Primes Generated by Recurrence Sequences

Graham Everest, Shaun Stevens, Duncan Tamsett, and Tom Ward

October 12, 2012

1. MERSENNE NUMBERS AND PRIMITIVE PRIME DIVISORS.

A notorious problem from elementary number theory is the "Mersenne Prime Conjecture." This asserts that the *Mersenne sequence* $M = (M_n)$ defined by

$$M_n = 2^n - 1 \quad (n = 1, 2, \dots)$$

contains infinitely many prime terms, which are 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 proper divisors. 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 forty-three are known, the largest of which, $M_{30,402,457}$, has over nine million decimal digits. This was discovered by a team at Central Missouri State University as part of the GIMPS project [26], which harnesses idle time on thousands of computers all over the world to run a distributed version of the Lucas–Lehmer test.

A paltry forty-three 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 discuss heuristic arguments for this and other more or less tractable problems in number theory.

Primitive prime divisors. In 1892, Zsigmondy [27] discovered a beautiful argument that shows that the sequence M does yield infinitely many prime numbers—but in a less restrictive sense. Given any integer sequence $S =$

$(S_n)_{n\geq 1}$, we define a *primitive divisor* of the term S_n ($\neq 0$) to be a divisor of S_n that is coprime to every nonzero term S_m with $m < n$. Any prime factor of a primitive divisor is called a *primitive prime divisor*. Factoring 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 no primitive divisor, but each

Table 1: Primitive divisors of (M_n).

n	M_n	Factorization
2	3	$\mathbf{3}$
3	7	$\mathbf{7}$
4	15	$3 \cdot \mathbf{5}$
5	31	$\mathbf{31}$
6	63	$3^2 \cdot 7$
7	127	$\mathbf{127}$
8	255	$3 \cdot 5 \cdot \mathbf{17}$
9	511	$7 \cdot \mathbf{73}$
10	1023	$3 \cdot \mathbf{11} \cdot 31$

of the other early terms has at least one. Zsigmondy [27] proved that all the terms M_n ($n > 6$) have primitive divisors. He also proved a similar result for more general sequences $U = (U_n)_{n\geqslant 1}$, namely, those of the form $U_n = a^n - b^n$, where a and b ($a > b$) are positive coprime integers: U_n has a primitive divisor unless $a = 2, b = 1$ and $n = 6$ or $a + b$ is a power of 2 and $n = 2$. These results generalized earlier work of Bang [2], who considered the case $b = 1$.

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

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

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 $(U_n/(a-b))$ might 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 [12, chap. 6]). It has also been applied in finite group theory (see Praeger [18], for example). Schinzel [19], [21] 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 greater than five.

2. RECURRENCE SEQUENCES. For most people their first introduction to the Fibonacci sequence

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

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 linear recurrence relations. The terms $U_n = a^n - b^n$ satisfy the recurrence

$$U_{n+2} = (a+b)U_{n+1} - abU_n \qquad (n = 1, 2, \dots).$$

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

$$U_n(u, v) = (u^n - v^n)/(u - v), \;\; V_n(u, v) = u^n + v^n.$$

For instance, the Fibonacci sequence is given by

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

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

$$U_{n+2} = (u+v)U_{n+1} - uvU_n \qquad (n = 1, 2, \dots),$$

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 [4] used methods from Diophantine analysis to prove that both $U_n(u, v)$ and $V_n(u, v)$ have primitive divisors once $n > 30$. Three striking aspects of this result are the uniform nature of the bound, its small numerical value and the fact that it is sharp (the sequence $U\left(\frac{1+\sqrt{-7}}{2}, \frac{1-\sqrt{-7}}{2}\right)$ attains the bound). In particular, for any given sequence it is easy to check the first thirty terms for primitive divisors, arriving at a complete picture. For example, an easy calculation reveals that the Fibonacci number A_n does not have a primitive divisor if and only if $n = 1, 2, 6$, or 12.

