#A75 INTEGERS 25 (2025)

A NUMBER-THEORETIC PROBLEM CONCERNING PSEUDO-REAL RIEMANN SURFACES

Gareth A. Jones
School of Mathematical Sciences, University of Southampton, Southampton, United Kingdom
G.A.Jones@maths.soton.ac.uk

Alexander K. Zvonkin
LaBRI, Université de Bordeaux, Talence, France
zvonkin@labri.fr

Received: 1/18/24 , Revised: 8/14/24, Accepted: 7/20/25, Published: 8/15/25

Abstract

Motivated by their research on automorphism groups of pseudo-real Riemann surfaces, Bujalance, Cirre and Conder have conjectured that there are infinitely many primes p such that $p+2$ has all its prime factors $q \equiv -1 \pmod 4$. We use theorems of Landau and Raikov to prove that the number of integers $n \leq x$ with only such prime factors q is asymptotic to $cx/\sqrt{\ln x}$ for a specific constant $c = 0.4865\ldots$. Heuristic arguments, following Hardy and Littlewood, then yield a conjecture that the number of such primes $p \leq x$ is asymptotic to $c' \int_2^x (\ln t)^{-3/2} dt$ for a constant $c' = 0.8981\ldots$. The theorem, the conjecture and a similar conjecture applying the Bateman–Horn Conjecture to other pseudo-real Riemann surfaces are supported by evidence from extensive computer searches.

1. Introduction

A compact Riemann surface is real (meaning definable, as a complex algebraic curve, over $\mathbb{R}$) if and only if it has an orientation-reversing automorphism of order 2; it is called *pseudo-real* if it has an orientation-reversing automorphism, but none of order 2. In [5], Bujalance, Cirre and Conder have proved that for each $g \geq 2$, the maximum order $M^+_{\mathrm{ab}}(g)$ of an abelian group of (orientation-preserving) automorphisms of a pseudo-real Riemann surface of genus g is at least g. To demonstrate the sharpness of this lower bound, their Theorem 4.8 presents a set $\mathcal{A}$ of integers g for which the cyclic group C_g of order g attains the upper bound $M^+_{\mathrm{ab}}(g) = g$. These have the form $g = p + 1$ where p is what we will call a BCC prime, defined as follows.

DOI: 10.5281/zenodo.16881820

Definition 1. A *BCC prime* is a prime number p such that $p+2$ has only prime factors $q \equiv -1 \pmod 4$.

The first sixteen BCC primes are

$$5, 7, 17, 19, 29, 31, 41, 47, 61, 67, 79, 97, 101, 127, 131 \text{ and } 137. \tag{1}$$

In [5], $\mathcal{A}$ is described as "a very large and possibly infinite set". If $\mathcal{A}$ is finite, then conceivably there is a better lower bound for $M^{+}_{\mathrm{ab}}(g)$, valid for all sufficiently large g (in particular, larger than all those in $\mathcal{A}$, which can then be regarded as small exceptions). To avoid this possibility, it is important to know whether $\mathcal{A}$ is infinite. We are therefore interested in the following problem.

Problem 1. Are there infinitely many BCC primes, or equivalently, is the set $\mathcal{A}$ infinite?

Of course, the answer will be positive if one can prove that there are infinitely many prime pairs $p, p+2$ with $p \equiv 1 \pmod 4$, but even without this extra congruence condition, the existence or otherwise of infinitely many prime pairs is a very difficult open problem. Nevertheless, this would follow from a proof of various conjectures in number theory, such as the Bateman–Horn Conjecture [3] or Schinzel's Hypothesis H [21]. Here we give a heuristic argument, supported by computational evidence for their asymptotic distribution, for the following.

Conjecture 1. There are infinitely many BCC primes.

It follows from a theorem of Iwaniec [10] that there are infinitely many primes p such that $p+2$ has only prime factors $q \equiv 1 \pmod 4$, whereas the corresponding result with $q \equiv -1 \pmod 4$ is unproved; see Subsection 3.2 for details.

In support of Conjecture 1 we use theorems of Landau [12] and of Raikov [18] to give, in Theorem 2, an asymptotic estimate of the form $cx/\sqrt{\ln x}$ (with c a specific constant) for the number of integers $n \le x$ with only prime factors $q \equiv -1 \pmod 4$. Heuristic arguments, inspired by Hardy and Littlewood [9], then allow us to make Conjecture 2, that the number of BCC primes $p \le x$ is asymptotic to

$$c' \int_2^x \frac{dt}{(\ln t)^{3/2}}$$

for a specific constant $c' = 0.8981751984\ldots$. We give numerical evidence, based on computer searches, to support both Theorem 2 and Conjecture 2. In addition, in Section 7, we briefly consider another set of primes, defined by similar but more complicated conditions, also arising from a construction in [5].

The formulae we obtain involve various multiplicative factors. These are defined in terms of well-known constants such as e, γ and π, together with constants such as $C(k,u)$ (defined later) which have been computed elsewhere to over 100 decimal places, so in principle these factors can be computed with similar accuracy. However, for simplicity we have usually presented numerical data to about ten significant figures, since this is adequate for the arguments we wish to present.

Although our main emphasis here is on number theory and computation, for readers interested in Riemann surfaces we have described the construction in [5] of pseudo-real surfaces in more detail in a final appendix. For the benefit of such readers, or any others unfamiliar with number theory, we have occasionally included explanations and citations for facts which may seem obvious to experts in that field.

2. Raikov's Theorem and Its Application

If $\mathcal{P}$ is any non-empty set of prime numbers, then let us define an integer $n \in \mathbb{N}$ to be a $\mathcal{P}$-*integer* if all its prime factors are elements of $\mathcal{P}$, and let $\mathcal{P}^*$ denote the set of all $\mathcal{P}$-integers. Note that 1 is a $\mathcal{P}$-integer, represented by the empty product. If we define $g_n = 1$ or 0 as $n \in \mathcal{P}^*$ or not, then the function

$$g(x) := \sum_{n \leq x} g_n \tag{2}$$

gives the number of $\mathcal{P}$-integers $n \leq x$. We will consider the asymptotic behavior of $g(x)$ as $x \to \infty$.

In [12] Landau showed that if, for some $k \in \mathbb{N}$, $\mathcal{P}$ is the set of all primes in the union of l distinct congruence classes of units (mod k), then

$$g(x) \sim \frac{cx}{(\ln x)^{1-l/\varphi(k)}} \quad \text{as} \quad x \to \infty$$

for some constant $c > 0$ depending on $\mathcal{P}$, where φ is Euler's totient function. Motivated by Problem 1, we take $k = 4$ and define $\mathcal{P}$ to be the set of primes in the congruence class $[-1] \in \mathbb{Z}_4^*$, so that Conjecture 1 asserts that $n-2$ is prime for infinitely many $n \in \mathcal{P}^*$. From Landau's result we see that

$$g(x) \sim \frac{cx}{\sqrt{\ln x}} \quad \text{as} \quad x \to \infty$$

for some $c > 0$. Our first aim is to determine this constant c, and then to compare the resulting estimates with the actual values of $g(x)$ for various x.

For any set $\mathcal{P}$ of primes, the corresponding Dirichlet series

$$F(s) := \sum_{n=1}^{\infty} \frac{g_n}{n^s} \tag{3}$$

converges absolutely for $\mathrm{Re}(s) > 1$ by comparison with the Riemann zeta function $\zeta(s)$, so that it represents a holomorphic function in this half plane. (For the basic properties of Dirichlet series, see [2, Ch. 11] or [11, Ch. IV, §2], for example.) The function $n \mapsto g_n$ is completely multiplicative, so $F(s)$ has an Euler product expansion

$$F(s) = \prod_{q \in \mathcal{P}} \left(1 - \frac{1}{q^s}\right)^{-1} = \prod_{q \in \mathcal{P}} \frac{1}{1 - q^{-s}}. \tag{4}$$

A theorem of Raikov [18], as given in [6, Theorem 2.4.1], states the following.

