Primitive Root Bias for Twin Primes

Stephan Ramon Garcia\textsuperscript{a}, Elvis Kahoro\textsuperscript{a}, and Florian Luca\textsuperscript{b,c,d}

\textsuperscript{a}Department of Mathematics, Pomona College, Claremont, CA, USA; \textsuperscript{b}School of Mathematics, University of the Witwatersrand, Wits, Johannesburg, South Africa; \textsuperscript{c}Max Planck Institute for Mathematics, Bonn, Germany; \textsuperscript{d}Department of Mathematics, Faculty of Sciences, University of Ostrava, Ostrava, Czech Republic

**ABSTRACT**

Numerical evidence suggests that for only about 2\% of pairs $p$, $p + 2$ of twin primes, $p + 2$ has more primitive roots than does $p$. If this occurs, we say that $p$ is \textit{exceptional} (there are only two exceptional pairs with $5 \leq p \leq 10,000$). Assuming the Bateman–Horn conjecture, we prove that at least 0.459\% of twin prime pairs are exceptional and at least 65.13\% are not exceptional. We also conjecture a precise formula for the proportion of exceptional twin primes.

**KEYWORDS**

prime; twin prime; primitive root; Bateman–Horn conjecture; Twin Prime Conjecture; Brun Sieve

**2010 AMS SUBJECT CLASSIFICATION**

11A07; 11A41; 11N36; 11N37

1. Introduction

Let $n$ be a positive integer. An integer coprime to $n$ is a \textit{primitive root} modulo $n$ if it generates the multiplicative group $(\mathbb{Z}/n\mathbb{Z})^\times$ of units modulo $n$. A famous result of Gauss states that $n$ possesses primitive roots if and only if $n$ is 2, 4, an odd prime power, or twice an odd prime power. If a primitive root modulo $n$ exists, then $n$ has precisely $\varphi(\varphi(n))$ of them, in which $\varphi$ denotes the Euler totient function. If $p$ is prime, then $\varphi(p) = p - 1$ and hence $p$ has exactly $\varphi(p - 1)$ primitive roots.

If $p$ and $p + 2$ are prime, then $p$ and $p + 2$ are \textit{twin primes}. The Twin Prime Conjecture asserts that there are infinitely many twin primes. While it remains unproved, recent years have seen an explosion of closely-related work [Castryck et al. 14, Zhang 14, Maynard 15]. Let $\pi_2(x)$ denote the number of primes $p$ at most $x$ for which $p + 2$ is prime. The first Hardy–Littlewood conjecture asserts that

$$\pi_2(x) \sim 2C_2 \int_2^x \frac{dt}{(\log t)^2}$$

(1.1)

in which

$$C_2 = \prod_{p \geq 3} \frac{p(p - 2)}{(p - 1)^2} = 0.660161815 \ldots$$

(1.2)

is the \textit{twin primes constant} [Hardy and Littlewood 04]. A simpler expression that is asymptotically equivalent to (1.1) is $2C_2 x/(\log x)^2$.

A casual inspection (see Table 1) suggests that if $p$ and $p + 2$ are primes and $p \geq 5$, then $p$ has at least as many primitive roots as $p + 2$; that is, $\varphi(p - 1) \geq \varphi(p + 1)$. If this occurs, then $p$ is \textit{exceptional}. The preceding inequality holds for all twin primes $p$, $p + 2$ with $5 \leq p \leq 10,000$, except for the pairs 2381, 2383 and 3851, 3853.

If $p$, $p + 2$ are primes with $p \geq 5$ and $\varphi(p - 1) < \varphi(p + 1)$, then $p$ is \textit{exceptional}. We do not regard $p = 3$ as exceptional for technical reasons. Let $\pi_e(x)$ denote the number of exceptional primes $p \leq x$; that is,

$$\pi_e(x) = \# \{ p \leq x : p \text{ and } p + 2 \text{ are prime and } \varphi(p - 1) < \varphi(p + 1) \}.$$

Computational evidence suggests that approximately 2\% of twin primes are exceptional; see Table 2. We make the following conjecture.

**Conjecture 1.** A positive proportion of the twin primes are exceptional. That is, $\lim_{x \to \infty} \pi_e(x)/\pi_2(x)$ exists and is positive.

We are able to prove Conjecture 1, if we assume the Bateman–Horn conjecture (stated below). Our main theorem is the following.

**Theorem 1.** Assume that the Bateman–Horn conjecture holds.

(a) The set of twin prime pairs $p$, $p + 2$ for which $\varphi(p - 1) < \varphi(p + 1)$ has lower density (as a subset of twin primes) at least 0.459\%.

(b) The set of twin prime pairs $p$, $p + 2$ for which $\varphi(p - 1) \geq \varphi(p + 1)$ has lower density (as a subset of twin primes) at least 65.13\%.

This article has been republished with minor changes. These changes do not impact the academic content of the article.

**CONTACT** Stephan Ramon Garcia (Stephan.Garcia@pomona.edu) Department of Mathematics, Pomona College, 610 N. College Ave., Claremont, CA 91711, USA.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uexm.

