

# Primes Generated by Recurrence Sequences

G. Everest, S. Stevens, D. Tamsett and T. Ward

20th January 2006

## 1 Mersenne Numbers and Primitive Prime Divisors

A notorious problem from elementary number theory is the Mersenne prime conjecture. This says that the *Mersenne sequence*  $M = (M_n)$ , defined by

$$M_n = 2^n - 1 \text{ for } n \geq 1,$$

will contain infinitely many prime terms — known as *Mersenne primes*.

The Mersenne prime conjecture is related to a classical problem in number theory concerning *perfect numbers*. A whole number is said to be perfect if, like  $6 = 1 + 2 + 3$  and  $28 = 1 + 2 + 4 + 7 + 14$ , it is equal to the sum of all its divisors apart from itself. Euclid pointed out that  $2^{k-1}(2^k - 1)$  is perfect whenever  $2^k - 1$  is prime. A much less obvious result, due to Euler, is a partial converse: if  $n$  is an *even* perfect number, then it must have the form  $2^{k-1}(2^k - 1)$  for some  $k$  with the property that  $2^k - 1$  is a prime. Whether there are any *odd* perfect numbers remains an open question. Thus finding Mersenne primes amounts to finding (even) perfect numbers.

The sequence  $M$  certainly produces some primes initially, for example

$$M_2 = 3, M_3 = 7, M_5 = 31, M_7 = 127, \dots$$

However, the appearance of Mersenne primes quickly thins out; only 43 are known, the largest of which,  $M_{30,402,457}$ , has over 9 million decimal digits. This was discovered by a team at Central Missouri State University as part of the GIMPS project [23], which harnesses idle time on thousands of computers all over the world to run a distributed version of the Lucas-Lehmer test.

Only 43 primes might seem rather a small return for such a huge effort. Anybody looking for gold or gems with the same level of success would surely abandon the search. It seems fair to ask why we should expect there to be infinitely many Mersenne primes. In the absence of a rigorous proof, our expectations may be informed by *heuristic* arguments. In Section 3 we will discuss heuristic arguments for this and other more or less tractable problems in Number Theory.

## 1.1 Primitive prime divisors

In 1892, Zsigmondy [24] discovered a beautiful argument which shows that the sequence  $M$  does yield infinitely many prime numbers – but in a less restrictive sense.

Given any integer sequence  $S = (S_n)_{n \geq 1}$ , define a *primitive divisor* of the term  $S_n \neq 0$  to be a divisor of  $S_n$  which is coprime to every nonzero term  $S_m$  with  $m < n$ . Any prime factor of a primitive divisor is called a *primitive prime divisor*. Factorizing the first few terms of the Mersenne sequence reveals several primitive divisors, shown in bold in Table 1. Notice that the term  $M_6$  has

Table 1: Primitive divisors of  $(M_n)$ .

$n$	$M_n$	Factorization
2	<b>3</b>	<b>3</b>
3	7	<b>7</b>
4	15	3 · <b>5</b>
5	31	<b>31</b>
6	63	3 <sup>2</sup> · 7
7	127	<b>127</b>
8	255	3 · 5 · <b>17</b>
9	511	7 · <b>73</b>
10	1023	3 · <b>11</b> · 31

no primitive divisor, but all the other early terms do have one. Zsigmondy [24] proved that all the terms  $M_n$  for  $n > 6$  have a primitive divisor. He also proved a similar result for more general sequences  $U = (U_n)_{n \geq 1}$ , of the form  $U_n = a^n - b^n$ , whenever  $a > b$  are positive coprime integers:  $U_n$  has a primitive divisor unless

$$a = 2, b = 1 \text{ and } n = 6$$

or

$$a + b \text{ is a power of 2 and } n = 2.$$

Away from the special situation where  $a - b = 1$ , it is not reasonable to expect the terms  $U_n = a^n - b^n$  to ever be prime, since the identity

$$a^n - b^n = (a - b)(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \dots + b^{n-1})$$

shows that  $U_n$  is divisible by  $a - b$ . However, it does seem likely that for any coprime starting values  $a$  and  $b$ , infinitely many terms of the sequence  $(U_n/(a - b))$  will be prime. Sadly, no proof of this plausible statement is known for even a single pair of starting values.

Although Zsigmondy's result is much weaker than the Mersenne Prime Conjecture, it initiated a great deal of interest in the arithmetic of such sequences (see [10, Chap. 6]). It has also been applied in finite group theory; see Praeger [15] for example. Schinzel [16], [18] extended Zsigmondy's result, giving further insight into the finer arithmetic of sequences like  $M$ . For example, he proved that  $M_{4k}$  has a composite primitive divisor for all odd  $k > 5$ .

## 2 Recurrence sequences

### 2.1 Linear recurrence sequences

For most people, their first introduction to the Fibonacci sequence

$$A_1 = 1, A_2 = 1, A_3 = 2, A_4 = 3, A_5 = 5, A_6 = 8, \dots$$

is through the (binary) linear recurrence relation

$$A_{n+2} = A_{n+1} + A_n.$$

Sequences such as the Mersenne sequence  $M$  and those considered by Zsigmondy are of particular interest because they also satisfy binary recurrence relations. The terms  $U_n = a^n - b^n$  satisfy the recurrence

$$U_{n+2} = (a + b)U_{n+1} - abU_n, \text{ for all } n \geq 1.$$

More generally, let  $u$  and  $v$  denote conjugate quadratic integers; in other words, zeros of a monic irreducible polynomial with integer coefficients. Consider the integer sequences  $U(u, v)$  and  $V(u, v)$  defined by

$$U_n(u, v) = (u^n - v^n)/(u - v) \quad \text{and} \quad V_n(u, v) = u^n + v^n.$$

For instance, the Fibonacci sequence is given by

$$A_n = U_n\left(\frac{1}{2}(1 + \sqrt{5}), \frac{1}{2}(1 - \sqrt{5})\right).$$

The sequence  $U(u, v)$  satisfies the recurrence relation

$$U_{n+2} = (u + v)U_{n+1} - uvU_n, \text{ for all } n \geq 1,$$

and  $V(u, v)$  satisfies the same relation.