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

$$S_{n+4}S_n = S_{n+3}S_{n+1} + S_{n+2}^2 \qquad (n = 1, 2, \dots),$$

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

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

Amazingly, 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 [23], and it is known to be associated with the arithmetic of elliptic curves (see [12, secs. 10.1, 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 are several primes: of those that we listed,

$$2, 3, 7, 23, 59, 8\,209, 620\,297$$

are prime. The 207th term (which has 1857 decimal digits) is also a prime. It seems natural to ask whether there are infinitely many prime terms in the Somos-4 sequence. More generally, consider integer sequences S satisfying relations of the type

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

where e and f are integral constants not both zero. Such sequences are often called *Somos sequences* (or *bilinear recurrence sequences*) and Christine Swart [24], 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 always satisfies some bilinear recurrence relation of this kind. We refer to a Somos sequence as *nonlinear* 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 [7] 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 [9], and it suggested that a nonlinear Somos sequence should contain only finitely many prime terms. See [11] for proofs in some special cases.

On the other hand, Silverman [22] established a qualitative analogue of Zsigmondy's result for elliptic curves that applies, in particular, to the Somos-4 sequence. An explicit form of this result proved by Everest, McLaren, and Ward [10] guarantees that from S_5 onwards all terms have primitive divisors. There are many nonlinear Somos sequences to which Silverman's proof does not apply. A version of Zsigmondy's theorem valid for these sequences awaits discovery.

Polynomials. Given the previous sections, it might be tempting to think that all integral recurrence sequences have primitive divisors from some point on. 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 that does not always produce primitive divisors. This is a rather trivial counterexample, so consider now the sequence P defined by

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

where β is a nonzero integer. The terms of this sequence satisfy the linear recurrence relation

$$P_{n+3} = 3P_{n+2} - 3P_{n+1} + P_n.$$

It has long been suspected that for any fixed β such that $-\beta$ is not a square the sequence P contains infinitely many prime terms. A proof is known for not even one value of β. It seems reasonable to ask the apparently simpler question about the existence of primitive divisors of terms. Clearly any prime term is itself 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. *If $-\beta$ is not a square, then there are infinitely many terms of the sequence P which do not have primitive divisors.*

We prove Theorem 2.1 in section 4. Computations suggest that the following stronger result should be true.

Conjecture 2.2. Suppose that $-\beta$ is not a square. If $\rho_\beta(N)$ denotes the number of terms P_n in the sequence P with $n < N$ that have primitive divisors, then

$$\rho_\beta(N) \sim cN$$

for some constant c satisfying $0 < c < 1$.

Here, and throughout the remainder of the article, for functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}_+$ we write $f \sim g$ to mean $f(x)/g(x) \to 1$ as $x \to \infty$.

In the last section of the article, we consider some approaches to bounding the number of terms in P that have primitive divisors. For example, we will furnish a simple proof that

$$\liminf_{N\to\infty} \frac{\rho_\beta(N)}{N} \geqslant \frac{1}{2}.$$

We have been unable to find a proof of Conjecture 2.2. In section 3 we show how other kinds of arguments can be marshalled in its support, and in section 4 we discuss briefly the nature of the constant c.

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

$$L_{n+k} = c_{k-1}L_{n+k-1} + \cdots + c_0 L_n \quad (n = 1, 2, \dots) \tag{2}$$

for constants $c_0, \dots, c_{k-1}$, but satisfies no shorter relation. When $k = 3$ (respectively, $k = 4$), the sequence L is called a *ternary (respectively, 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 nonpolynomial sequences that cannot satisfy Zsigmondy will now be presented.

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

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

is always a linear recurrence sequence.

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

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

and it is a ternary sequence satisfying

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

From the seventh term on, all the terms of the sequence seem to have primitive divisors.

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

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

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 reveals 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 [21] of Zsigmondy's result to the algebraic setting. (We are indebted to Professor Győry for communicating to us the remarks about $W(u, v)$.)

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

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

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

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

The algebraic numbers $\alpha_1, \ldots, \alpha_d$, 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

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

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

The roots of the sequences in Examples 1 and 2 are quite different in character. In general, if $uv = 1$, then $W(u, v)$ is a ternary linear recurrence sequence with roots $1, u$, and v. When $uv = -1$, $W(u, v)$ is quaternary with roots $1, -1, u$, and 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 greater than one will have primitive divisors from some point on provided that its roots are distinct and no quotient α_i/α_j of different roots is a root of unity.