© 2017 Taylor & Francis Group, LLC
Table 1. For twin primes \( p, p + 2 \) with \( 5 \leq p \leq 2000 \), the difference \( \delta(p) = \varphi(p - 1) - \varphi(p + 1) \) is nonnegative. That is, \( p \) has at least as many primitive roots as does \( p + 2 \).

<table>
<thead>
<tr>
<th>( p )</th>
<th>( \varphi(p - 1) )</th>
<th>( \varphi(p + 1) )</th>
<th>( \delta(p) )</th>
<th>( p )</th>
<th>( \varphi(p - 1) )</th>
<th>( \varphi(p + 1) )</th>
<th>( \delta(p) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td></td>
<td>5</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>4</td>
<td></td>
<td>11</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>6</td>
<td>6</td>
<td></td>
<td>17</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>12</td>
<td>8</td>
<td></td>
<td>29</td>
<td>12</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>16</td>
<td>12</td>
<td></td>
<td>41</td>
<td>16</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>28</td>
<td>16</td>
<td></td>
<td>59</td>
<td>28</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>24</td>
<td>24</td>
<td></td>
<td>71</td>
<td>24</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>40</td>
<td>32</td>
<td></td>
<td>101</td>
<td>40</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>52</td>
<td>36</td>
<td></td>
<td>107</td>
<td>52</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>137</td>
<td>64</td>
<td>44</td>
<td></td>
<td>137</td>
<td>64</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>149</td>
<td>72</td>
<td>40</td>
<td></td>
<td>149</td>
<td>72</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>179</td>
<td>88</td>
<td>48</td>
<td></td>
<td>179</td>
<td>88</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>197</td>
<td>72</td>
<td>64</td>
<td></td>
<td>197</td>
<td>72</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>227</td>
<td>84</td>
<td>60</td>
<td></td>
<td>227</td>
<td>84</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>96</td>
<td>64</td>
<td></td>
<td>239</td>
<td>96</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>269</td>
<td>122</td>
<td>72</td>
<td></td>
<td>269</td>
<td>122</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>281</td>
<td>96</td>
<td>92</td>
<td></td>
<td>281</td>
<td>96</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>311</td>
<td>120</td>
<td>96</td>
<td></td>
<td>311</td>
<td>120</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>347</td>
<td>172</td>
<td>112</td>
<td></td>
<td>347</td>
<td>172</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>419</td>
<td>180</td>
<td>96</td>
<td></td>
<td>419</td>
<td>180</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>431</td>
<td>168</td>
<td>144</td>
<td></td>
<td>431</td>
<td>168</td>
<td>144</td>
<td></td>
</tr>
<tr>
<td>461</td>
<td>176</td>
<td>120</td>
<td></td>
<td>461</td>
<td>176</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>521</td>
<td>192</td>
<td>168</td>
<td></td>
<td>521</td>
<td>192</td>
<td>168</td>
<td></td>
</tr>
<tr>
<td>569</td>
<td>280</td>
<td>144</td>
<td></td>
<td>569</td>
<td>280</td>
<td>144</td>
<td></td>
</tr>
<tr>
<td>599</td>
<td>264</td>
<td>160</td>
<td></td>
<td>599</td>
<td>264</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td>617</td>
<td>240</td>
<td>204</td>
<td></td>
<td>617</td>
<td>240</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td>641</td>
<td>256</td>
<td>212</td>
<td></td>
<td>641</td>
<td>256</td>
<td>212</td>
<td></td>
</tr>
<tr>
<td>659</td>
<td>276</td>
<td>160</td>
<td></td>
<td>659</td>
<td>276</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td>809</td>
<td>400</td>
<td>216</td>
<td></td>
<td>809</td>
<td>400</td>
<td>216</td>
<td></td>
</tr>
</tbody>
</table>

Computations suggest that the value of the limit in Conjecture 1 is approximately 2%; see Figures 1 and 2. A value for the limiting ratio is proposed in Section 5.

It is also worth pointing out that this bias is specific to the twin primes since the set of primes \( p \) for which \( \varphi(p - 1) - \varphi(p + 1) \) is positive (respectively, negative) has density 50% as a subset of the primes [Garcia and Luca]. That is, if we remove the assumption that \( p + 2 \) is also prime, then the bias completely disappears. Although only tangentially related to the present discussion, it is worth mentioning the exciting preprint [Lemke Oliver and Soundararajan] which concerns a peculiar and unexpected bias in the primes.

2. The Bateman–Horn conjecture

The proof of Theorem 1 is deferred until Section 4. We first require a few words about the Bateman–Horn conjecture. Let \( f_1, f_2, \ldots, f_m \) be a collection of distinct irreducible polynomials with positive leading coefficients. An integer \( n \) is prime generating for this collection if each \( f_1(n), f_2(n), \ldots, f_m(n) \) is prime. Let \( P(x) \) denote the number of prime-generating integers at most \( x \) and suppose that \( f = f_1 f_2 \cdots f_m \) does not vanish identically modulo any prime. The Bateman–Horn conjecture is

\[
P(x) \sim \frac{C}{D} \int_{2}^{x} \frac{dt}{(\log t)^m},
\]

in which

\[
D = \prod_{i=1}^{m} \deg f_i \quad \text{and} \quad C = \prod_{p} \frac{1 - N_f(p)/p}{(1 - 1/p)^m},
\]

where \( N_f(p) \) is the number of solutions to \( f(n) \equiv 0 \pmod{p} \) [Bateman and Horn 62].