Some powerful generalizations of Zsigmondy's Theorem have been obtained for these sequences. Bilu, Hanrot and Voutier [3] used methods from Diophantine analysis to prove that both  $U_n(u, v)$  and  $V_n(u, v)$  have a primitive divisor for any  $n > 30$ . The two striking aspects of this result are the uniform nature of the bound and its small numerical value. In particular, for any given sequence it is easy to check the first 30 terms for primitive divisors, arriving at a complete picture. For example, an easy calculation reveals that the Fibonacci sequence  $A_n$  does not have a primitive divisor if and only if  $n = 1, 2, 6$  or  $12$ .

### 2.2 Bilinear recurrence sequences

The theory of linear recurrence sequences has a *bilinear* analogue. For example, the Somos-4 sequence  $S = (S_n)$  satisfies the bilinear recurrence relation

$$S_{n+4}S_n = S_{n+3}S_{n+1} + S_{n+2}^2, \text{ for all } n \geq 1$$

with the initial condition  $S_1 = S_2 = S_3 = S_4 = 1$ . This sequence begins

1, 1, 1, 1, 2, 3, 7, 23, 59, 314, 1 529, 8 209, 833 313, 620 297, 7 869 898, . . . .

Remarkably, all the terms are integers even though calculating  $S_{n+4}$  *a priori* involves dividing by  $S_n$ . This sequence was discovered by Michael Somos [20] and it is known to be associated to the arithmetic of elliptic curves. See [10, Sect. 10.1 and 11.1] for a summary of this, and further references, including a remarkable observation due to Propp *et al.* that the terms of the sequence must be integers because they count matchings in a sequence of graphs.

Amongst the early terms of  $S$ , several are prime: of those listed above,

2, 3, 7, 23, 59, 8 209 and 620 297

are prime. It seems natural to ask if there are infinitely many prime terms. More generally, consider integer sequences  $S$  satisfying a relation

$$S_{n+4}S_n = eS_{n+3}S_{n+1} + fS_{n+2}^2, \tag{1}$$

where  $e$  and  $f$  are integral constants not both zero. Such sequences are often called *Somos Sequences* or *bilinear recurrence sequences* and Christine Swart [21], building on earlier remarks of Nelson Stephens, showed how they are related to the arithmetic of elliptic curves. Some care is needed because, for example, a binary linear recurrence sequence will always satisfy a bilinear recurrence relation of this kind: we will refer to a Somos sequence as *non-linear* if it does not satisfy any linear recurrence relation. These are natural generalizations of linear recurrence sequences, so perhaps we should expect them to contain infinitely many prime terms. Computational evidence in [5] tended to support that belief because of the relatively large primes discovered. However, a heuristic argument (discussed later) using the Prime Number Theorem was adapted in [7] and it suggested that a non-linear Somos sequence should contain only finitely many prime terms. See [9] for proofs in some special cases.

On the other hand Silverman [19] showed a qualitative analogue of Zsigmondy's result for elliptic curves which applies, in particular, to the Somos-4 sequence. An explicit form of this result proved by Everest, McLaren and Ward [8] guarantees that from  $S_5$  onwards, all terms have a primitive divisor. There are many non-linear Somos sequences to which Silverman's proof does not apply. A version of Zsigmondy's Theorem valid for these sequences awaits discovery.

### 2.3 Polynomials

Given the previous sections, it might be tempting to think that all integral recurrence sequences have primitive divisors from some point. However it is easy to write down counterexamples. The sequence  $T = (T_n)$  defined by  $T_n = n$ , which satisfies

$$T_{n+2} = 2T_{n+1} - T_n$$

is a binary linear recurrence sequence which does not always produce primitive divisors. This is a rather trivial counterexample, so consider now the sequence  $P$  defined by

$$P_n = n^2 + \beta.$$

The terms of this sequence satisfy a linear recurrence relation,

$$P_{n+3} = 3P_{n+2} - 3P_{n+1} + P_n \text{ for all } n \geq 1.$$

It has long been suspected that for any fixed  $\beta$  with  $-\beta$  not a square, the sequence  $P$  will contain infinitely many prime terms. No proof is known even for one value of  $\beta$ . It seems reasonable to ask the apparently simpler question about the existence of primitive divisors of terms. Clearly any prime term will itself be a primitive divisor – but do the composite terms have primitive divisors? Using a result of Schinzel about the largest prime factor of the terms in polynomial sequences it is fairly easy to prove the following.

**Theorem 2.1.** *Suppose  $-\beta$  is not a square. There are infinitely many terms of the sequence  $P$  which do not have a primitive divisor.*

Theorem 2.1 will be proved in Section 4. Computations suggest that the following stronger result should be true.

**Conjecture 2.2.** *Suppose  $-\beta$  is not a square and let  $\rho_\beta(N)$  denote the number of terms  $P_n$ , with  $n < N$ , having a primitive divisor in the sequence  $P$ . Then*

$$\rho_\beta(N) \sim cN \text{ for some constant } 0 < c < 1.$$

Here, and throughout the article, for functions

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

and

$$g : \mathbb{R} \rightarrow \mathbb{R}_+,$$

we write  $f \sim g$  to mean  $f(x)/g(x) \rightarrow 1$  as  $x \rightarrow \infty$ .