3. HEURISTIC ARGUMENTS. There are a number of ways that mathematicians form views about 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 proofs. These are called *heuristic* arguments—the word comes from the Greek root *Ευρηκα* (*Eureka*), meaning "I have found it." 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 $\{1, \ldots, N\}$ is chosen at random using a fair N-sided die, then the probability ρ_N that the number chosen is prime satisfies $\rho_N \log N \to 1$ as $N \to \infty$. This crude estimate has been used several times to argue heuristically in favor of the plausibility of conjectured solutions of difficult problems. Some examples follow. In each case the argument presented does not have the force of a proof, yet it still seems to have some predictive power and has suggested lines of attack.

Fermat primes. Hardy and Wright [13, sec. 2.5] argued along these lines that there ought to be only finitely many Fermat primes. A *Fermat prime* is a prime number in the sequence (F_n) of *Fermat numbers*:

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

Fermat demonstrated that F_1, F_2, F_3, and F_4 are all primes; Euler showed that F_5 is composite by using congruence arguments. Since then, many Fermat numbers have been shown to be composite and quite a few have been completely factored. Wilfrid Keller maintains a web site [14] 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

$$\sum_{n<N} \frac{1}{\log F_n} \sim \sum_{n<N} \frac{1}{2^n \log 2} < \frac{1}{\log 2}.$$

Statements like this cannot be taken too literally, for 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.

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 be prime only 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

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

where q is the largest prime less than $2p$. This would lead one to think 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's theorem (see (4) for a precise statement), which suggests that there are infinitely many Mersenne primes. Wagstaff [25] and then Pomerance and Lenstra [16] 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 γ is the Euler-Mascheroni constant. Caldwell's Prime Page [5] gives more details about these arguments and about the hunt for new Mersenne primes.

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

$$\sum_{n<N} \frac{1}{\log S_n} \sim \frac{1}{h} \sum_{n<N} \frac{1}{n^2} \leqslant \frac{\pi^2}{6h}.$$

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 [11]. 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 nonlinear integral Somos sequences. Thus the style of this heuristic argument intimates that perhaps the total number of prime terms is uniformly bounded across all such sequences. Extensive calculation has failed to yield a sequence with many more than a dozen prime terms.

Quadratic polynomials. Suppose β is an integer that 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 that there are roughly

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

prime terms in the sequence P with $n < N$, assuming again that 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 β there will be $dN/\log N$ prime terms with $n < N$, where $d = d(\beta)$ is a constant that depends upon β. Bateman and Horn [3] offered a heuristic argument and provided numerical evidence to the effect 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 modulo 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, we assume from now on that $n > |\beta|$. The terms P_n with $n \leqslant |\beta|$ are not guaranteed to exhibit the behavior described in this section.

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 dividing P_n with $p < n$. Then, by assumption, $P_n \equiv 0 \pmod{p}$, so $P_m \equiv 0 \pmod{p}$ for some m smaller than p simply by choosing m to be the residue of n modulo p. Because $p < n$, $m < n$. In others 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 that are greater than n. (Note that n does not divide P_n, as $n > |\beta|$.) We can say more: we can guarantee a solution of $P_m \equiv 0 \pmod{p}$ for some m satisfying $m \leqslant p/2$. Thus, to find primitive divisors we have to look only amongst the prime divisors that are bigger than $2n$ (i.e., a prime p dividing P_n is a primitive divisor only if $p > 2n$).

Conversely, suppose that p is a prime dividing P_n that 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 $m \pm n \equiv 0 \pmod{p}$. In particular,

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

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

We call an integer k *biased* if it has a prime factor q with $q > 2\sqrt{k}$. Thus any prime greater than three is biased. The numbers $22, 26$, and 34 are biased, whereas 24 and 28 are not.

Proposition 4.2. *When $n > |\beta|$, the term P_n has a primitive divisor if and only if P_n is biased. If $n > |\beta|$ and 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

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

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

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

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 of can be as large, and hence no other prime divisor can be primitive. ☐

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 $7, 10, \ldots$, so the second term has two primitive prime divisors. The kind of results discussed here are asymptotic results, which makes this restriction unimportant.