Theorem 1. *Let $F(s) = \sum_{n \geq 1} a_n/n^s$ be a Dirichlet series with non-negative coefficients, converging for $\mathrm{Re}(s) > 1$. Suppose that $F(s)$ extends analytically at all points on $\mathrm{Re}(s) = 1$ apart from $s = 1$, and that at $s = 1$ we can write*

$$F(s) = \frac{H(s)}{(s-1)^{1-\alpha}} \tag{5}$$

for some $\alpha \in \mathbb{R}$ and some $H(s)$ holomorphic in the region $\mathrm{Re}(s) \geq 1$ and nonzero there. Then

$$\sum_{n \leq x} a_n \sim \frac{cx}{(\ln x)^\alpha} \tag{6}$$

as $x \to \infty$, with

$$c = \frac{H(1)}{\Gamma(1-\alpha)} \tag{7}$$

where Γ is the Gamma function.

In the case of Problem 1 we take

$$\mathcal{P} := \{q \mid q \text{ is prime and } q \equiv -1 \pmod 4\}.$$

Thus the $\mathcal{P}$-integers $n \leq 100$ are

$$1, 3, 7, 9, 11, 19, 21, 23, 27, 31, 33, 43, 47, 49, 57, 59, 63, 67, 69, 71, 77, 79, 81, 83, 93, 99. \tag{8}$$

These 26 integers represent just over half of the odd integers $n \leq 100$. However, this proportion decreases towards 0 as the upper bound increases: see Theorem 2 and Table 1.

Next we will determine the function H and hence the constants α and c for the sequence (g_n) corresponding to this set $\mathcal{P}$. We will do this by expressing the corresponding function F in terms of ζ and related functions, so that the analyticity conditions for F required in Theorem 1 will follow from similar properties of these functions.

The L-function

$$L(s) = L(s, \chi) := \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod_{q \text{ prime}} \left(1 - \frac{\chi(q)}{q^s}\right)^{-1} = \prod_{q \text{ prime}} \frac{1}{1 - \chi(q) q^{-s}} \tag{9}$$

corresponding to the character χ (mod 4) generated by $\chi(3) = -1$ is holomorphic on the half plane $\mathrm{Re}(s) > 1$; analytic continuation extends it to an entire function on $\mathbb{C}$ (see [2, Theorem 12.5], for example).

Comparing the Euler product expansions for ζ and L shows that

$$\frac{(1-2^{-s})\,\zeta(s)}{L(s)} = \prod_{q\equiv-1(4)} \frac{1+q^{-s}}{1-q^{-s}} = \prod_{q\equiv-1(4)} \frac{1-q^{-2s}}{(1-q^{-s})^2} = \frac{F(s)^2}{F(2s)} \tag{10}$$

for $\mathrm{Re}(s) > 1$, where the products are over the indicated primes q. Writing this as

$$F(s)^2 = \frac{(1-2^{-s})\zeta(s)}{L(s)} F(2s) \tag{11}$$

gives an analytic continuation of $F(s)^2$ to the half plane $\mathrm{Re}(s) > 1/2$; it is meromorphic there, with a simple pole at $s = 1$, and it may have poles in the strip $1/2 < \mathrm{Re}(s) < 1$ arising from possible non-trivial zeros of $L(s)$ away from its critical line. It may have zeros in this strip arising from possible non-trivial zeros of $\zeta(s)$, but it has no zeros with $\mathrm{Re}(s) = 1$. It follows that there is a simply connected region $\Omega \subset \mathbb{C}$, containing the half plane $\mathrm{Re}(s) \geq 1$, such that $h(s) := (s-1)F(s)^2$ is a non-vanishing holomorphic function on Ω, and hence $H(s) := \sqrt{h(s)} = \sqrt{s-1}F(s)$ is a non-zero holomorphic function on Ω, where $\sqrt{\cdot}$ has the usual branch. Thus $F(s)$ satisfies the hypotheses of Theorem 1, with $\alpha = 1/2$.

In order to find the constant c in Theorem 1 we need to find $H(1)$. To make the transition from $\mathrm{Re}(s) > 1$ to $s = 1$ we will use Abel's continuity theorem, with real $s \to 1+$. From (10),

$$\begin{aligned}\frac{F(s)^2}{\zeta(s)} &= \left(1-\frac{1}{2^s}\right)\cdot \prod_{q\equiv-1(4)} \left(1-\frac{1}{q^s}\right)^{-2} \cdot \prod_{q\equiv\pm1(4)} \left(1-\frac{1}{q^s}\right) \\ &= \left(1-\frac{1}{2^s}\right)\cdot \prod_{q\equiv\pm1(4)} \left(1-\frac{1}{q^s}\right)^{\pm1}\end{aligned}$$

for $\mathrm{Re}(s) > 1$, so that in this half plane

$$(s-1)F(s)^2 = \left(1-\frac{1}{2^s}\right)\cdot \prod_{q\equiv\pm1(4)} \left(1-\frac{1}{q^s}\right)^{\pm1} \cdot (s-1)\zeta(s). \tag{12}$$

We need to check that the product on the right-hand side converges when $s = 1$, and to find its limit. The first and last factors have values $1/2$ and 1 (the residue of $\zeta(s)$ at its simple pole $s = 1$).

To deal with the infinite product in (12), we can use some results of Languasco and Zaccagnini [13, 14, 15]. For each integer $k \geq 3$ and each unit u (mod k) they define a non-zero Mertens-type constant $C(k,u)$ by the asymptotic estimate

$$\begin{aligned} P(k,u;x) &:= \prod_{x \geq q \equiv u(k)} \left(1 - \frac{1}{q}\right) \\ &= \frac{C(k,u)}{(\ln x)^{1/\varphi(k)}} + O\left(\frac{1}{(\ln x)^{1+1/\varphi(k)}}\right) \quad \text{as} \quad x \to \infty, \end{aligned} \tag{13}$$

where the product is over all primes $q \equiv u \pmod{k}$ such that $q \leq x$. In [14, Equation (2)] they show that

$$C(k,u)^{\varphi(k)} = e^{-\gamma} \prod_{q \text{ prime}} \left(1 - \frac{1}{q}\right)^{\alpha(q)} \tag{14}$$

where γ is the Euler–Mascheroni constant, the product is now over all primes q, and $\alpha(q) := \varphi(k) - 1$ or -1 as $q \equiv u \pmod{k}$ or not. By taking $k = 4$ and $u = 1$ in (14), we see that

$$\prod_{q \equiv \pm 1(4)} \left(1 - \frac{1}{q}\right)^{\pm 1} = e^{\gamma} C^2 \left(1 - \frac{1}{2}\right) = \frac{e^{\gamma} C^2}{2},$$

where $C := C(4,1)$. Thus, when $s = 1$ the right-hand side of (12) converges to $e^{\gamma} C^2/4$. It therefore follows from Abel's theorem that

$$H(1)^2 = \lim_{s \to 1+} (s-1) F(s)^2 = \frac{e^{\gamma}}{4} C^2 \neq 0,$$

so that this particular instance of (5) becomes $F(s) = H(s)/(s-1)^{1/2}$ and hence

$$\alpha = \frac{1}{2} \quad \text{and} \quad H(1) = \sqrt{e^{\gamma}} \cdot \frac{C}{2}.$$

Since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, we therefore have

$$c = \frac{H(1)}{\Gamma(\frac{1}{2})} = \sqrt{\frac{e^{\gamma}}{\pi}} \cdot \frac{C}{2}, \tag{15}$$

giving the following result[1].