In the last section of the article, we consider some approaches to bounding the number of terms which have a primitive divisor. For example, we will demonstrate a simple proof that

$$\liminf_{N \rightarrow \infty} \frac{\rho_\beta(N)}{N} \geq \frac{1}{2}.$$

We cannot find a proof of Conjecture 2.2 so in the next section we will show how other kinds of arguments can be marshalled in its support, as well as discussing briefly the nature of the constant  $c$ .

## 2.4 Linear recurrence sequences

To set matters in a more general context, define  $L = (L_n)_{n \geq 1}$  to be a *linear recurrence sequence of order  $k$*  if it satisfies a relation

$$L_{n+k} = c_{k-1}L_{n+k-1} + \cdots + c_0L_n, \text{ for all } n \geq 1 \quad (2)$$

for constants  $c_0, \dots, c_{k-1}$ , but satisfies no shorter relation. When  $k = 3$  (respectively 4), the sequence  $L$  is called a *ternary (quaternary) linear recurrence sequence*. For example, the sequences  $P$  considered in the previous section are all ternary linear recurrence sequences. Theorem 2.1 shows that Zsigmondy's theorem cannot extend to these quadratic sequences. Some non-polynomial sequences that cannot satisfy Zsigmondy will now be presented.

With  $u$  and  $v$  again denoting conjugate quadratic integers, the integer sequence  $W(u, v) = (W_n(u, v))_{n \geq 1}$  defined by

$$W_n(u, v) = (u^n - 1)(v^n - 1)$$

is always a linear recurrence sequence.

**Example 2.3.** 1. The sequence  $B = -W(2 + \sqrt{3}, 2 - \sqrt{3})$  begins

$$2, 12, 50, 192, 722, 2700, 10082, 37632, 140450, 524172, \dots$$

and it is a ternary sequence satisfying

$$B_{n+3} = 5B_{n+2} - 5B_{n+1} + B_n.$$

All of the terms of  $B$  have primitive divisors.

2. The sequence  $C = -W(1 + \sqrt{2}, 1 - \sqrt{2})$  begins

$$2, 4, 14, 32, 82, 196, 478, 1152, 2786, 6724, \dots$$

and it is a quaternary sequence satisfying

$$C_{n+4} = 2C_{n+3} + 2C_{n+2} - 2C_{n+1} - C_n.$$

In contrast to the previous example, the terms  $C_{2k}$ , for odd  $k$ , do not have primitive divisors.

In general, when  $uv = -1$ , the terms  $W_{2k}(u, v)$ , for odd  $k$ , fail to yield primitive divisors. This is because an easy calculation shows that

$$W_{2k}(u, v) = -W_k(u, v)^2$$

when  $k$  is odd. On the other hand, we recommend the following as an exercise: when  $uv = 1$ , the terms  $W_n(u, v)$  do produce primitive divisors from some point on. As far as we can tell, to establish this requires Schinzel's extension [18] of Zsigmondy's result to the algebraic setting; we are indebted to Professor Györy for communicating to us the remarks about  $W(u, v)$ .

All of these special cases can be subsumed into a wider picture. Write

$$f(x) = x^k - c_{k-1}x^{k-1} - \dots - c_0$$

for the *characteristic polynomial* of the linear recurrence relation in (2). Then  $f$  can be factorized over  $\mathbb{C}$ ,

$$f(x) = (x - \alpha_1)^{e_1} \dots (x - \alpha_d)^{e_d}.$$

The algebraic numbers  $\alpha_1, \dots, \alpha_d$ , are known as the *characteristic roots* (or just *roots*) of the sequence. The terms  $L_n$  of any sequence  $L$  satisfying the relation (2) can be written

$$L_n = \sum_{i=1}^d g_i(n) \alpha_i^n$$

for polynomials  $g_1, \dots, g_d$  of degrees  $e_1 - 1, \dots, e_d - 1$  with algebraic coefficients.

The roots of the two sequences in Example 2.3 are quite different in character. In general, if  $uv = 1$  then  $W(u, v)$  is a ternary linear recurrence sequence with roots  $1, u, v$ . When  $uv = -1$  then  $W(u, v)$  is quaternary with roots  $1, -1, u, v$ . For the quadratic sequence defined by  $P_n = n^2 + \beta$ ,  $\alpha = 1$  is a triple root of the associated characteristic polynomial.

It seems reasonable to conjecture that the terms of an integral linear recurrence sequence of order  $k \geq 2$  will have a primitive divisor from some point on provided the roots are distinct and no pairwise quotient of distinct roots is a root of unity.

### 3 Heuristic Arguments

There are a number of ways that mathematicians form a view on which statements are likely to be true. These views inform research directions and help to concentrate effort on the most fruitful areas of enquiry.

The only certainty in mathematics comes from rigorous proofs that adhere to the rules of logic; the discourse of *logos*. When such a proof is not available, other kinds of arguments can make mathematicians expect that statements will be true, even though these arguments fall well short of a proof. These are called *heuristic* arguments – the word comes from the Greek root *Ευρηκα* (*Eureka*) meaning, ‘I have found it’. It usually means the principles used to make decisions in the absence of complete information or the ability to examine all possibilities. In informal ways, mathematicians use heuristic arguments all the time when they discuss mathematics, and these are part of the *mythos* discourse in mathematics.