Proof of Theorem 2.1. A result of Schinzel [20, Theorem 13] shows that for any positive α the largest prime factor of P_n is bounded above by n^α for infinitely many n. Taking $\alpha = 1$, we conclude that P_n is not biased infinitely often. By Proposition 4.2, P_n fails to have a primitive divisor infinitely often. ☐

In section 5, 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 4.2 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 also use the following: given functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}_+$, we write $f = O(g)$ to mean that $|f(x)|/g(x)$ is bounded and $f = o(g)$ to signify that $f(x)/g(x) \to 0$ as $x \to \infty$.

Theorem 4.3. *If $\pi_\imath(N)$ denotes the number of biased numbers less than or equal to N, then $\pi_\imath(N) \sim N \log 2$.*

Proof. Write a biased number as qm, where q is its largest prime factor. The biased condition then translates to $q > 4m$. 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. When $p < 2\sqrt{N}$ there are $\lfloor p/4 \rfloor$ biased integers pm smaller than N (here $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x). When $p \geqslant 2\sqrt{N}$, each number pm smaller than N is biased, so there are $\lfloor N/p \rfloor$ biased integers pm. Hence the total number is

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

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

$$N \sum_{2\sqrt{N} \leqslant p < N} \tfrac{1}{p}$$

by an amount that is $O(N/\log N)$ by the prime number theorem. To estimate this sum we use Mertens's formula, which can be found in Apostol's book [1, Theorem 4.12]:

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

Applying (4) thus estimates $\pi_\imath(N)$ as

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

which is asymptotically $N \log 2$. $\square$

Theorem 4.3 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 $N \log 2$. Computations furnish evidence that the number of biased terms in $n^2 + \beta$ is asymptotically cN for some constant c. When $|\beta| < 10$, they suggest that the constant c is reasonably close to $\log 2$ in each case, although convergence appears slow.

5. COUNTING PRIMITIVE DIVISORS. We conclude with some simple estimates for $\rho_\beta(N)$, the number of terms P_n in the sequence P with $n < N$ that have primitive divisors. The proofs use little aside 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*

$$\rho_\beta(N) < N - \frac{CN}{\log N} \tag{5}$$

holds for all sufficiently large N. There is a constant $D > 0$ such that

$$\frac{N}{2} - \frac{DN}{\log N} < \rho_\beta(N) \tag{6}$$

is true for all sufficiently large N.

Both of the statements in Theorem 5.1 can be strengthened along the following lines: any choice of constants C or 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 that 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 establish 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 4.2 it is sufficient to bound $\omega(Q_N)$ because, with finitely many exceptions, a primitive divisor is unique.

Proof of the lower bound. Define

$$\mathcal{S} = \{p \in \mathcal{R} : p|Q_N \text{ and } p < 2N\}, \quad \mathcal{S}' = \{p \in \mathcal{R} : p|Q_N \text{ and } p \geqslant 2N\}.$$

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[1] on primes in arithmetic progression implies that asymptotically half the primes lie in $\mathcal{R}$, so

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

[1] 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 [8], 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 $\pi_a(X) \to \infty$ as $X \to \infty$. 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. It states that $\pi_a(X) \sim X/\phi(a) \log X$, where ϕ is the Euler totient function. This was shown by de la Vallée Poussin; a proof can be found in the book of Prachar [17, chap. 5, sec. 7]. It is this result that we are using here.

Therefore it is sufficient to estimate s' from below.

Proof of inequality (6). From the definition of Q_N,

$$\log Q_N = \sum_{n=1}^{N} \log |n^2 + \beta| = 2\sum_{n=1}^{N} \left(\log n + O\left(\tfrac{1}{n^2}\right)\right) = \left(2\sum_{n=1}^{N} \log n\right) + O(1),$$

so by Stirling's formula

$$\log Q_N = 2N \log N - 2N + O(1). \tag{8}$$

On the other hand, we can write

$$\sum_{p|Q_N} e_p \log p = \log Q_N \tag{9}$$

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

$$\sum_{p\in\mathcal{S}, p<N} e_p \log p + \sum_{p\in\mathcal{S}, p\geqslant N} e_p \log p + \sum_{p\in\mathcal{S}'} \log p = \log Q_N, \tag{10}$$

noting that $e_p = 1$ whenever $p \geqslant 2N$. The second term in the decomposition is $O(N)$, since $e_p \leqslant 2$ for p in $\mathcal{S}$ with $p > N$, each term in the sum is no larger than $\log 2N$, and 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