If \( f_1(t) = t \) and \( f_2(t) = t + 2 \), then \( f(t) = t(t + 2) \), \( N_f(2) = 1 \), and \( N_f(p) = 2 \) for \( p \geq 3 \). In this case, Bateman–Horn predicts \( (1 - 1) \), the first Hardy–Littlewood conjecture, which in turn implies the Twin Prime Conjecture.

Although weaker than the Bateman–Horn conjecture, the Brun sieve [Tenenbaum 15, Thm. 3, Section I.4.2] has the undeniable advantage of being proven. It says that there exists a constant \( B \) that depends only on \( m \) and \( D \) such that

\[
P(x) \leq \frac{BC}{D} \int_{2}^{x} \frac{dt}{(\log t)^m} = (1 + o(1)) \frac{BC}{D} \frac{x}{(\log x)^m}
\]

for sufficiently large \( x \). In particular,

\[
\pi_2(x) \leq \frac{Kx}{(\log x)^2}
\]

for some constant \( K \) and sufficiently large \( x \). The best known \( K \) in the estimate above is \( K = 4.5 \) [Wu 04].

3. An heuristic argument

We give an heuristic argument which suggests that \( \varphi(p - 1) \geq \varphi(p + 1) \) for an overwhelming proportion of twin primes \( p, p + 2 \). It also identifies specific conditions under which \( \varphi(p - 1) < \varphi(p + 1) \) might occur. This informal reasoning can be made rigorous under the assumption of the Bateman–Horn conjecture (see Section 4).

Observe that each pair of twin primes, aside from 3,5, is of the form \( 6n - 1, 6n + 1 \). Thus, if \( p, p + 2 \) are twin primes with \( p \geq 3 \), then \( 2 | (p - 1) \) and \( 6 | (p + 1) \). We use this in the following lemma to obtain an equivalent characterization of (un)exceptionality.

Lemma 2. If \( p \) and \( p + 2 \) are prime and \( p \geq 5 \), then

\[
\varphi(p - 1) \geq \varphi(p + 1) \iff \frac{\varphi(p - 1)}{p - 1} \geq \frac{\varphi(p + 1)}{p + 1}. \tag{3-1}
\]

Proof. The forward implication is straightforward arithmetic, so we focus on the reverse. If the inequality on the right-hand side of (3-1) holds, then

\[
0 \leq \frac{\varphi(p - 1)}{p - 1} + \frac{\varphi(p + 1)}{p + 1} = \frac{p}{p} \varphi(p - 1) + \frac{1}{p} \varphi(p + 1) + \frac{1}{p + 1} \varphi(p - 1) + \frac{1}{p + 1} \varphi(p + 1)
\]

\[
< \varphi(p - 1) - \varphi(p + 1) + \frac{1}{p} \varphi(p + 1)
\]

since \( 2 | (p - 1) \) and \( 6 | (p + 1) \). For the preceding to hold, the integer \( \varphi(p - 1) - \varphi(p + 1) \) must be nonnegative.\( \square \)
In light of (3.1) and the formula (in which $q$ is prime)

$$\frac{\varphi(n)}{n} = \prod_{q \mid n} \left(1 - \frac{1}{q}\right),$$

it follows that $p$ is exceptional if and only if $p + 2$ is prime and

$$\frac{1}{2} \prod_{q \mid (p-1), q \geq 5} \left(1 - \frac{1}{q}\right) < \frac{1}{3} \prod_{q \mid (p+1), q \geq 5} \left(1 - \frac{1}{q}\right) \quad (3.2)$$

because $2|(p - 1), 3 \nmid (p - 1)$ and $6|(p + 1).$ The condition (3.2) can occur for $p - 1$ divisible by only small primes. For example, if $5, 7, 11 | (p - 1),$ then $5, 7, 11 \nmid (p + 1)$ and the quantities in (3.2) become

$$\frac{24}{77} \prod_{q \mid (p-1), q \geq 13} \left(1 - \frac{1}{q}\right)$$

and

$$\frac{1}{3} \prod_{q \mid (p+1), q \geq 13} \left(1 - \frac{1}{q}\right).$$

Since

$$\frac{24}{77} \approx 0.3117 < \frac{1}{3}$$

and $2 \cdot 5 \cdot 7 \cdot 11 = 770,$ one expects (3.2) to hold occasionally if $p = 770n + 1.$ Dirichlet’s theorem on primes in arithmetic progressions ensures that $p + 2 = 770n + 3$ is prime $1/\varphi(770) = 1/240 = 0.4167\%$ of the time. Thus, we expect a small proportion of twin prime pairs to satisfy (3.2). For
Numerical evidence suggests that \( \lim_{x \to \infty} \frac{\pi_e(x)}{\pi_2(x)} \) exists and is slightly larger than 2\%. The horizontal axis denotes the number of exceptional twin prime pairs. The vertical axis represents the ratio \( \pi_e/\pi_2 \).

For example, among the first 100 exceptional pairs (see Table 2), the following values of \( p \) have the form 770\( n + 1 \):

- 3851, 20021, 26951, 47741, 50051, 52361, 70841, 87011, 98561, 117041, 165551, 167861, 197891, 225611, 237161, 241781, 274121, 278741, 301841, 315701, 322631, 345731, 354971, 357281, 361901, 371141, 410411, 424271, 438131, 440441, 470471.