One consequence of the Prime Number Theorem is the following statement: the probability that  $N$  is prime is roughly  $1/\log N$ . What this means is that if an element of the set  $\{1, \dots, N\}$  is chosen at random using a fair  $N$ -sided die for each  $N$ , then the probability  $\rho_N$  that the number chosen is prime satisfies  $\rho_N \log N \rightarrow 1$  as  $N \rightarrow \infty$ . This crude estimate has been used several times

to argue heuristically in favour of the plausibility of difficult problems. Some examples follow — in each case the argument presented falls well short of a proof, yet it still seems to have some predictive power and has suggested lines of attack.

### 3.1 Fermat primes

Hardy and Wright [11, Sect. 2.5] argued along these lines that there ought to be only finitely many Fermat primes. A Fermat prime is a prime number of the form

$$F_n = 2^{2^n} + 1.$$

Fermat demonstrated that  $F_1, \dots, F_4$  are all primes; Euler showed that  $F_5$  is composite using congruence arguments. Since then many Fermat numbers have been shown to be composite and quite a few have been completely factorized. Wilfrid Keller maintains a web site [12] with details of the current state of knowledge on factorization of Fermat numbers. The number of Fermat primes  $F_n$  with  $n < N$ , if they are no more or less likely to be prime than a random number of comparable size, should be roughly

$$\sum_{n < N} \frac{1}{\log F_n} \sim \sum_{n < N} \frac{1}{2^n \log 2} < \frac{1}{\log 2} \quad \text{for all } N \geq 1.$$

Statements like this cannot be taken too literally — the numbers  $F_n$  have many special properties, not all of which are understood. However, this kind of argument tends to support the belief that there are only finitely many Fermat primes, and would incline many mathematicians to attempt to prove that statement rather than its negation. Massive advances in computing power suggest that we know — indeed, that Fermat knew — all the Fermat primes.

### 3.2 Mersenne primes

The Prime Number Theorem can also be used to argue in support of the Mersenne Prime Conjecture. A heuristic argument of the following form is used. First,  $2^k - 1$  can only be prime for  $k$  a prime, so assume now that  $k$  is a prime  $p$ . We would like to estimate the probability that  $2^p - 1$  is prime. The prime number theorem suggests that a random number of the size of  $2^p - 1$  is prime with probability  $1/\log(2^p - 1)$ , which is around  $1/p \log 2$ . However,  $2^p - 1$  is far from random: it is not divisible by 2, nor by 3, and indeed not by any prime smaller than  $2p$ . Arguing in this way suggests that the probability that  $2^p - 1$  is prime is approximately

$$\rho_p = \frac{1}{p \log 2} \cdot \frac{2}{1} \cdot \frac{3}{2} \cdot \frac{5}{4} \cdots \frac{q}{q-1} \tag{3}$$

where  $q$  is the largest prime less than  $2p$ . This suggests that the expected number of Mersenne primes  $M_n$  with  $n < N$  is roughly  $\sum_{p < N} \rho_p$ .

Since  $\rho_p > 1/p \log 2$ , the sum diverges by Mertens' Theorem (see (4) for a precise statement), which suggests that there are infinitely many Mersenne primes. Wagstaff [22], then Pomerance and Lenstra [13] have extended this heuristic argument by including estimates for the product of rationals in (3) to obtain an asymptotic estimate that closely matches the available evidence. On the basis of these heuristics, they conjecture that the number of Mersenne primes  $M_n$  with  $n < N$  is asymptotically

$$\frac{e^\gamma}{\log 2} \log N,$$

where  $\gamma$  is the Euler-Mascheroni constant. Caldwell's Prime Page [4] gives more details about these arguments and about the hunt for new Mersenne primes.

### 3.3 Bilinear recurrence sequences

Consider now the Somos sequences defined by the recurrence (1). General results about heights on elliptic curves show that the growth rate of  $S_n$  is quadratic-exponential. In other words,

$$\log S_n \sim hn^2,$$

where  $h$  is a positive constant. Thus the expected number of prime terms with  $n < N$  should be approximately

$$\sum_{n < N} \frac{1}{\log S_n} \sim \frac{1}{h} \sum_{n < N} \frac{1}{n^2} \leq \frac{\pi^2}{6h} \text{ for all } N.$$

This resembles the argument of Hardy and Wright, and suggests that only finitely many prime terms should be expected. Proofs of the finiteness in many special cases have subsequently been found [9] and the search for these proofs was motivated in part by the heuristic arguments. Interestingly, it is known that the constant  $h$  is uniformly bounded below across all non-linear integral Somos sequences. Thus the style of heuristic argument suggests that perhaps the total number of prime terms is uniformly bounded across all such sequences. Extensive calculation has failed to yield a sequence with more than a dozen prime terms.

### 3.4 Quadratic polynomials

Suppose  $\beta$  is an integer which is not the negative of a square and recall the sequence  $P$  given by  $P_n = n^2 + \beta$ . Again, the Prime Number Theorem predicts roughly

$$\sum_{n < N} \frac{1}{\log P_n}$$

prime terms in the sequence  $P$  with  $n < N$ , if  $P_n$  is neither more nor less likely to be prime than a random number of that size. The sum is asymptotically  $N/2 \log N$  which supports the belief that there are infinitely many prime

terms in the sequence  $P$ . Computation suggests that for fixed  $\beta$  there will be  $dN/\log N$  prime terms with  $n < N$ , where  $d = d(\beta)$  is a constant which depends upon  $\beta$ . Bateman and Horn [2] gave a heuristic argument and numerical evidence to suggest that

$$d = \frac{1}{2} \prod_p \left(1 - \frac{1}{p}\right)^{-1} \left(1 - \frac{w(p)}{p}\right),$$

where the product is taken over all primes and  $w(p)$  denotes the number of solutions  $x \pmod{p}$  to the congruence  $x^2 \equiv -\beta \pmod{p}$ .