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

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). □

Proof of equation (11). For each p in $\mathcal{S}$

$$e_p \geqslant \left\lfloor \tfrac{2N}{p} \right\rfloor,$$

so the left-hand side of (11) is bounded below by

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

By Apostol [1, Theorem 7.3],

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

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

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

Then

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

We now show that the second term is asymptotically negligible. For p in $\mathcal{S}$ the number $-\beta$ has two p-adic square roots, and $\mathrm{ord}_p(P_n) = r+1$ 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

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

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

Proof of the upper bound. This proof is similar to that for the lower bound. However, it relies on a finer partition of the set $\mathcal{R}$. Given integers $K > 2$ and $N > K$, we split $\mathcal{S}'$ into the sets

$$\mathcal{T} = \{p \in \mathcal{R} : p|Q_N, 2N < p < KN\}, \quad \mathcal{U} = \{p \in \mathcal{R} : p|Q_N, KN < p\}.$$

Proof of inequality (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 earlier that $e_p = 1$ whenever $p > 2N$. Equations (8), (10), and (11) reveal that

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

for some positive a. The left-hand side is greater than

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

so we 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 that K is fixed, with $C = \log K - a > 0$. Dividing through by $\log(KN)$ leads to

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

The inequality $-1/(1+x) < -1+x$ holds when $x > 0$. We apply this with $x = \log K / \log N$ to the second term on the right of (13) to obtain the inequality

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

whose last term is asymptotically negligible. The last term on the right of (13) can be estimated by appealing to Dirichlet's theorem again, yielding

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

which is also asymptotically negligible. Hence for any C' $(0 < C' < C)$,

$$\omega(Q_N) \sim t + u < N - \frac{C'N}{\log N}$$

for all large N. $\square$

A slightly stronger result is provable with these methods, namely, that

$$\rho_\beta(N) < N - \frac{N \log \log N}{\log N}$$

for all sufficiently large N. We leave this as an exercise to the interested reader.

Note added in proof. We are grateful to Prof. Schinzel for pointing out that some of our results are already in the literature. Chowla and Todd [6] call n *reducible* if n^2+1 is not biased, and show that the density of biased numbers is $\log 2$. Knödel [15] showed that if $f(x) = o(x \log x)$ and $\pi_f(N)$ denotes the number of $n < N$ with the property that n^2+1 has no prime factor exceeding $f(n)$, then $\limsup_{N\to\infty} \pi_f(N)/N \leqslant 0.5$. One of our results reduces (for the case $\beta = 1$) to Knödel's result with $f(N) = 2N$.

References

1. T. M. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, New York, 1976.

2. A. S. Bang, Taltheoretiske Undersøgelser, *Tidskrift f. Math.* **5** (1886), 70–80 and 130–137.

3. P. T. Bateman and R. A. Horn, A heuristic asymptotic formula concerning the distribution of prime numbers, *Math. Comp.* **16** (1962) 363–367.

4. 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; appendix by M. Mignotte.

5. C. Caldwell, *The Prime Pages*, `www.utm.edu/research/primes/`.

6. S. D. Chowla and John Todd, The density of reducible integers, *Canadian J. Math.* **1** (1949) 297–299.

7. 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) 385–434.

8. P. G. Lejeune Dirichlet, Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sing, unendlich viele Primzahlen enthält, *Abhand. Ak. Wiss. Berlin* **48–51** (1837); reprinted in *Werke*, vol. 1, G. Reimer, Berlin, 1889, pp.315–342.

9. M. Einsiedler, G. R. Everest, and T. Ward, Primes in elliptic divisibility sequences, *LMS J. Comput. Math.* **4** (2001) 1–15.