Theorem 2. *The function $g(x)$ satisfies*

$$g(x) \sim \frac{cx}{\sqrt{\ln x}} = \sqrt{\frac{e^{\gamma}}{\pi}} \cdot \frac{C}{2} \cdot \frac{x}{\sqrt{\ln x}} \quad \textit{as} \quad x \to \infty. \tag{16}$$

[1] Here we cannot resist expressing our pleasure at the appearance of the constant $\sqrt{e^{\gamma}/\pi}$, in which the three basic constants e, π and γ of analysis are united by three of the basic operations (division, exponentiation and taking square roots) of algebra; its application to number theory adds to the pleasure.

In [14, 15] Languasco and Zaccagnini have evaluated many of these constants $C(k,u)$ to over 100 decimal places. These include

$$C = C(4,1) = 1.29230415712868860710913838987043206534 29\ldots, \qquad (17)$$

which, together with

$$\sqrt{\frac{e^{\gamma}}{\pi}} = 0.75294950604642059593549975758760481083 86\ldots,$$

allows us to evaluate

$$c = 0.48651988838589099712724564058682340553 82\ldots \qquad (18)$$

We use this value in Table 1 to test the accuracy of the resulting estimates given by Theorem 2. The second column gives the number $g(x)$ of $\mathcal{P}$-integers $n \le x$ for $x = 10^k$ with $k = 1, \ldots, 10$. For example, the 26 such integers $n \le 10^2$ are listed in (8). Just 12 of these correspond to BCC primes $p = n - 2$, namely the first 12 primes displayed in (1), with 12 associated genera $g = p + 1 = n - 1 \in \mathcal{A}$. The third column of Table 1 gives the estimates $cx/\sqrt{\ln x}$ for $g(x)$, and the fourth column gives their errors.

x	$g(x)$	$cx/\sqrt{\ln x}$	error	$E(x)$	error
10	4	3.21	−19.75 %	3.08	−23.00 %
10^2	26	22.67	−12.81 %	25.58	−1.62 %
10^3	201	185.11	−7.91 %	202.61	1.80 %
10^4	1 680	1 603.11	−4.58 %	1 710.35	1.80 %
10^5	14 902	14 338.63	−3.78 %	15 069.0	1.12 %
10^6	135 124	130 893.21	−3.13 %	136 274.75	0.85 %
10^7	1 243 370	1 211 835.68	−2.54 %	1 253 639.87	0.83 %
10^8	11 587 149	11 335 684.78	−2.17 %	11 672 710.45	0.74 %
10^9	108 941 388	106 873 861.02	−1.90 %	109 666 579.94	0.67 %
10^{10}	1 031 330 156	1 013 894 469.43	−1.69 %	1 037 530 754.16	0.60 %

Table 1: Estimates $cx/\sqrt{\ln x}$ and $E(x) = c\int_2^x \frac{dt}{\sqrt{\ln t}}$ for $g(x)$.

An error of 1.69% for $x = 10^{10}$ is not impressive, but then the simple form $\pi(x) \sim x/\ln x$ of the Prime Number Theorem, which can also be viewed as an instance of Raikov's Theorem, has an error of about 5% here. Motivated by the much greater accuracy of $\mathrm{Li}(x) = \int_2^x (\ln t)^{-1} dt$ as an estimate for $\pi(x)$, we considered whether

$$E(x) := c\int_2^x \frac{dt}{\sqrt{\ln t}}$$

might give a better estimate for $g(x)$. The evidence is given in the fifth and sixth columns of Table 1.

More generally, the estimate (6) in Raikov's Theorem can be restated as

$$\sum_{n\le x} a_n \sim c\int_2^x \frac{dt}{(\ln t)^{\alpha}},$$

where c and α are as defined before, since the two estimates are asymptotically equivalent. Our experience, together with heuristic arguments based on the expected values of certain random variables, suggests that this integral form will usually give more accurate approximations.

3. Two Digressions

3.1. An Alternative Evaluation of c

For an alternative approach to evaluating c, we will use the Dedekind zeta function $\zeta_K(s)$ of the field $K=\mathbb{Q}(i)$, defined (for any algebraic number field K) by

$$\zeta_K(s) := \sum_I \frac{1}{N(I)^s} = \prod_P \left(1-\frac{1}{N(P)^s}\right)^{-1} = \prod_P \frac{1}{1-N(P)^{-s}}, \tag{19}$$

where I and P range over all ideals and prime ideals of $\mathcal{O}_K$, and $N(\cdot)$ denotes their norm. We claim that if $K=\mathbb{Q}(i)$ then

$$\zeta_K(s) = \zeta(s)L(s). \tag{20}$$

To see this, when $K=\mathbb{Q}(i)$, so that $\mathcal{O}_K=\mathbb{Z}[i]$, the prime $q=2$ ramifies, primes $q\equiv 1 \pmod 4$ split into two prime ideals $(a\pm ib)$ with norm $q=a^2+b^2$, and primes $q\equiv -1 \pmod 4$ are inert, each giving one prime ideal with norm q^2, so that for $\mathrm{Re}(s)>1$ we have

$$\begin{aligned}
\zeta_K(s) &= \frac{1}{1-2^{-s}}\cdot\prod_{q\equiv 1(4)}\frac{1}{(1-q^{-s})^2}\cdot\prod_{q\equiv -1(4)}\frac{1}{1-q^{-2s}}\\
&= \left(\frac{1}{1-2^{-s}}\prod_{q\equiv 1(4)}\frac{1}{1-q^{-s}}\prod_{q\equiv -1(4)}\frac{1}{1-q^{-s}}\right)\\
&\times \left(\prod_{q\equiv 1(4)}\frac{1}{1-q^{-s}}\prod_{q\equiv -1(4)}\frac{1}{1+q^{-s}}\right) = \zeta(s)L(s).
\end{aligned}$$

If K is any algebraic number field then $\zeta_K(s)$ converges absolutely for $\mathrm{Re}(s)>1$, and has a meromorphic extension to $\mathbb{C}$ with a unique pole at $s=1$. This pole is simple, with residue $\rho_K = \lim_{s\to 1}(s-1)\zeta_K(s)$ given by the class number formula (see [4, Ch. 5, §1.1] or [11, Ch IV, Theorem 2.12], for example)

$$\rho_K = \lim_{s\to 1}(s-1)\zeta_K(s) = \frac{2^{r_1}(2\pi)^{r_2}Rh}{w\sqrt{|D|}},$$

where r_1 is the number of real embeddings of K in $\mathbb{C}$, $2r_2$ is the number of complex embeddings, R is the regulator, h is the class number, w is the number of roots of 1 in K, and D is the discriminant. Thus $\zeta = \zeta_{\mathbb{Q}}$ has residue $\rho_{\mathbb{Q}} = 1$ at $s = 1$, whereas when $K = \mathbb{Q}(i)$ we have $r_1 = 0$, $r_2 = 1$ (the identity and complex conjugation), $R = 1$ (since $r_1 = 0$), $h = 1$ (all ideals in $\mathcal{O}_K = \mathbb{Z}[i]$ are principal), $w = 4$ (the roots of 1 are the powers of i) and $D = -4$, so that $\rho_K = \pi/4$. (There is also a geometric proof of this, counting lattice points $a + ib \in \mathbb{Z}[i]$ in the first quadrant of disks $a^2 + b^2 \le N$: see [17, Example 10.1.7 and Exercise 10.1.8] for details, or [22] for an outline.)