## 4 Biased Numbers

We now return to the problem of looking for primitive prime factors in the sequence given by  $P_n = n^2 + \beta$  with  $-\beta$  not a square. Since we are mainly interested in asymptotic behaviour, assume from now on that  $n > |\beta|$ . The terms  $P_n$  with  $n \leq |\beta|$  are not guaranteed to fit the behaviour described below.

**Lemma 4.1.** *A prime  $p$  is a primitive divisor of  $P_n$  if and only if  $p$  divides  $P_n$  and  $p > 2n$ .*

*Proof.* Consider first a prime  $p < n$  dividing  $P_n$ . Then, by assumption,  $P_n \equiv 0 \pmod{p}$ , so  $P_m \equiv 0 \pmod{p}$  for some  $m < p$  simply by choosing  $m$  to be the residue of  $n \pmod{p}$ . By assumption,  $p \leq n$  so  $m < n$ . In other words,  $p$  is not a primitive divisor of  $P_n$ .

This means that to find primitive divisors of  $P_n$  we have to look for prime divisors which are greater than  $n$ . We can say more: We can guarantee a solution of  $P_m \equiv 0 \pmod{p}$  for some  $m \leq p/2$ . Thus, to find primitive divisors we have only to look amongst the prime divisors which are bigger than  $2n$ . Thus a prime  $p$  dividing  $P_n$  is a primitive divisor only if  $p > 2n$ .

Conversely, suppose that  $p$  is a prime dividing  $P_n$  which is not a primitive divisor. Then  $n^2 + \beta \equiv 0 \pmod{p}$  and there is an integer  $m < n$  with  $m^2 + \beta \equiv 0 \pmod{p}$  so (by subtracting the two congruences)  $m^2 - n^2 \equiv 0 \pmod{p}$ . It follows that  $m \pm n \equiv 0 \pmod{p}$ , and in particular

$$p \leq m + n < 2n.$$

It follows that a prime  $p$  is a primitive divisor of  $P_n$  if and only if  $p$  divides  $P_n$  and  $p > 2n$ .  $\square$

**Definition 4.2.** *An integer  $k$  is biased if it has a prime factor  $q$  with  $q > 2\sqrt{k}$ .*

Thus any prime greater than 3 is biased; 22, 26, 34 are biased, while 24 and 28 are not.

**Proposition 4.3.** *For all  $n > |\beta|$ , the term  $P_n$  has a primitive divisor if and only if  $P_n$  is biased. For all  $n > |\beta|$ , if  $P_n$  has a primitive divisor, then that primitive divisor is a prime and it is unique.*

*Proof.* Part of the first statement comes from Lemma 4.1. To complete the proof of the first statement we claim that, for  $n$  greater than  $|\beta|$ ,  $P_n$  has such a prime divisor if and only if  $P_n$  is biased: If  $p$  is a prime dividing  $P_n$  and  $p > 2n$  then

$$p \geq 2n + 1 > 2\sqrt{n^2 + n} > 2\sqrt{n^2 + \beta}.$$

Conversely, if  $p > 2\sqrt{n^2 + \beta}$  then

$$p \geq 2\sqrt{n^2 - n + 1} > 2n - 1$$

so  $p > 2n$  (since  $2n$  cannot be prime).

The uniqueness of the primitive divisor follows at once. If  $p$  is a prime dividing  $P_n$  and  $p > 2\sqrt{P_n}$ , then no other prime divisor can be as large, so cannot be primitive.  $\square$

The requirement  $n > |\beta|$  is necessary: if  $|\beta|$  is prime then  $P_{|\beta|}$  has primitive divisor  $|\beta|$  but is not biased. Also, terms with small  $n$  may have more than one primitive divisor: for example, the sequence of values of the polynomial  $n^2 + 6$  begins 7, 10, ... so the second term has two primitive prime divisors. The kind of results discussed here are asymptotic results, so this restriction is unimportant.

*Proof of Theorem 2.1.* Results of Schinzel [17, Th. 13] show that for any  $\alpha > 0$ , the largest prime factor of  $P_n$  is bounded above by  $n^\alpha$  for infinitely many  $n$ . Taking  $\alpha = 1$ , it follows that  $P_n$  is not biased infinitely often. By Proposition 4.3,  $P_n$  fails to have a primitive divisor infinitely often.  $\square$

In the final section of the article, we consider quantitative information about the frequency with which, rather than the extent to which,  $P_n$  is not biased.

Support for Conjecture 2.2 follows from Proposition 4.3 because an asymptotic formula can be obtained for the distribution of biased numbers. Alongside the earlier notation for describing the growth rates of various functions, we will also use the following: Given functions

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

and

$$g : \mathbb{R} \rightarrow \mathbb{R}_+,$$

write  $f = O(g)$  to mean that  $|f(x)|/g(x)$  is bounded, and  $f = o(g)$  to mean that  $f(x)/g(x) \rightarrow 0$  as  $x \rightarrow \infty$ .