10. G. Everest, G. McLaren, and T. Ward, Primitive divisors of elliptic divisibility sequences, *J. Number Theory* **118** (2006) 71–89.

11. G. Everest, V. Miller, and N. Stephens, Primes generated by elliptic curves, *Proc. Amer. Math. Soc.* **132** (2004) 955–963.

12. G. R. Everest, A. J. van der Poorten, I. Shparlinski, and T. Ward, *Recurrence Sequences*, Mathematical Surveys and Monographs, no. 104, American Mathematical Society, Providence, 2003.

13. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5th ed., Clarendon Press, Oxford University Press, New York, 1979.

14. W. Keller, *Prime Factors* $k \cdot 2^n + 1$ *of Fermat Numbers* F_m *and Complete Factoring Status*, `www.prothsearch.net/fermat.html`.

15. W. Knödel, Reduzible Zahlen, *Monatsh. Math.* **54** (1950) 308–312.

16. H. W. Lenstra, Jr., Primality testing, in *Mathematics and Computer Science (Amsterdam, 1983)*, CWI monogr., vol. 1, North-Holland, Amsterdam, 1986, pp. 269–287.

17. K. Prachar, *Primzahlverteilung*, Springer-Verlag, Berlin, 1978.

18. C. E. Praeger, Primitive prime divisor elements in finite classical groups, in *Groups St. Andrews 1997 in Bath, II,* Cambridge University Press, Cambridge, 1999, pp. 605–623.

19. A. Schinzel, On primitive prime factors of $a^n - b^n$, *Proc. Cambridge Philos. Soc.* **58** (1962) 555–562.

20. ———, On two theorems of Gelfond and some of their applications, *Acta Arith.* **13** (1967/1968) 177–236.

21. ———, Primitive divisors of the expression $A^n - B^n$ in algebraic number fields, *J. Reine Angew. Math.* **268/269** (1974) 27–33.

22. J. H. Silverman, Wieferich's criterion and the *abc*-conjecture, *J. Number Theory* **30** (1988) 226–237.

23. M. Somos, Problem 1470, *Crux Mathematicorum* **15** (1989) 208.

24. C. Swart, *Elliptic Curves and Related Sequences*, Ph.D. thesis, University of London, 2003.

25. S. S. Wagstaff, Jr., Divisors of Mersenne numbers, *Math. Comp.* **40** (1983) 385–397.

26. G. W. Woltman, *GIMPS*, `www.mersenne.org/prime.htm`.

27. K. Zsigmondy, Zur Theorie der Potenzreste, *Monatsh. Math.* **3** (1892) 265–284.

GRAHAM EVEREST received a Ph.D. from King's College London in 1983 and joined the University of East Anglia mathematics department the same year. He works in number theory, specializing in the relationship between Diophantine problems and elliptic curves. He was ordained as a minister in the Church of England in 2005.
School of Mathematics, University of East Anglia, Norwich NR4 7TJ, U.K.
g.everest@uea.ac.uk

SHAUN STEVENS received a Ph.D. from King's College London in 1998. After research positions in Orsay, Münster, and Oxford he joined the University of East Anglia mathematics department in 2002. He works in representation theory, on the local Langlands program.
School of Mathematics, University of East Anglia, Norwich NR4 7TJ, U.K.
shaun.stevens@uea.ac.uk

DUNCAN TAMSETT received a Ph.D. in Marine Geophysics from the University of Newcastle upon Tyne in 1984. He is best described as a professional odd-jobs man, currently contributing to a software system for planning oil well paths, as well as looking after a system for enhancing side-scan sonar images of the sea-bed and characterizing/classifying sonar image texture. He recently moved to Inverness for the Scottish Highland and Island experience.
5 Drummond Crescent, Inverness IV2 4QW, U.K.
duncan@tamsetts.freeserve.co.uk

TOM WARD received a Ph.D. from the University of Warwick in 1989. After research positions at the University of Maryland College Park and Ohio State University, he joined the University of East Anglia mathematics department in 1992. He has been head of the department since 2002. He works in ergodic theory and its interactions with number theory.
School of Mathematics, University of East Anglia, Norwich NR4 7TJ, U.K.
t.ward@uea.ac.uk