Since ζ and ζ_K have meromorphic continuations in a neighborhood of 1, it follows that the function $L = \zeta_K/\zeta$ is meromorphic there, with

$$L(1) = \lim_{s\to 1}\frac{\zeta_K(s)}{\zeta(s)} = \frac{\rho_K}{\rho_{\mathbb{Q}}} = \frac{\pi}{4}. \tag{21}$$

If we multiply (11) by $s - 1$ and then take limits as $s \to 1$ we see that

$$H(s)^2 = (s-1)F(s)^2 \to \frac{2F(2)}{\pi}$$

since $(s-1)\zeta(s) \to 1$ and $L(s) \to L(1) = \pi/4$. Thus in Raikov's Theorem we have $\alpha = 1/2$ and $H(1) = \sqrt{2F(2)/\pi}$, so

$$c = \frac{H(1)}{\Gamma(\frac{1}{2})} = \frac{1}{\pi}\sqrt{2F(2)}$$

where

$$F(2) = \prod_{q\equiv -1(4)}\left(1 - \frac{1}{q^2}\right)^{-1} = \sum_{n=1}^{\infty}\frac{g_n}{n^2} = 1 + \frac{1}{3^2} + \frac{1}{7^2} + \cdots .$$

The partial sums $\sum_{n=1}^{N} g_n/n^2$ of this series converge to $F(2)$ from below as $N \to \infty$. To obtain a sequence converging to $F(2)$ from above we use

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = \zeta(2) = \frac{\pi^2}{6}$$

(see [7, Exercise 9.7], for example), so

$$\sum_{\text{odd}\, n}\frac{1}{n^2} = \zeta(2) - \frac{1}{2^2}\zeta(2) = \frac{\pi^2}{8}$$

and hence

$$F(2) = \frac{\pi^2}{8} - \sum_{n=1}^{\infty} \frac{h_n}{n^2} = \frac{\pi^2}{8} - \left(\frac{1}{5^2} + \frac{1}{13^3} + \cdots \right)$$

where $h_n = 1 - g_n$ or 0 as n is odd or even, so that the sum is over all odd $n \notin \mathcal{P}^*$. Thus

$$L := \sum_{n=1}^{N} \frac{g_n}{n^2} < F(2) < \frac{\pi^2}{8} - \sum_{n=1}^{N} \frac{h_n}{n^2} \ := U \tag{22}$$

for all N, with both bounds converging monotonically to $F(2)$ as $N \to \infty$, so one can evaluate $F(2)$ to any desired accuracy by taking N sufficiently large.

For example, taking $N = 10^9$ in (22) and evaluating terms to 30 decimal places gives upper and lower bounds

$$U = 1.168075585806680823515746 08710,$$

$$L = 1.168075585306680823515746 05022,$$

with $U - L = 5.0000000000000003688 \times 10^{-10}$. For the constant

$$c = \frac{1}{\pi}\sqrt{2F(2)},$$

this gives upper and lower bounds

$$U' = 0.486519888468395460325765388132,$$

$$L' = 0.486519888364266948773906811661,$$

with $U' - L' = 1.04128511551858576471 \times 10^{-10}$. The value

$$c = 0.48651988838589099712724564058682340553 82\ldots$$

given in (18) lies between these two bounds, and they agree with it in their first nine significant figures.

Remark 1. The values of L and U in (22) for any given N depend on partial sums of the Dirichlet series for the sequences (g_n) and (h_n), representing odd integers n which are or are not $\mathcal{P}$-integers. For small N the former predominate, so that the lower bounds L are closer to $F(2)$ than the upper bounds U are; since both converge monotonically to $F(2)$, this means that U decreases faster than L increases, especially as non-$\mathcal{P}$-integers eventually start to predominate as N increases.

To illustrate this, let us denote by L_k and U_k the lower and upper bounds for $F(2)$ computed over the segment $n \leq N = 10^k$, and let R_k denote the ratio between the values of their rate of change, that is,

$$R_k = \frac{L_k - L_{k-1}}{U_{k-1} - U_k}.$$

For comparison, let r_k denote the ratio $g(x)/h(x)$ of $\mathcal{P}$-integers to non-$\mathcal{P}$-integers among the odd integers $n \le x = 10^k$. These ratios R_k and r_k are shown, for $k = 1, \ldots, 9$, in Table 2. It follows from Theorem 2 that R_k and r_k converge to 0 as $k \to \infty$, and Table 2 shows that they do so "in step".

k	R_k	r_k
1	3.5966238347	4.0000000000
2	0.9856093680	1.0833333333
3	0.6579817478	0.6722408026
4	0.5213884648	0.5060240963
5	0.4445379068	0.4245825972
6	0.3809371214	0.3703285499
7	0.3382979524	0.3309801604
8	0.3076691054	0.3016477220
9	0.2834484269	0.2785807156

Table 2: Comparison of ratios R_k and r_k.

3.2. Bias

Motivated by curiosity rather than problems involving Riemann surfaces, we note that if one defines

$$\mathcal{P}^+ := \{q \mid q \text{ is prime and } q \equiv 1 \pmod 4\}$$

then, by essentially the same argument as that used in Section 2 for $\mathcal{P}$, the corresponding counting function $g^+(x)$ for $\mathcal{P}^+$-integers $n \le x$ is estimated by a similar formula

$$g^+(x) \sim \frac{c^+ x}{\sqrt{\ln x}} = \sqrt{\frac{e^\gamma}{\pi}} \cdot \frac{C^+}{2} \cdot \frac{x}{\sqrt{\ln x}} \quad \text{as} \quad x \to \infty,$$

where

$$C^+ := C(4, -1) = 0.86892776823432382990915277910465291229 39\ldots$$

according to [14, 15], so that

$$c^+ = 0.32712936694102638240023286902839279969 51\ldots\ .$$

Since $c/c^+ = C(4,1)/C(4,-1) = 1.4872400265\ldots$, this shows that $\mathcal{P}$-integers appear nearly 50% more frequently than $\mathcal{P}^+$-integers. This is despite the fact that the numbers $\pi^\pm(x)$ of primes $q \le x$ satisfying $q \equiv \pm 1$ (mod 4) satisfy

$$\pi^+(x) \sim \pi^-(x) \sim \pi(x)/2 \quad \text{as} \quad x \to \infty$$

by de la Vallée-Poussin's quantified form of Dirichlet's theorem on primes in an arithmetic progression (see [4, Ch. 5, §3.2], for example). The reason for this surprisingly large bias is that the smallest primes $q \equiv 1 \pmod 4$, such as $5, 13, 17, 29, \ldots$, are almost always larger than the corresponding primes congruent to $-1 \pmod 4$, such as $3, 7, 11, 19, \ldots$ (see [8] for a very readable account of this and other similar phenomena), so that $\mathcal{P}^+$-integers tend to be larger and hence less frequently found than $\mathcal{P}$-integers. For example, the first ten $\mathcal{P}$-integers are 3, 7, 9, 11, 19, 21, 23, 27, 31, 33, while the corresponding $\mathcal{P}^+$-integers are 5, 13, 17, 25, 29, 37, 41, 53, 61, 65.

Table 3 gives the numbers of $\mathcal{P}$- and $\mathcal{P}^+$-integers $n \leq x$ for various x. It looks plausible that the ratios in the last column of the table tend to $c/c^+ = 1.4872400265\ldots$.

x	$g(x)$	$g^+(x)$	$g(x)/g^+(x)$
10	4	2	2.0000000000
10^2	26	15	1.7333333333
10^3	201	123	1.6341463414
10^4	1 680	1 074	1.5642458100
10^5	14 902	9 623	1.5485815234
10^6	135 124	87 882	1.5375618305
10^7	1 243 370	814 183	1.5271382477
10^8	11 587 149	7 618 317	1.5209591567
10^9	108 941 388	71 838 469	1.5164770284

Table 3: Numbers $g(x)$ and $g^+(x)$ of $\mathcal{P}$- and $\mathcal{P}^+$-integers $n \leq x$.

It follows from a theorem of Iwaniec [10] that for each integer $A \neq 0$ the number of primes $p \leq x$ of the form $\xi^2 + \eta^2 + A$ with $\gcd(\xi, \eta) = 1$ has order of magnitude $x/(\ln x)^{3/2}$. Now, a positive integer n is properly represented by the quadratic form $\xi^2 + \eta^2$, that is, with $\gcd(\xi, \eta) = 1$, if and only if n is not divisible by 4 or by any prime $q \equiv -1 \pmod 4$, or equivalently, $n = m$ or $2m$ where m is a $\mathcal{P}^+$-integer. Taking $A = -2$ we see that there are infinitely many primes p such that $p + 2$ is a $\mathcal{P}^+$-integer. Unfortunately, there is no quadratic form which plays a similar role for $\mathcal{P}$-integers.