**Theorem 4.4.** *Let  $\pi_i(N)$  denote the number of biased numbers less than or equal to  $N$ . Then*

$$\pi_i(N) \sim N \log 2 \quad \text{as } N \rightarrow \infty.$$

*Proof.* Write a biased number as  $qm$  where  $q$  is the largest prime factor, so that the biased condition is  $q > 4m$ . To compute the number of biased numbers below  $N$ , note that the counting can be achieved by dividing the set into two parts. Let  $p$  denote a variable prime. For  $p < 2\sqrt{N}$  there are  $\lfloor p/4 \rfloor$  biased

integers  $pm < N$ . For  $p \geq 2\sqrt{N}$ , each  $pm < N$  is biased so there are  $\lfloor \frac{N}{p} \rfloor$  biased integers  $pm$ . Hence the total number is

$$\sum_{p < 2\sqrt{N}} \left\lfloor \frac{p}{4} \right\rfloor + \sum_{2\sqrt{N} \leq p < N} \left\lfloor \frac{N}{p} \right\rfloor.$$

The first sum is  $O(N/\log N)$  so it can be ignored asymptotically. The second sum differs from

$$N \sum_{2\sqrt{N} \leq p < N} \frac{1}{p}$$

by an amount which is  $O(N/\log N)$  by the Prime Number Theorem. To estimate this sum use Mertens' Formula, which may be found in Apostol's book [1, Th. 4.12],

$$\sum_{p < x} \frac{1}{p} = \log \log x + A + o(1). \quad (4)$$

Applying this gives

$$N \left[ \log \log N + A - \log(\log \sqrt{N} + \log 2) - A + o(1) \right]$$

which is asymptotically  $N \log 2$ . □

Theorem 4.4 can be applied to give the following heuristic argument in support of Conjecture 2.2. The probability that a large integer is biased is roughly  $\log 2$ . Hence the expected number of biased values of  $n^2 + \beta$  with  $n < N$  is asymptotically  $N \log 2$ . Computational evidence suggests that the number of biased terms in  $n^2 + \beta$  is asymptotically  $cN$  for some constant  $c$ . Computations with  $|\beta| < 10$  suggest the constant  $c$  looks reasonably close to  $\log 2$  in each case, although convergence appears slow.

## 5 Terms With and Terms Without Primitive Divisors

The article concludes with some simple estimates for  $\rho_\beta(N)$ , the number of terms  $P_n$  with  $n < N$  having a primitive divisor in the sequence  $P$ . The proofs use little apart from well-known estimates for sums over primes, which can be found in the book of Apostol [1].

**Theorem 5.1.** *There is a constant  $C > 0$  such that*

$$\rho_\beta(N) < N - \frac{CN}{\log N} \quad \text{for all sufficiently large } N. \quad (5)$$

*There is a constant  $D > 0$  such that*

$$\frac{N}{2} - \frac{DN}{\log N} < \rho_\beta(N) \quad \text{for all sufficiently large } N. \quad (6)$$

Both of the statements in Theorem 5.1 can be strengthened along the following lines: any choice of sufficiently large constant  $C$  or sufficiently small constant  $D$  could be made – as each constant varies, so does the smallest value of  $N$  beyond which the inequalities become valid.

Apart from a finite number of primes, any prime  $p$  that divides  $n^2 + \beta$  has the property that  $-\beta$  is a quadratic residue modulo  $p$ . Let  $\mathcal{R}$  denote the set of odd primes for which  $-\beta$  is a quadratic residue. Notice that  $\mathcal{R}$  comprises the intersection of a finite union of arithmetic progressions with the set of primes, and this finite union of arithmetic progressions in turn comprises exactly half of the residue classes modulo  $4|\beta|$ . We will prove the two parts of Theorem 5.1 in reverse order, because the upper bound (5) arises by specializing the argument used to prove the lower bound (6).

Write

$$Q_N = \prod_{n=1}^N |P_n|$$

and denote by  $\omega(Q_N)$  the number of distinct prime divisors of  $Q_N$ . By Proposition 4.3 it is sufficient to bound  $\omega(Q_N)$  because, with finitely many exceptions, a primitive divisor is unique.

## 5.1 Proof of the Lower Bound

Define

$$\mathcal{S} = \{p \in \mathcal{R} \mid p|Q_N \text{ and } p < 2N\}$$

and

$$\mathcal{S}' = \{p \in \mathcal{R} \mid p|Q_N \text{ and } p \geq 2N\}.$$

Let  $s = |\mathcal{S}|$  and  $s' = |\mathcal{S}'|$ . We seek a lower bound for  $s + s'$ , since

$$s + s' = \omega(Q_N).$$

The asymptotic form of Dirichlet's Theorem<sup>1</sup> on primes in arithmetic progression implies that asymptotically half the primes lie in  $\mathcal{R}$ , so

$$s \sim \frac{N}{\log 2N}. \tag{7}$$

Therefore it is sufficient to estimate  $s'$  from below.

---

<sup>1</sup>In 1826 Dirichlet proved that if  $a$  and  $b$  are positive integers with no common factor, then there are infinitely many primes of the form  $ax + b$  with  $x \in \mathbb{N}$ . This result, which appeared in a memoir published in 1837 [6], was proved using methods from analysis, thus laying the foundations for the subject now called analytic number theory. Writing  $\pi_a(X)$  for the number of primes of the form  $ax + b$  with  $x < X$ , Dirichlet proved that  $\pi_a(X) \rightarrow \infty$  as  $X \rightarrow \infty$ . There is also what might be called a prime number theorem for arithmetic progressions, which gives an asymptotic estimate for the number of such primes; this states that  $\pi_a(X) \sim X/\phi(a) \log X$  where  $\phi(a)$  is the Euler  $\phi$ -function of  $a$ . This was shown by de la Vallée Poussin; a proof may be found in the book of Prachar [14, Sect. V.7]. It is this result that we are using here.