This accounts for 31\% of the first 100 exceptional pairs. We now make this heuristic argument rigorous, under the assumption that the Bateman–Horn conjecture holds.

### 4. Proof of Theorem 1

Assume that the Bateman–Horn conjecture holds. We first prove statement (a) of Theorem 1. In what follows, \( p, q, r \) denote prime numbers.

**Proof of (a).** Consider twin primes \( p, p + 2 \) such that \( 5, 7, 11 | (p - 1) \). Let \( \pi'_2(x) \) be the number of such \( p \leq x \).

![Figure 1](image1.png) **Figure 1.** Numerical evidence suggests that \( \lim_{x \to \infty} \frac{\pi_e(x)}{\pi_2(x)} \) exists and is slightly larger than 2\%. The horizontal axis denotes the number of exceptional twin prime pairs. The vertical axis represents the ratio \( \pi_e/\pi_2 \).

![Figure 2](image2.png) **Figure 2.** Plots in the \( xy \)-plane of ordered pairs \( (p, \varphi(p - 1)) \) (in red) and \( (p + 2, \varphi(p + 1)) \) (in cyan) for twin primes \( p, p + 2 \). There are no exceptional pairs visible in Figure 2a; that is, \( \varphi(p - 1) \geq \varphi(p + 1) \) in each case. The exceptional pairs 2381, 2383 and 3851, 3853 are visible in Figure 2b. A smattering of exceptional pairs emerge as more twin primes are considered.
**Step 1.** Since $5 \cdot 7 \cdot 11 = 385$, the desired primes are precisely those of the form

$$n = 385k + 1 \leq x$$

such that $n + 2 = 385k + 3$ is prime.

In the Bateman–Horn conjecture, let

$$f_1(t) = 385t + 1, \quad f_2(t) = 385t + 3, \quad \text{and} \quad f = f_1 f_2.$$

Then

$$N_f(p) = \begin{cases} 1 & \text{if } p = 2, \\ 2 & \text{if } p = 3, \\ 0 & \text{if } p = 5, 7, 11, \\ 2 & \text{if } p \geq 13. \end{cases} \quad (4-1)$$

Since $p \leq x$, we must have $k \leq (x - 1)/385$. For sufficiently large $x$, the Bateman–Horn conjecture predicts that the number of such $k$ is

$$\pi'_2(x) = (1 + o(1)) \frac{(x - 1)/385}{\log((x - 1)/385)} \prod_{p \geq 2} \left( 1 - \frac{N_f(p)/p}{(1 - 1/p)^2} \right)$$

$$= (1 + o(1)) \frac{2x}{385 \log(x)^2} \prod_{p \geq 3} \left( 1 - \frac{N_f(p)/p}{(1 - 1/p)^2} \right)$$

$$= (1 + o(1)) \frac{2x}{385 \log(x)^2} \prod_{p \geq 3} \left( 1 - \frac{2/p}{(1 - 1/p)^2} \right)$$

$$\times \prod_{p=5,7,11} \left( 1 - 2/p \right)^{-1} \quad (4-2)$$

$$= (1 + o(1)) \frac{2x}{135 \log(x)^2} \prod_{p \geq 3} \left( 1 - \frac{2/p}{(1 - 1/p)^2} \right)$$

Then $N_f(p)$ is as in (4-1) except for $p = r$, in which case

$$N_f(r) = 0. \quad \text{Indeed,}$$

$$f_1(t) \equiv b_r \equiv 385k_0 + 1 \equiv -1 \pmod{r}$$

and

$$f_2(t) \equiv b_r + 2 \equiv 1 \pmod{r}$$

for all $t$. As $x \to \infty$, the Bateman–Horn conjecture predicts that the number of such $\ell$ is

$$\pi'_2, (x) = (1 + o(1)) \frac{(x - b_r)/(385r)}{\log((x - b_r)/(385r))^2}$$

$$\times \prod_{p \geq 2} \left( 1 - \frac{N_f(p)/p}{(1 - 1/p)^2} \right)$$

$$= (1 + o(1)) \frac{x}{385r \log(x)^2} \prod_{p \geq 2} \left( 1 - \frac{N_f(p)/p}{(1 - 1/p)^2} \right)$$

$$= (1 + o(1)) \frac{2x}{385r \log(x)^2} \prod_{p \geq 3} \left( 1 - \frac{2/p}{(1 - 1/p)^2} \right)$$

$$\times \prod_{p=5,7,11,r} \left( 1 - \frac{2/p}{(1 - 1/p)^2} \right) \quad (4-4)$$

**Step 2.** Fix a prime $r \geq 13$. Let $\pi'_2, (x)$ be the number of primes $p \leq x$ such that $p$, $p + 2$ are prime, $5, 7, 11 | (p - 1)$, and $r | (p + 1)$. The desired primes are precisely those of the form

$$n = 385k + 1 \leq x$$

such that $n + 2 = 385k + 3$ is prime and $r | (385k + 2)$.