4. The Twin Prime Conjecture

It is useful now to recall the heuristic arguments used by Hardy and Littlewood to justify their work in [9] on, among various other problems, the Twin Prime Conjecture, which asserts that there are infinitely many pairs of twin primes $p, p+2$.

By the Prime Number Theorem, the probability of a natural number $p \leq x$

being prime is asymptotic to $1/\ln x$, so the numbers of primes and of pairs of primes $p, p' \leq x$ are asymptotically equivalent to

$$\mathrm{Li}(x) := \int_2^x \frac{dt}{\ln t} \quad \text{and} \quad I_2(x) := \int_2^x \frac{dt}{(\ln t)^2}.$$

However, if p and p' are related by an equation such as $p' = p + 2$, rather than chosen independently, the estimate $I_2(x)$ will be incorrect. Since natural numbers are uniquely determined by their residues (mod q) for all primes q, it makes sense to apply the restriction $p' = p + 2$ through its effect on these residues. For each prime q, we need to consider the probabilities that p and p', chosen independently or with $p' = p+2$, being both prime, are both coprime to q (we neglect the vanishingly small probability that either of them is equal to q). By a simple count of congruence classes (mod q), we see that these probabilities are, respectively,

$$\left(1 - \frac{1}{q}\right)^2 \quad \text{and} \quad 1 - \frac{\omega_f(q)}{q},$$

where $\omega_f(q)$ is the number of roots of the polynomial $f(t) = t(t + 2)$ (mod q). In order to replace the first probability, implicit in the estimate $I_2(x)$, with the second more correct probability, we therefore multiply $I_2(x)$ by the correction factor

$$C_2(q) = \left(1 - \frac{1}{q}\right)^{-2} \left(1 - \frac{\omega_f(q)}{q}\right).$$

Doing this for each prime q, we multiply $I_2(x)$ by a correction term, called a *Hardy–Littlewood constant*,

$$C(f) = \prod_{q \text{ prime}} C_2(q) = \prod_{q \text{ prime}} \left(1 - \frac{1}{q}\right)^{-2} \left(1 - \frac{\omega_f(q)}{q}\right).$$

Clearly $\omega_f(2) = 1$ while $\omega_f(q) = 2$ for each prime $q > 2$, so we have an estimate for the number $\pi_2(x)$ of twin prime pairs $p, p + 2 \leq x$ of the form

$$E_2(x) = C(f)I_2(x) = 2C_2 I_2(x) = 2C_2 \int_2^x \frac{dt}{(\ln t)^2}\,, \tag{23}$$

where the initial factor 2 is the correction factor $C_2(2)$ corresponding to the prime $q = 2$, and

$$C_2 := \prod_{q>2} \left(1 - \frac{1}{q}\right)^{-2} \left(1 - \frac{2}{q}\right) = \prod_{q>2} \left(1 - \frac{1}{(q-1)^2}\right) = 0.6601618158\ldots, \tag{24}$$

with the product over all odd primes q.

The conjecture is that

$$\pi_2(x) \sim E_2(x) \quad \text{as} \quad x \to \infty.$$

Although it is unproved, there is strong evidence for this conjecture. For example, it is known that

$$\pi_2(10^{18}) = 808\,675\,888\,577\,436,$$

while Maple evaluates

$$E_2(10^{18}) = 808\,675\,901\,493\,606.3\ldots.$$

The relative error is 0.0000016%.

In 1962 Bateman and Horn [3] made a wide-ranging generalization of the Hardy–Littlewood estimate, giving a similar conjectured estimate $E(x)$ for the number $Q(x)$ of natural numbers $t \le x$ at which a given finite set of polynomials $f_1(t), \ldots, f_k(t) \in \mathbb{Z}[t]$ simultaneously take prime values. It assumes that these polynomials satisfy the following conditions, which are obviously necessary for there to be infinitely many such t (Schinzel's Hypothesis [21] asserts that they are also sufficient):

1. each $f_i(t)$ has a positive leading coefficient;
2. each $f_i(t)$ is irreducible in $\mathbb{Z}[t]$;
3. the product $f(t) := f_1(t) \ldots f_k(t)$ is not identically zero modulo any prime.

The Bateman–Horn Conjecture (BHC) asserts that

$$Q(x) \sim E(x) := C\int_a^x \frac{dt}{\ln f_1(t) \ldots \ln f_k(t)} \quad \text{as} \quad x \to \infty, \tag{25}$$

where C is the *Hardy–Littlewood constant*

$$C = C(f_1, \ldots, f_k) = \prod_{q \text{ prime}} \left(1 - \frac{1}{q}\right)^{-k} \left(1 - \frac{\omega_f(q)}{q}\right), \tag{26}$$

$\omega_f(q)$ is the number of roots of $f \pmod{q}$, and the lower limit a in the integral in (25) is chosen so that the integral avoids singularities where $\ln f_i(t) = 0$ for some i. If conditions (1) to (3) are satisfied then the infinite product in (26) converges to a limit $C > 0$ [1, §5], while the definite integral in (25) diverges to $+\infty$ with x, so if the BHC is true then $Q(x) \to +\infty$ and Schinzel's Hypothesis, that the polynomials f_i simultaneously take prime values for infinitely many $t \in \mathbb{N}$, is verified.

This includes the cases of the twin primes and the Sophie Germain primes conjectures, where $f_1(t) = t$ and $f_2(t) = t + 2$ or $2t + 1$, respectively. The BHC has been proved only in the case of a single polynomial of degree 1: this is the quantified version, due to de la Vallée Poussin, of Dirichlet's Theorem on primes in an arithmetic progression $at+b$. Nevertheless, the estimates produced by the BHC, in a wide range of applications, agree remarkably well with observed counts obtained by primality-testing. As an example, with $f_1(t) = 4t+1$ and $f_2(t) = 4t+3$, Table 4 shows the BHC estimates $E_2^+(x)$ for the numbers $\pi_2^+(x)$ of twin primes $p, p+2 \leq x$ with $p \equiv 1 \pmod 4$, those yielding BCC primes p.

x	$\pi_2^+(x)$	$E_2^+(x)$	error
10	1	1.148985018	14.8985 %
10^2	4	5.498634634	37.4659 %
10^3	19	21.62864106	13.8350 %
10^4	105	105.8363607	0.79653 %
10^5	604	623.0852586	3.15981 %
10^6	4 046	4 122.745734	1.89683 %
10^7	29 482	29 375.63915	−0.3608 %
10^8	220 419	220 182.6280	−0.1072 %
10^9	1 712 731	1 712 652.809	−0.0046 %
10^{10}	13 706 592	13 705 706.99	−0.0065 %

Table 4: Numbers of pairs of twin primes $p, p+2 \leq x$ with $p \equiv 1 \pmod 4$.

5. In the Footsteps of Hardy and Littlewood

We will now adapt the heuristic arguments used by Hardy and Littlewood, and later by Bateman and Horn, to the slightly different context of our BCC primes problem.

Using the Prime Number Theorem and Theorem 2 to give the distributions of prime numbers p and of $\mathcal{P}$-integers n, we first consider

$$I(x) := \int_2^x \frac{c\,dt}{(\ln t)^{3/2}}$$

as an estimate for the number $a(x)$ of BCC primes $p \leq x$, where c is as in (18). Of course, this treats p and n as independent random variables, and takes no account of the fact that $n = p+2$. In following the example of Hardy and Littlewood, with a similar heuristic justification, we now apply correction factors only for primes $q \notin \mathcal{P}$, that is, $q = 2$ or $q \equiv 1 \pmod 4$, since it is for such q that we need to replace the probability that independent variables p and n are both coprime to q with the corresponding probability for p and $p+2$. We therefore multiply this estimate $I(x)$

by $2C_2^+$ where

$$C_2^+ := \prod_{q\equiv 1(4)}\left(1-\frac{1}{q}\right)^{-2}\left(1-\frac{2}{q}\right) = \prod_{q\equiv 1(4)}\left(1-\frac{1}{(q-1)^2}\right), \qquad (27)$$

with the extra factor 2 corresponding to the prime $q=2$.