*Proof of (6).* From the definition of  $Q_N$ ,

$$\begin{aligned}\log Q_N &= \sum_{n=1}^N \log |n^2 + \beta| \\ &= 2 \sum_{n=1}^N \left( \log n + O\left(\frac{1}{n^2}\right) \right) \\ &= \left( 2 \sum_{n=1}^N \log n \right) + O(1),\end{aligned}$$

so by Stirling's Formula

$$\log Q_N = 2N \log N - 2N + O(1). \quad (8)$$

On the other hand, we may write

$$\sum_{p|Q_N} e_p \log p = \log Q_N, \quad (9)$$

corresponding to the prime decomposition  $\prod_{p|Q_N} p^{e_p}$  of  $Q_N$ , for positive integers  $e_p$ . The first step in the proof is to identify a subset of  $\mathcal{R}$  which contributes a fixed amount to the main term in (8). The sum on the left-hand side of (9) may be decomposed to give

$$\sum_{p \in \mathcal{S}, p < N} e_p \log p + \sum_{p \in \mathcal{S}, p \geq N} e_p \log p + \sum_{p \in \mathcal{S}'} \log p = \log Q_N, \quad (10)$$

noting that  $e_p = 1$  whenever  $p \geq 2N$ . The second term in the decomposition is  $O(N)$ , since  $e_p \leq 2$  for  $p \in \mathcal{S}$  with  $p > N$ , each term in the sum is no larger than  $\log 2N$ , and finally the prime number theorem implies that there are  $O(N/\log N)$  terms. Thus the second term does not contribute to the asymptotic behaviour.

Assume for the moment that

$$\sum_{p \in \mathcal{S}, p < N} e_p \log p = N \log N + O(N). \quad (11)$$

Combining (8), (10) and (11) gives

$$N \log N + O(N) = \sum_{p \in \mathcal{S}'} \log p < s' \log P_N = s' \log(N^2 + \beta).$$

Thus (6) follows at once, subject to the proof of (11).  $\square$

*Proof of Equation (11).* For every  $p \in \mathcal{S}$ ,

$$e_p \geq \left\lfloor \frac{2N}{p} \right\rfloor.$$

Hence there is a constant  $c > 0$  such that the left-hand side of (11) is bounded below by

$$\sum_{p \in \mathcal{S}, p < N} \left\lfloor \frac{2N}{p} \right\rfloor \log p.$$

By Apostol [1, Th. 7.3],

$$\sum_{p \in \mathcal{S}, p < N} \left\lfloor \frac{2N}{p} \right\rfloor \log p = N \log N + O(N). \quad (12)$$

For each  $p \in \mathcal{S}$  and  $k \in \mathbb{N}$ , denote by  $\text{ord}_p(k)$  the index of the greatest power of  $p$  dividing  $k$  and put

$$\mathcal{B}_p(N) = \{n < N \mid \text{ord}_p(P_n) > 1\}.$$

Then

$$\sum_{p \in \mathcal{S}, p < N} e_p \log p = \sum_{p \in \mathcal{S}, p < N} \left\lfloor \frac{2N}{p} \right\rfloor \log p + \sum_{p \in \mathcal{S}, p < N} \left( \sum_{n \in \mathcal{B}_p(N)} \text{ord}_p(P_n) - 1 \right) \log p.$$

We will now show the second term is asymptotically negligible. For each  $p \in \mathcal{S}$ ,  $-\beta$  has two  $p$ -adic square roots, and  $\text{ord}_p(P_n) = r + 1$  if and only if the  $p$ -adic expansion of  $n$  agrees with one of these square roots up to the term in  $p^r$  and no further. Hence

$$\begin{aligned} \sum_{p \in \mathcal{S}, p < N} \left( \sum_{n \in \mathcal{B}_p(N)} \text{ord}_p(P_n) - 1 \right) \log p &\leq \sum_{p \in \mathcal{S}, p < N} \left( \sum_{r=1}^{\frac{\log P_N}{\log p}} r \cdot 2 \left\lfloor \frac{N}{p^{r+1}} \right\rfloor \right) \log p \\ &< 2N \sum_{p \in \mathcal{S}, p < N} \frac{\log p}{(p-1)^2} + 2s \log P_N, \end{aligned}$$

which is  $O(N)$  since the sum converges and  $s = O(N/\log N)$  by (7). Putting this together with (12) shows (11) as required.  $\square$

## 5.2 Proof of the Upper Bound

This proof is similar; however it relies on a finer partition of the set  $\mathcal{R}$ . Given integers  $K > 2$  and  $N > K$ , split  $\mathcal{S}'$  into the sets

$$\begin{aligned} \mathcal{T} &= \{p \in \mathcal{R} \mid p|Q_N, 2N < p < KN\}; \\ \mathcal{U} &= \{p \in \mathcal{R} \mid p|Q_N, KN < p\}. \end{aligned}$$