In particular, $k$ must be of the form

$$k = k_0 + r\ell,$$

in which $k_0$ is the smallest positive integer with $k_0 \equiv 2(385)^{-1} \pmod{r}$. Let $b_r = 385k_0 + 1$. Then

$$n = 385r\ell + b_r \quad \text{and} \quad n + 2 = 385r\ell + (b_r + 2) \quad (4-3)$$

are both prime, $n \leq x$, and

$$\ell \leq x - b_r \quad \frac{385r}{385r}.$$

In the Bateman–Horn conjecture, let

$$f_1(t) = 385rt + b_r, \quad f_2(t) = 385rt + (b_r + 2),$$

and $f = f_1 f_2$.

**Step 3.** Suppose that $p$ is counted by $\pi'_2(x)$; that is, suppose that $p$, $p + 2$ are prime and that $5, 7, 11 | (p - 1)$. Then $6 | (p + 1), \quad 5, 7, 11 | (p + 1)$, and

$$\frac{\varphi(p - 1)}{p - 1} \leq \prod_{q=2,5,7,11} \left( 1 - \frac{1}{q} \right) = \frac{24}{77}.$$
If the pair $p$ is unexceptional, then Lemma 2 ensures that
\[
\frac{1}{3} \prod_{r|p+1 \atop r \geq 13} \left( 1 - \frac{1}{r} \right) = \frac{\varphi(p+1)}{p+1} \leq \frac{\varphi(p-1)}{p-1} \leq \frac{24}{77}.
\]

Consequently,
\[
\prod_{r|p+1 \atop r \geq 13} \left( 1 + \frac{1}{r-1} \right) \geq \frac{77}{72},
\]
in which $r$ is prime. Let
\[
F(p) = \sum_{r|p+1 \atop r \geq 13} \log \left( 1 + \frac{1}{r-1} \right).
\]

**Step 4.** We want to count the twin primes pairs $p, p+2$ with $p \leq x$, $F(p) \geq \log(77/72)$, and $5, 7, 11|(p-1)$. To do this, we sum up $F(p)$ over all twin primes $p$ counted by $\pi'_2(x)$ and change the order of summation to obtain
\[
A(x) = \sum_{p \text{ counted by } \pi'_2(x)} F(p).
\]

\[
= \sum_{r \geq 13} \pi'_{2,r}(x) \log \left( 1 + \frac{1}{r-1} \right)
\]
\[
\leq \sum_{13 \leq z \leq x} \pi'_{2,z}(x) \log \left( 1 + \frac{1}{r-1} \right)
\]
\[
+ \sum_{r \geq 13} \pi'_{2,r}(x) \log \left( 1 + \frac{1}{r-1} \right)
\]
\[
= A_1(x) + A_2(x) + A_3(x),
\]

in which $z$ is to be determined later. We bound the three sums separately.

(a) If $13 \leq r \leq z$, then (4–4) asserts that
\[
\pi'_{2,r}(x) = (1 + o(1)) \frac{\pi_2(x)}{135(r-2) - 1}
\]
uniformly for $r \in [13, z]$ as $x \to \infty$. For sufficiently large $x$ we have\footnote{Since \(\log(1+t) \leq t\) for \(t > 0\), the terms of the series are \(O(1/r^2)\) and hence it converges rapidly enough for reliable numerical evaluation. Mathematica provides the value \(0.025496772370458\ldots\).}
\[
A_1(x) \leq (1 + o(1)) \frac{\pi_2(x)}{135} \left( \sum_{13 \leq z \leq x} \frac{1}{(r-2)} \log \left( 1 + \frac{1}{r-1} \right) \right)
\]
\[
\leq (1 + o(1)) \frac{\pi_2(x)}{135} \left( \sum_{r \geq 13} \frac{1}{(r-2)} \log \left( 1 + \frac{1}{r-1} \right) \right)
\]
\[
\leq (1 + o(1)) \frac{\pi_2(x)}{135} \left( \frac{135}{2} \right) \leq \frac{0.025496772370458}{135} \pi_2(x) < 0.00018865 \pi_2(x).
\]

(b) If $z < r \leq (\log x)^3$, we use the Brun sieve and manipulations similar to those used to obtain (4–4) to find an absolute constant $K$ such that
\[
\pi'_{2,r}(x) \leq \frac{K(x/(135r))}{(\log(x/(135r)))^2}
\]
for sufficiently large $x$. Since $r \leq (\log x)^3$,
\[
\log(x/(135r)) \geq \log(x^{1/2}) \geq (\log x)/2
\]
holds if $x \geq 10^{14}$. Then (1–1) ensures that
\[
\pi'_{2,r}(x) \leq \frac{4Kx}{135r(\log x)^2} \leq \frac{5K\pi_2(x)}{135(r-2)}
\]
for sufficiently large $x$. Now we fix $z$ such that $5K/(135(z-2)) < 10^{-9}$. Since \(\log(1+t) < t\) for \(t > 0\), for sufficiently large $x$ we obtain
\[
A_2(x) = \sum_{x \leq r \leq (\log x)^3} \pi'_{2,r}(x) \log \left( 1 + \frac{1}{r-1} \right)
\]
\[
\leq \frac{5K\pi_2(x)}{135} \sum_{x \leq r \leq (\log x)^3} \frac{1}{r-2} \log \left( 1 + \frac{1}{r-1} \right)
\]
\[
< \frac{5K\pi_2(x)}{135} \sum_{x \leq r \leq (\log x)^3} \left( \frac{1}{r-2} - \frac{1}{(r-1)} \right)
\]
\[
= \frac{5K\pi_2(x)}{135} \sum_{x \leq r \leq (\log x)^3} \left( \frac{1}{r-2} - \frac{1}{r-1} \right)
\]
\[
< 10^{-9} \pi_2(x).
\]