This leads to the following conjecture.

Conjecture 2. The function $a(x)$ which counts BCC primes $p \le x$ satisfies

$$a(x) \sim 2C_2^+ I(x) = c'\int_2^x \frac{dt}{(\ln t)^{3/2}} \quad \text{as} \quad x \to \infty, \qquad (28)$$

where

$$c' = 2C_2^+ c = \prod_{q\equiv 1\,(4)}\left(1-\frac{1}{(q-1)^2}\right)\cdot\sqrt{\frac{e^\gamma}{\pi}}\cdot C(4,1). \qquad (29)$$

6. Computations

The various estimates we have discussed can be tested computationally by using Maple. The definite integrals, such as those appearing in (23) and (28), are evaluated accurately and rapidly, even for large values of x, by numerical integration. The Hardy–Littlewood constants, such as C_2 in (24) and C_2^+ in (27), are defined as infinite products which converge slowly; good approximations can be found by taking partial products over large initial segments of the relevant primes, but on a laptop these calculations can take a matter of hours. (An alternative method of approximating these constants is discussed at the end of this section.) Having evaluated the various estimates, one can compare them with the actual numbers of terms being counted by using the primality test within Maple.

Using Maple to evaluate (27) for primes $q \le 10^{10}$, we found that

$$C_2^+ \approx 0.923061132219503810946175811877\ldots\ ;$$

the computation took 7 hours and 53 minutes with an Intel i7 processor. The values of $\sqrt{e^\gamma/\pi}$ and $C(4,1)$ given in Section 2 allow us to deduce that

$$c' \approx 0.8981751984\ldots \qquad (30)$$

To test the value for c' in (30) let us define $C(x)$ by

$$a(x) = C(x)\int_2^x (\ln t)^{-3/2}dt,$$

so we expect that $C(x) \to c'$ as $x \to \infty$. Using Maple to evaluate $a(x)$ and $\int_2^x (\ln t)^{-3/2} dt$, we found the values of $C(x)$ shown in Tables 5 and 6. We note that, after some initial instability due to the relatively small numbers of BCC primes appearing, the values of $C(x)$ decrease almost monotonically towards c' from $x = 10^7$ to $x = 35 \cdot 10^9$.

x	$a(x)$	$C(x)$
10^1	2	0.4688555840
10^2	12	0.7153107401
10^3	65	0.8472363117
10^4	388	0.8706477077
10^5	2 708	0.9004062930
10^6	19 969	0.9024565742
10^7	155 369	0.9040719795
10^8	1 250 182	0.9023589943
10^9	10 345 920	0.9011815839
10^{10}	87 545 946	0.9010054301

Table 5: Values of the coefficients $C(x)$ for $x = 10^k$, $k = 1, \ldots, 10$.

Beyond this point, in view of the modest computing facilities available, we considered segments $[x, y] \subset \mathbb{R}$ of length $y - x = 10^8$ or 10^7, starting at values $x = 10^{15}, 10^{20}, \ldots, 10^{50}$. The results are shown in Table 7, where $a(x, y) = a(y) - a(x)$ is the number of BCC primes $p \in [x, y]$, and $C(x, y)$ is defined by the equation

$$a(x, y) = C(x, y) \cdot \int_x^y \frac{dt}{(\ln t)^{3/2}}.$$

Initially we see a further monotonic decrease towards c', but then $C(x, y) < c'$ when $x = 10^{30}$ and $y = 10^{30} + 10^8$. We are not very concerned about this: for instance, Littlewood [16] famously showed that the error $\mathrm{Li}(x) - \pi(x)$ in the Prime Number Theorem changes sign infinitely many times (see also [19] for a similar phenomenon related to Mertens's Third Theorem), so why not here? The instability in the last four rows is perhaps due to the relatively small numbers of BCC primes appearing in these shorter intervals: for instance, compare the values of $a(x, y)$ with that of $a(x)$ for the equivalent interval $[0, 10^7]$.

Here we emphasize an important distinction. The estimate $g(x) \sim cx/(\ln x)^{1/2}$ in Theorem 2, including the evaluation of c in (18), is proved, as a consequence of Raikov's Theorem. The estimate $a(x) \sim c' \int_2^x (\ln t)^{-3/2} dt$ in Conjecture 2, including the formula for c' given there, is just a conjecture, and we do not expect to see a proof for it soon. Nevertheless, we feel that the computational data presented in this section give plausible evidence for the validity of this estimate, and hence for the infinitude of BCC primes.

x	$a(x)$	$C(x)$
$11 \cdot 10^9$	95 675 252	0.9010083668
$12 \cdot 10^9$	103 758 501	0.9010266288
$13 \cdot 10^9$	111 794 166	0.9010090565
$14 \cdot 10^9$	119 795 477	0.9010345605
$15 \cdot 10^9$	127 757 388	0.9010373485
$16 \cdot 10^9$	135 683 004	0.9010241227
$17 \cdot 10^9$	143 578 133	0.9010192899
$18 \cdot 10^9$	151 444 525	0.9010197264
$19 \cdot 10^9$	159 283 669	0.9010228308
$20 \cdot 10^9$	167 092 278	0.9010018195
$21 \cdot 10^9$	174 878 590	0.9009949342
$22 \cdot 10^9$	182 641 563	0.9009888293
$23 \cdot 10^9$	190 382 424	0.9009836658
$24 \cdot 10^9$	198 100 186	0.9009699991
$25 \cdot 10^9$	205 796 174	0.9009503295
$26 \cdot 10^9$	213 474 480	0.9009388161
$27 \cdot 10^9$	221 136 287	0.9009360946
$28 \cdot 10^9$	228 778 913	0.9009277506
$29 \cdot 10^9$	236 403 538	0.9009161678
$30 \cdot 10^9$	244 014 302	0.9009145456
$31 \cdot 10^9$	251 607 900	0.9009079931
$32 \cdot 10^9$	259 184 139	0.9008943241
$33 \cdot 10^9$	266 743 992	0.9008756612
$34 \cdot 10^9$	274 293 778	0.9008715586
$35 \cdot 10^9$	281 827 468	0.9008601041

Table 6: Values of the coefficients $C(x)$ for $x = m \cdot 10^9$, $m = 11, \ldots, 35$.

x	$y - x$	$a(x, y)$	$C(x, y)$
10^{15}	10^8	442 649	0.8985037831
10^{20}	10^8	287 429	0.8982538979
10^{25}	10^8	205 949	0.8994835673
10^{30}	10^8	156 398	0.8979178604
10^{35}	10^7	12 389	0.8963174519
10^{40}	10^7	10 299	0.9103503807
10^{45}	10^7	8 554	0.9022181293
10^{50}	10^7	7 507	0.9273527313

Table 7: Estimation over segments $[x, y]$.

An anonymous referee has suggested an alternative method of approximating the infinite product (27) by taking logarithms, splitting the resulting infinite series into those terms involving primes $q \leq x$ and the rest, and then applying Equation (4.14) in [20] to the latter. Using this method with $x = 10^{10}$, and working to 40 decimal places, we obtained upper and lower bounds

$$U = 0.92306113222237775452345671972710771393934,$$

$$L = 0.9230611322152300766578062881426292862307$$

for C_2^+, which confirm the first ten decimal places of the value obtained above. This computation took 12 hours and 13 minutes.