*Proof of (5).* Write  $t = |\mathcal{T}|$  and  $u = |\mathcal{U}|$ . As before, the contribution from  $s$  is negligible. Thus we wish to bound the expression  $t + u$  from above. The sum on the left-hand side of (9) decomposes according to the definitions of  $\mathcal{S}$ ,  $\mathcal{T}$  and  $\mathcal{U}$  to give

$$\sum_{p \in \mathcal{S}} e_p \log p + \sum_{p \in \mathcal{T}} \log p + \sum_{p \in \mathcal{U}} \log p = \log Q_N,$$

noting as before that  $e_p = 1$  whenever  $p > N$ . Equations (8), (10) and (11) show that

$$\sum_{p \in \mathcal{T}} \log p + \sum_{p \in \mathcal{U}} \log p < N \log N + aN$$

for some  $a > 0$ . The left-hand side is greater than

$$t \log N + u \log(KN)$$

so add  $t \log K$  to both sides to obtain

$$(t + u) \log(KN) < N \log N + aN + t \log K.$$

Rearranging the right-hand side gives

$$(t + u) \log(KN) < N \log(KN) + (a - \log K)N + t \log K.$$

Assume  $K$  is fixed so that  $C = \log K - a > 0$ . Dividing through by  $\log(KN)$  gives

$$(t + u) < N - \frac{CN}{\log(KN)} + \frac{t \log K}{\log(KN)}. \quad (13)$$

For any  $x > 0$ ,  $-\frac{1}{1+x} < -1 + x$ . Apply this with  $x = \log K / \log N$  to the second term on the right of (13) to give

$$-\frac{CN}{\log(KN)} < -\frac{CN}{\log(N)} + O\left(\frac{N}{(\log N)^2}\right),$$

whose last term is asymptotically negligible. The last term on the right of (13) can be estimated using Dirichlet's Theorem again, giving

$$\frac{t \log K}{\log(KN)} = O\left(\frac{t}{\log N}\right) = O\left(\frac{N}{(\log N)^2}\right)$$

which is also asymptotically negligible. Hence, if  $C > 0$  is sufficiently large, then

$$\omega(Q_N) \sim t + u < N - \frac{CN}{\log N}$$

for all large  $N$ . □

A slightly stronger result is provable with these methods, namely

$$\rho_\beta(N) < N - \frac{N \log \log N}{\log N} \text{ for all sufficiently large } N.$$

We leave this as an exercise to the interested reader.

## References

1. T. M. Apostol, *Introduction to analytic number theory*, Springer-Verlag, New York, 1976, Undergraduate Texts in Mathematics.
2. P. T. Bateman and R. A. Horn, A heuristic asymptotic formula concerning the distribution of prime numbers, *Math. Comp.* **16** (1962), 363–367.
3. Y. Bilu, G. Hanrot, and P. M. Voutier, Existence of primitive divisors of Lucas and Lehmer numbers, *J. Reine Angew. Math.* **539** (2001), 75–122, With an appendix by M. Mignotte.
4. C. Caldwell, *The Prime Pages*, [www.utm.edu/research/primes/](http://www.utm.edu/research/primes/).
5. D. V. Chudnovsky and G. V. Chudnovsky, Sequences of numbers generated by addition in formal groups and new primality and factorization tests, *Adv. in Appl. Math.* **7** (1986), no. 4, 385–434.
6. P. G. Lejeune Dirichlet, Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält, *Abhand. Ak. Wiss. Berlin* **48–51** (1837). [Reprinted in *Werke* Vol. I, 315–342; G. Reimer, Berlin (1889)].
7. M. Einsiedler, G. R. Everest, and T. Ward, Primes in elliptic divisibility sequences, *LMS J. Comput. Math.* **4** (2001), 1–15.
8. G. Everest, G. McLaren, and T. Ward, Primitive divisors of elliptic divisibility sequences, *J. Number Theory* (2006), to appear.
9. G. Everest, V. Miller, and N. Stephens, Primes generated by elliptic curves, *Proc. Amer. Math. Soc.* **132** (2004), no. 4, 955–963.
10. G. R. Everest, A. J. van der Poorten, I. Shparlinski, and T. Ward, *Recurrence sequences*, Mathematical Surveys and Monographs, vol. 104, American Mathematical Society, Providence, RI, 2003.
11. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, fifth ed., The Clarendon Press Oxford University Press, New York, 1979.
12. W. Keller, *Prime factors  $k \cdot 2^n + 1$  of Fermat numbers  $F_m$  and complete factoring status*, [www.prothsearch.net/fermat.html](http://www.prothsearch.net/fermat.html).
13. H. W. Lenstra, Jr., *Primality testing*, Mathematics and computer science (Amsterdam, 1983), CWI Monogr., vol. 1, North-Holland, Amsterdam, 1986, pp. 269–287.
14. K. Prachar, *Primzahlverteilung*, Springer-Verlag, Berlin-New York, 1978. [Reprint of the 1957 original].

15. C. E. Praeger, Primitive prime divisor elements in finite classical groups, in *Groups St. Andrews 1997 in Bath, II*, Cambridge Univ. Press, Cambridge, 1999, pp. 605–623.
16. A. Schinzel, On primitive prime factors of  $a^n - b^n$ , *Proc. Cambridge Philos. Soc.* **58** (1962), 555–562.
17. A. Schinzel, On two theorems of Gelfond and some of their applications, *Acta Arith.* **13** (1967/1968), 177–236.
18. A. Schinzel, Primitive divisors of the expression  $A^n - B^n$  in algebraic number fields, *J. Reine Angew. Math.* **268/269** (1974), 27–33.
19. J. H. Silverman, Wieferich’s criterion and the *abc*-conjecture, *J. Number Theory* **30** (1988), no. 2, 226–237.
20. M. Somos, Problem 1470, *Cruce Mathematicorum* **15** (1989), 208.
21. C. Swart, *Elliptic Curves and Related Sequences*, PhD Thesis, University of London (2003).
22. S. S. Wagstaff, Jr., Divisors of Mersenne numbers, *Math. Comp.* **40** (1983), no. 161, 385–397.
23. G. W. Woltman, *GIMPS*, [www.mersenne.org/prime.htm](http://www.mersenne.org/prime.htm).
24. K. Zsigmondy, Zur Theorie der Potenzreste, *Monatsh. Math.* **3** (1892), 265–284.