(c) Suppose that $(\log x)^3 < r \leq x$. By (4–3), the primes counted by $\pi'_{2,r}(x)$ lie in an arithmetic progression modulo $385r$. Thus, their number is at most
\[
\pi_{2,r}(x) \leq \left\lfloor \frac{x}{385r} \right\rfloor + 1 \leq \frac{x}{385r} + 1.
\]
Since \(\log(1+t) < t\) for \(t > 0\), for sufficiently large $x$ we obtain
\[
A_3(x) = \sum_{(\log x)^3 < r \leq x} \pi'_{2,r}(x) \log \left( 1 + \frac{1}{r-1} \right)
\]
\[
\leq \sum_{(\log x)^3 < r \leq x} \frac{1}{(r-1)} \left( \frac{x}{385r} + 1 \right)
\]
\[
\leq \frac{x}{385} \sum_{r > (\log x)^3} \frac{1}{r(r-1)} + \sum_{(\log x)^3 < r \leq x} \frac{1}{r-1}
\]
\[ \leq \frac{x}{385} \sum_{r>(\log x)^3} \left( \frac{1}{r-1} - \frac{1}{r} \right) + \int_{(\log x)^3-2}^{x} \frac{dt}{t} \]
\[ \leq \frac{x}{385((\log x)^3 - 1)} + \left( \log t \right)_{t=(\log x)^3-2}^{t=x} \]
\[ \leq \frac{2x}{385(\log x)^3} + \log x \]
\[ = \left( \frac{1}{385c_2 \log x} + \frac{(\log x)^3}{2c_2x} \right) \frac{2C_2x}{(\log x)^2} \]
\[ = (1 + o(1)) \left( \frac{1}{385c_2 \log x} + \frac{(\log x)^3}{2c_2x} \right) \pi_2(x) \]
\[ < 10^{-9} \pi_2(x). \]

**Step 5.** Returning to (4–5) and using the preceding three estimates, we have
\[
A(x) = A_1(x) + A_2(x) + A_3(x) \\
< 0.000188866 \pi_2(x) + 10^{-9} \pi_2(x) + 10^{-9} \pi_2(x) \\
< 0.000188866 \pi_2(x).
\]

for sufficiently large \( x \).

**Step 6.** Let \( U(x) \) be the set of primes \( p \) counted by \( \pi'_2(x) \) that are unexceptional; that is, \( \varphi(p-1)/(p-1) \geq \varphi(p+1)/(p+1) \) by Lemma 2. As we have seen, if \( p \in U(x) \), then \( F(p) \geq \log(77/72) \). Thus,
\[
0 \leq \#U(x) \log(77/72) \leq \sum_{p \in U(x)} F(p) \leq A(x) \]
\[ \leq 0.000188866 \pi_2(x), \]
from which we deduce that
\[
\#U(x) \leq \left( \frac{0.000179}{\log(77/72)} \right) \pi_2(x) < 0.00281306 \pi_2(x). \]

The primes \( p \) counted by \( \pi'_2(x) \) which are not in \( U(x) \) are exceptional; that is \( \varphi(p-1)/(p-1) < \varphi(p+1)/(p+1) \). By (4–2) and the preceding calculation, for large \( x \) there are at least
\[
\pi'_2(x) - \#U(x) > (0.00740740 - 0.00281306) \pi_2(x) \\
> 0.00459 \pi_2(x)
\]
such primes. This completes the proof of statement (a) from Theorem 1.

**Proof of (b).** This is similar to the preceding, although it is much simpler. As before, \( p, q, r \) denote primes. If \( p, p+2 \) are prime and \( p \) is exceptional, then
\[
\frac{1}{2} \prod_{r|(p-1)} \left( 1 - \frac{1}{r} \right) = \frac{\varphi(p-1)}{p-1} \leq \frac{\varphi(p+1)}{p+1} \leq \frac{1}{3}
\]
since \( 3 \mid (p-1) \) and \( 6 \mid (p+1) \). If we let
\[
G(p) = \sum_{r|(p+1)} \log \left( 1 + \frac{1}{r-1} \right),
\]
then \( G(p) \geq \log(3/2) \) holds for all exceptional primes \( p \). Let \( \pi'_e(x) \) denote the number of exceptional primes \( p \leq x \). Then
\[
\pi'_e(x) \log(3/2) \leq \sum_{p \text{ counted by } \pi'_e(x)} \log \left( 1 + \frac{1}{r-1} \right) \leq \sum_{r \equiv 1(\mod r)} 1 \]
\[ \leq (1 + o(1)) \pi'_e(x) \sum_{r \equiv 1(\mod r)} \log \left( 1 + \frac{1}{r-1} \right) \]
\[ < 0.14137 \pi_2(x), \]
which shows that there are at least
\[
\pi'_2(x) - \pi'_e(x) \geq \pi_2(x) \left( 1 - \frac{0.14137}{\log(3/2)} \right) > 0.6513 \pi_2(x)
\]
unexceptional primes at most \( x \). \( \square \)