7. A Similar Problem

In the construction of Riemann surfaces used by Bujalance, Cirre, and Conder in [5] and considered here, it is important for group-theoretic reasons that the prime p and the $\mathcal{P}$-integer n should differ by 2. However, it is clear that in the arguments we have used one could generalize this relationship, and still obtain similar estimates for the distribution of such primes. For instance, one could replace the difference 2 here with any non-zero even integer. We now give a less trivial example, also arising from [5].

Having obtained an Accola–Maclachlan-type lower bound $M(g) \geq 4(g-1)$ for the largest possible order $M(g)$ of the automorphism group of a pseudo-real Riemann surface of odd genus $g \geq 3$, the authors of [5] have presented a set $\mathcal{G}$ of genera g for which this bound is sharp. Note that the sharpness of the corresponding bound $M(g) \geq 2g$ for even genera $g \geq 2$ is left as an open problem. For this they give a similar construction (see the remark following their Theorem 5.2) of a family of pseudo-real Riemann surfaces of genus $g = 2p + 1$ where p is what we will call a BCC2 prime, defined as follows.

Definition 2. A *BCC2 prime* is a prime number p such that

$$p \equiv 3 \pmod 8), \; p \equiv 2 \text{ or } 5 \pmod 9) \text{ and } p \not\equiv 5 \pmod 7), \tag{31}$$

and

$$\begin{gathered} n := p + 1 \text{ is not divisible by } 11, 23, 47 \text{ or by} \\ \text{any prime } q \equiv 1 \pmod 3) \text{ or } q \equiv 1 \pmod 4). \end{gathered} \tag{32}$$

This raises the following problem, which appears to be even more challenging than Problem 1.

Problem 2. Are there infinitely many BCC2 primes?

These primes are much rarer than the BCC primes considered earlier: for example, compare Table 8, which shows the first sixteen of them, with the corresponding list of BCC primes in (1).

p	factors of $n = p + 1 = 12m$	$m \pmod 7$	$r = m \pmod{84}$
11	$2^2 \cdot 3$		
1283	$2^2 \cdot 3 \cdot 107$	2	23
1571	$2^2 \cdot 3 \cdot 131$	5	47
2003	$2^2 \cdot 3 \cdot 167$	6	83
3011	$2^2 \cdot 3 \cdot 251$	6	83
7043	$2^2 \cdot 3 \cdot 587$	6	83
7907	$2^2 \cdot 3 \cdot 659$	1	71
8627	$2^2 \cdot 3 \cdot 719$	5	47
9923	$2^2 \cdot 3 \cdot 827$	1	71
10 067	$2^2 \cdot 3 \cdot 839$	6	83
15 107	$2^2 \cdot 3 \cdot 1259$	6	83
15 683	$2^2 \cdot 3 \cdot 1307$	6	83
17 123	$2^2 \cdot 3 \cdot 1427$	6	83
17 987	$2^2 \cdot 3 \cdot 1499$	1	71
18 131	$2^2 \cdot 3 \cdot 1511$	6	83
19 427	$2^2 \cdot 3 \cdot 1619$	2	23

Table 8: The first sixteen BCC2 primes p.

First let us restate the definition of BCC2 primes in terms of n. The congruences (mod 8) and (mod 9) in condition (31) are equivalent to $n = p + 1 \equiv 4 \pmod 8$ and $n \equiv 3$ or $6 \pmod 9$, that is, $n = 12m$ for some m coprime to 12; when this is satisfied the condition $p \not\equiv 5 \pmod 7$ in (31) is equivalent to $m \not\equiv 4 \pmod 7$. Thus condition (31) is equivalent to

$$n = 12m \text{ where } (12, m) = 1 \text{ and } m \not\equiv 4 \,(\mathrm{mod}\, 7). \tag{33}$$

Similarly, when this is satisfied, condition (32) is equivalent to

$$\text{each prime } q \text{ dividing } m \text{ satisfies } 47 < q \equiv -1 \,(\mathrm{mod}\ 12). \tag{34}$$

Thus conditions (31) and (32), taken together, are equivalent to (33) and (34), also taken together. Recall that in addition we require $n - 1$ $(= p)$ to be prime (and greater than 7), which implies that $m \not\equiv 3 \pmod 7$ since otherwise we would have $12m - 1 \equiv 36 - 1 \equiv 0 \pmod 7$, while conditions (33) and (34) imply that $m \not\equiv 4$ and $m \not\equiv 0 \pmod 7$, so we have

$$m \equiv 1, 2, 5 \text{ or } 6 \pmod 7. \tag{35}$$

In Table 8, apart from $p = 11$, all entries have the property that $n = 12m$ for some prime m. Indeed, the first exception to this is $p = 41\,771$, with $n = 12 \cdot 59^2$. Moreover, among the $172\,515$ BCC2 primes $p \leq 10^9$, there are only $41\,711$ for which m is not prime. Motivated by this, instead of attempting to estimate the distribution of the whole set of BCC2 primes, as we did in the case of the BCC primes, we will apply the Bateman–Horn Conjecture (see Section 4) to give strong evidence that there are infinitely many of them for which m is prime. This is analogous to the argument at the end of Section 4, where we considered those BCC primes p which are members of twin primes $p, p+2$.

Let us therefore assume that $n = 12m$ for some prime $m > 47$ such that $m \equiv -1 \pmod{12}$ and $m \equiv 1, 2, 5$ or $6 \pmod 7$, or equivalently $m \equiv r \pmod{84}$ where $r = 71, 23, 47$ or 83 by the Chinese Remainder Theorem. Clearly, any such choice of m satisfies conditions (33), (34) and (35). For each such r, we are therefore looking for integers $t \geq 0$ such that the polynomials

$$f_1(t) = 84t + r$$

and

$$f_2(t) = 12f_1(t) - 1 = 1008t + 12r - 1$$

both take prime values, namely m and p. These two polynomials are irreducible, with positive leading coefficients, and their product f is not identically zero modulo any prime, so we have four instances of the Bateman–Horn Conjecture, one for each value of r. In each case the BHC gives an asymptotic estimate $E_r(x)$ for the number $Q_r(x)$ of integers $t \leq x$ such that $f_1(t)$ and $f_2(t)$ are both prime. We have $\omega_f(q) = 0$ for the primes $q = 2, 3$ and 7, and $\omega_f(q) = 2$ for all other primes, so the Hardy–Littlewood constant $C = C(f_1, f_2)$, which is independent of r, is positive. In fact, comparing the infinite products for C and for $C_{\mathrm{twins}} := 2C_2$ (see Section 4), which differ at only these three primes, shows that

$$C = \frac{42}{5} \cdot C_{\mathrm{twins}} = \frac{42}{5} \cdot 1.3203236316\ldots = 11.0907185062\ldots.$$

It follows that $E_r(x) \to +\infty$ as $x \to \infty$, giving evidence that for each r there are infinitely many prime values m and p for f_1 and f_2, and thus infinitely many BCC2 primes p.

To test this approach we found that of the $172\,515$ BCC2 primes $p \leq 10^9$, the number with m prime, as we have been assuming, is $Q = 130\,804$. The four BHC estimates

$$E_r(x) = C \cdot \int_2^x \frac{dt}{\ln f_1(t) \ln f_2(t)}$$

were computed by taking $x = 992\,062$; beyond this point the primes $p = 1008t + 12r - 1$ begin to be greater than 10^9. The four values obtained were almost identical, ranging from $32\,709.13$ to $32\,709.29$. Their sum gave an estimate of $E = 130\,836.81$ for Q, with an error of 0.025%.

