p_{max}}(n) \approx C_f \left(\frac{e^{-\gamma}}{\log p_{max}}\right)^m n.$$

For quad sieve, we have

$$\text{cand}_{p_{max}}(n) \approx \frac{0.82504063\, n}{(\log p_{max})^4}.$$

If N has no prime factor $p \leq p_{max}$ then the likelihood that N is prime is

$$\mathscr{P}(N) = \prod_{p\leq p_{max}} (1-1/p)^{-1} \frac{1}{\log N} \approx \frac{\log p_{max}}{e^{-\gamma} \log N}.$$

Let S be a set of pairs (k, n). A quad sieve is applied to this set. The expected number of primes is

$$\pi_{k\,2^n-1} \approx \frac{1.4694571}{(\log p_{max})^3} \sum_S \frac{1}{\log(k\,2^n-1)}.$$

If n is fixed and k is odd, $k_1 \leq k \leq k_2$ and $1 \ll k\,2^n$, we have:

$$\pi_s \approx \frac{0.73472856}{(\log p_{max})^3} \int_{k_1}^{k_2} \frac{dt}{\log(t\,2^n)}.$$

Applying $\mathscr{P}(N)$ to twin primes:

$$\pi_t \approx \frac{1.30860477}{(\log p_{max})^2} \int_{k_1}^{k_2} \frac{dt}{\log(t\,2^n)^2},$$

and $\pi_{SG} \approx 2\,\pi_t$.

Quad sieve prime occurrences:

$p_{max} = 10^4$, $n = 100$, k is odd, $1 \le k \le 10^9$

	Actual	Expected.
candidates	57069	57324.8
primes	10461	10562.8
twin	1950	1946.6
Sophie Germain	3825	3893.2

$p_{max} = 10^5$, $n = 100$, k is odd, $1 \le k \le 10^9$

	Actual	Expected.
candidates	23660	23480.2
primes	5419	5408.2
twin	1280	1245.8
Sophie Germain	2481	2491.6

$p_{max} = 10^5$, $n = 1000$, k is odd, $1 \le k \le 10^{10}$

	Actual	Expected.
candidates	234241	234802.4
primes	6784	6732.2
twin	175	193.0
Sophie Germain	392	386.1

If $n = 3321925$ then $k\,2^n - 1$ has more than one million digits. For $1 \le k \le 2^{52}$ the expected values are

p_{max}	candidates	primes	twin	SG
10^{12}	3,187,000,000	68,120	1.46	2.91
10^{13}	2,314,000,000	53,578	1.24	2.48
10^{14}	1,720,000,000	42,898	1.07	2.14
10^{15}	1,306,000,000	34,877	0.93	1.86

In practice, we can set $3321925 \le n \le 3321925 + 63$ and $1 \le k \le 2^{46}$.