**5. Conjectured density**

Below we conjecture a value for the density of the exceptional primes relative to the twin primes. In what follows, we let \( P(n) \) denote the largest prime factor of \( n \) and let \( p(n) \) denote the smallest. We let \( \mu \) denote the Möbius function and remind the reader that \( \mu^2(ab) = 1 \) ensures that \( ab \) is a product of distinct prime factors. The restriction \( 5 \leq p(ab) \leq P(ab) \leq \frac{1}{\epsilon} \) implies that only finitely many prime factors are available to form \( a \) and \( b \). In principle, the right-hand side of (5–1) can be evaluated to arbitrary accuracy by taking \( \epsilon \) sufficiently small. Unfortunately, the number of terms
involved in the sum grows rapidly as \( \varepsilon \) shrinks and we are unable to obtain a reliable numerical estimate from (5–1).

As a brief “sanity check,” we also remark that the limit in (5–1), if it exists, is at most 1. Without the condition

\[
\frac{\varphi(a)}{2a} < \frac{\varphi(b)}{3b},
\]

the inner sum in (5–1) is

\[
\sum_{\substack{a, b \mid \mu^2(a)b = 1 \\ s \leq p(a) \leq q(a) \leq \frac{1}{2}}} \frac{1}{p - 4} = \sum_{\substack{s \leq p \leq q \\ \mu^2(s) = 1}} 2^{\nu(s)} \prod_{p \mid \mu^2(s)} \left( 1 + \frac{2}{p - 4} \right)
\]

\[
= \sum_{s \leq \frac{1}{2}} \left( p - 2 \right),
\]

which precisely offsets the first product in (5–1).

To proceed, we need to generalize the functions \( F \) and \( G \) that appeared in the proof of Theorem 1. Let \( \varepsilon > 0 \) and define

\[
F_\varepsilon(p) = \sum_{\substack{r \mid (p+1) \\ r \geq \frac{1}{2}}} \log \left( 1 + \frac{1}{r - 1} \right)
\]

\[
G_\varepsilon(p) = \sum_{\substack{r \mid (p-1) \\ r \geq \frac{1}{2}}} \log \left( 1 + \frac{1}{r - 1} \right).
\]

Particular instances of these functions have appeared in the proof of Theorem 1 with \( \varepsilon = 1/5 \) for \( F_\varepsilon \) (called \( F \)) and \( \varepsilon = 1/13 \) for \( G_\varepsilon \) (called \( G \)), respectively.

**Lemma 3.** For \( \varepsilon > 0 \), the number of twin primes \( p \leq x \) such that \( F_\varepsilon(p) > \varepsilon \) is \( O((\log(x))^1/4)\pi_2(x) \). The same conclusion holds with \( F_\varepsilon \) replaced by \( G_\varepsilon \).

**Proof.** The argument is essentially already in the proof of Theorem 1. We do it only for \( F_\varepsilon(p) \) since the argument for \( G_\varepsilon(p) \) is similar. We sum \( F_\varepsilon(p) \) for \( p \leq x \) with \( p, p + 2 \) prime and use the fact that \( \log(1 + t) \leq t \) to obtain

\[
\sum_{p \leq x \atop p, p + 2 \prime} F_\varepsilon(p) \leq \sum_{p \leq x \atop q(p-1) \atop q > \frac{1}{2}} \frac{1}{q - 1} = \sum_{q > \frac{1}{2}} \frac{1}{q - 1} \sum_{p \leq x \atop p, p + 2 \prime \atop q(p-1) \equiv 1 \pmod{q}} 1 = \sum_{q > \frac{1}{2}} \pi_2(x, q, 1) \frac{1}{q - 1},
\]

in which \( \pi_2(x; q, 1) \) denotes the number of primes \( p \leq x \) with \( p, p + 2 \) prime and \( p \equiv 1 \pmod{q} \). By the usual argument, the number of twin primes \( p, p + 2 \) with \( p \leq x \) and \( p \equiv 1 \pmod{q} \) equals the number of \( t \leq x/q \) such that \( qt + 1 \) and \( qt + 3 \) are prime. The number of them is, by the Brun sieve,

\[
\pi_2(x; q, 1) \ll \frac{x}{(q-1)(\log x)^2}.\]

The Prime Number Theorem and Abel summation reveal that

\[
\sum_{p \leq x \atop p, p + 2 \prime} F_\varepsilon(p) \ll \frac{x}{(\log x)^2} \sum_{q > \frac{1}{2}} \frac{1}{q - 1} \ll \varepsilon \pi_2(x),
\]

If we let

\[ A_\varepsilon = \{ p : p, p+2 \ prime and F_\varepsilon(p) > \varepsilon \}, \]

then

\[
\#A_\varepsilon(x) \varepsilon \ll \sum_{p \leq x \atop p, p+2 \prime} F_\varepsilon(p) \ll \varepsilon \left( \log \left( \frac{1}{\varepsilon} \right) \right)^{-1} \pi_2(x),
\]

which gives \( \#A_\varepsilon(x) = O((\log(x))^{-1})\pi_2(x) \).