After we submitted this paper, another similar problem, arising from automorphism groups of maps, came to our attention. This asks for the distribution of those $n \in \mathbb{N}$ for which some $x \in \mathbb{Z}$ satisfies $x^2 + x + 1 \equiv 0 \pmod{n}$. It follows from the Chinese Remainder Theorem that the set S of such n is $\mathcal{P}^* \,\dot{\cup}\, 3\mathcal{P}^*$, where $\mathcal{P}$ is now the set of primes $q \equiv 1 \pmod 3$. Arguments similar to those used in Section 2 show that the number $\bar{g}(x)$ of $\mathcal{P}$-integers $n \le x$ satisfies $\bar{g}(x) \sim \bar{c}/\sqrt{\ln x}$ where

$$\bar{c} = \frac{2}{3}\sqrt{\frac{e^\gamma}{\pi}}\, C(3, -1) = 0.3012165545\ldots\ .$$

The number of $n \le x$ in S is $\bar{g}(x) + \bar{g}(x/3) \sim 4\bar{g}(x)/3$.

Acknowledgments. The authors are grateful to Emilio Bujalance, Javier Cirre and Marston Conder for creating the problems addressed in this paper, to Kevin Ford for some helpful advice on the number theory literature, and to an anonymous referee for generously suggesting a number of significant improvements to the paper.

References

[1] S. L. Aletheia-Zomlefer, L. Fukshansky, and S. R. Garcia, The Bateman–Horn conjecture: heuristic, history, and applications, *Expo. Math.* **38** (2020), 430–479.

[2] T. M. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, New York, Heidelberg, Berlin, 1976.

[3] P. T. Bateman and R. A. Horn, A heuristic asymptotic formula concerning the distribution of prime numbers, *Math. Comp.* **16** (1962), 220–228.

[4] Z. I. Borevich and I. R. Shafarevich, *Number Theory*, Academic Press, New York and London, 1966.

[5] E. Bujalance, F. J. Cirre, and M. D. E. Conder, Bounds on the orders of groups of automorphisms of a pseudo-real surface of given genus, *J. Lond. Math. Soc.* (2) **101** (2020), 877–906.

[6] A. C. Cojocaru and M. R. Murty, *Introduction to Sieve Methods and their Applications*, London Math. Soc. Student Texts, Cambridge University Press, 2005.

[7] G. Everest and T. Ward, *Introduction to Number Theory*, Graduate Texts in Math. **232**, Springer, 2005.

[8] A. Granville and G. Martin, Prime number races, *Amer. Math. Monthly* **113** (2006), 1–33.

[9] G. H. Hardy and J. E. Littlewood, Some problems of 'Partitio numerorum'; III: On the expression of a number as a sum of primes, *Acta Math.* **114** (1923), 215–273.

[10] H. Iwaniec, Primes of the type $\varphi(x, y) + A$, where φ is a quadratic form, *Acta Arithmetica* **21** (1972), 203–234.

[11] G. J. Janusz, *Algebraic Number Fields* (2nd ed.), Amer. Math. Soc., Providence RI, 1996.

[12] E. Landau, Lösung des Lehmer'schen Problems, *Amer. J. Math.* **31** (1909), 86–102.

[13] A. Languasco and A. Zaccagnini, On the constant in the Mertens product for arithmetic progressions, I. Identities, *Funct. Approx. Comment. Math.* **42** (2010), 17–27.

[14] A. Languasco and A. Zaccagnini, On the constant in the Mertens product for arithmetic progressions, II. Numerical values, *Math. Comp.* **78** (2009), 315–326.

[15] A. Languasco and A. Zaccagnini, Computing the Mertens and Meissel–Mertens constants for sums over arithmetic progressions, preprint, `arXiv.math:0906.2132v1`.

[16] J. E. Littlewood, Sur la distribution des nombres premiers, *C. R. Math. Acad. Sci. Paris* **158** (1914), 1869–1872.

[17] M. R. Murty and J. Esmonde, *Problems in Algebraic Number Theory* (2nd ed.), Springer Graduate Texts in Math. **190**, Springer-Verlag, New York, 2005.

[18] D. A. Raikov, Generalization of a theorem of Ikehara–Landau (Russian), *USSR Mat. Sbornik* **45** (3) (1938), 559–568.

[19] G. Robin, Sur l'ordre maximum de la fonction somme des diviseurs, *Séminaire Delange–Pisot–Poitou, Théorie des Nombres* (1981–1982)*, Progress in Mathematics* **38** (1983), 233–244.

[20] J. B. Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers, *Illinois J. Math.* **6** (1962), 64–94.

[21] A. Schinzel and W. Sierpiński, Sur certaines hypothèses concernant les nombres premiers, *Acta Arith.* **4** (2) (1958), 185–208.

[22] Wikipedia, `Class number formula`.

Appendix

Here we briefly summarize the construction of the pseudo-real Riemann surfaces in [5] which give rise to Problems 1 and 2. In the first case, for any even $g \geq 2$ let Γ be an NEC group with signature $(1; -; [2, 2, g]; \{-\})$. This means that Γ acts as a group of isometries of the hyperbolic plane $\mathbb{H}$, with canonical generators d (a glide reflection) and x_1, x_2, x_3 (elliptic elements), and defining relations

$$x_1^2 = x_2^2 = x_3^g = d^2 x_1 x_2 x_3 = 1,$$

so that the quotient-surface $\mathbb{H}/\Gamma$ is the real projective plane with three cone-points of orders $2, 2$ and g.

Now let $\theta : \Gamma \to \mathrm{C}_{2g} = \langle u \mid u^{2g} = 1\rangle$ be the surface-kernel epimorphism defined by

$$d \mapsto u,\ x_1 \mapsto u^g,\ x_2 \mapsto u^g,\ x_3 \mapsto u^{-2},$$

and let $K = \ker\theta$. Then K is contained in the orientation-preserving subgroup Γ^+ of index 2 in Γ, so $S := \mathbb{H}/K$ is a compact Riemann surface, of genus g by the Riemann–Hurwitz formula. There is a natural action of Γ/K and hence of C_{2g} on S, with the elements of the subgroup $\langle u^2\rangle \cong \mathrm{C}_g$ and its complement acting as conformal and anticonformal automorphisms of S, and with $S/\mathrm{C}_{2g} \cong \mathbb{H}/\Gamma$; the canonical double cover $S/\mathrm{C}_g \cong \mathbb{H}/\Gamma^+$ of this orbifold is the Riemann sphere with four cone-points of order 2 and two of order g.

The generator u of C_{2g} acts as an anticonformal automorphism of S of order $2g$, whereas the only involution in G (namely u^g) acts conformally. One can choose the cone-points so that Γ is a maximal NEC group, in which case S has no further automorphisms (conformal or anticonformal), and is therefore pseudo-real. A detailed argument in [5, Theorem 4.8] shows that, provided g satisfies various other conditions (namely those defining the set $\mathcal{A}$, together with $g > 30$), no pseudo-real surface of genus g can have an abelian group of conformal automorphisms with more than g elements, so that $M^+_{\mathrm{ab}}(g) = g$. For example, since $p+2$ has no prime factors $q \equiv 1 \pmod 4$, $\mathrm{Aut}\,\mathrm{C}_{p+2}$ (of order $\varphi(p+2)$) has no elements of order 4, a fact used in subcase (2c) of the proof of Theorem 4.8 to eliminate $\mathrm{C}_{g+1} = \mathrm{C}_{p+2}$.

The construction of the surfaces S giving rise to Problem 2 is similar, except that in this case Γ has signature $(1; -; [2,2,2]; \{-\})$, S has genus $g = 2p+1$ with a full automorphism group

$$\langle u, v \mid u^2 = v^{4p} = 1, v^u = v^{2p-1}\rangle \cong \mathrm{C}_{4p} \rtimes \mathrm{C}_2$$

of order $8p = 4(g-1)$, and θ is given by

$$d \mapsto uv,\ x_1 \mapsto u,\ x_2 \mapsto u,\ x_3 \mapsto v^{2p}.$$

It is shown in [5, Theorem 5.3] that if p is a prime satisfying the conditions stated in Section 7 then no pseudo-real surface of genus g has a larger automorphism group, so that $M(g) = 4(g-1)$.