To justify our conjecture, we look at the \( \frac{1}{2} \)-part of \( p^2 - 1 \). We first let \( \varepsilon \leq 0.5 \). We note that \( 2|(p-1), 2|(p+1) \) and \( 3|(p+1) \) for all twin primes \( p \geq 5 \). For two co prime square-free numbers \( a, b \) with \( 5 \leq p(ab) \leq P(ab) \leq \frac{1}{2} \), we say that the twin prime \( p \) is of \( \frac{1}{2} \)-type \( (a, b) \) if

\[
p - 1 = 2^a \prod_{q \mid a} q^{\alpha_q} \prod_{q > \frac{1}{2}} q^{\alpha_q} \quad \text{and} \quad p + 1 = 2^b \prod_{q \mid b} q^{\beta_q} \prod_{q > \frac{1}{2}} q^{\beta_q}
\]

for some positive \( \alpha, \beta, \gamma, \alpha_q \) and \( \beta_q \) for \( q \mid ab \) and non-negative \( \gamma_q, \delta_q \) for \( q \geq \frac{1}{2} \). That is, the prime factors of \( p - 1 \) that are \( \leq \frac{1}{2} \) are exactly the ones dividing \( 2a \) and the prime factors of \( p + 1 \) that are \( \leq \frac{1}{2} \) are exactly the ones dividing \( 6b \).

Given \( \varepsilon \) and \( (a, b) \), let

\[
\varepsilon_{a,b} = \prod_{q \leq \frac{1}{2}} q.
\]

\[
\frac{\varphi(p-1)}{p-1} = \frac{\varphi(a)}{2} \prod_{q \mid (p-1) \atop q > \frac{1}{2}} \left( 1 - \frac{1}{q} \right) \quad \text{and} \quad \frac{\varphi(p+1)}{p+1} = \frac{\varphi(b)}{3} \prod_{q \mid (p+1) \atop q > \frac{1}{2}} \left( 1 - \frac{1}{q} \right).
\]

Since

\[
e^{-2y} < 1 - y < e^{-y} \quad \text{for} \ y < \frac{1}{2},\]
it follows that
\[ 1 - 4\varepsilon < e^{-2\varepsilon} < e^{-E(p)} = \prod_{\substack{q|\mu(p-1) \\ q \neq \frac{1}{2}}} \left( 1 - \frac{1}{q} \right) \]

hold for all twin primes \( p \leq x \) except the ones in \( A_\varepsilon(x) \), a set of cardinality \( O((\log(\frac{1}{\varepsilon}))^{-1}\pi_2(x)) \). Consequently,
\[ (1 - 4\varepsilon)^{-\varphi(a)} \leq \frac{\varphi(p-1)}{p-1} \]

holds for all but \( O((\log(\frac{1}{\varepsilon}))^{-1}\pi_2(x)) \) twin primes \( p \leq x \). Thus, the inequality
\[ \frac{\varphi(p-1)}{p-1} \leq \frac{\varphi(p+1)}{p+1} \]

implies that
\[ \frac{\varphi(a)}{2a} \leq (1 - 4\varepsilon)^{-1}\frac{\varphi(b)}{3b} . \]

Let us consider twin primes for which
\[ \frac{\varphi(b)}{3b} < \frac{\varphi(a)}{2a} < (1 - 4\varepsilon)^{-1}\frac{\varphi(b)}{3b} \quad (5-2) \]

occurs. Since
\[ \frac{\varphi(a)}{2a} = \frac{\varphi(p-1)}{p-1} (1 + O(\varepsilon)) \quad \text{and} \]
\[ \frac{\varphi(b)}{3b} = \frac{\varphi(p+1)}{p+1} (1 + O(\varepsilon)) \]

for all \( p \leq x \) with \( O((\log(\frac{1}{\varepsilon}))^{-1}\pi_2(x)) \) exceptions, it follows that twin primes \( p \leq x \) for which (5-2) holds have the additional property that
\[ \left| \frac{\varphi(p-1)}{p-1} - \frac{\varphi(p+1)}{p+1} \right| = O(\varepsilon) . \quad (5-3) \]

Let \( B_\varepsilon \) be the set of twin primes for which (5-3) holds. We make the following additional assumption.

**Additional assumption.** The number of twin primes \( p \leq x \) for which (5-3) holds is \( O(h(\varepsilon)\pi_2(x)) \) for some function \( h(y) \) with \( h(y) \to 0 \) as \( y \to 0 \).

Assumption (5-3) has been shown to hold when \( p \) is only a prime [Garcia and Luca]. That is, the number of primes \( p \leq x \) such that (5-3) holds is at most \( O(h(\varepsilon)\pi_2(x)) \), where \( h(\varepsilon) \) tends to zero when \( \varepsilon \to 0 \). In fact, this was a crucial step in showing that \( \varphi(p-1) - \varphi(p+1) \) has no bias if only \( p \) is assumed to be prime.

Proving this for primes uses the Turán–Kubilius theorem about the number of prime factors \( q \leq y \) of \( p \pm 1 \) when \( p \) is prime as the parameter \( y \) tends to infinity and also Sperner’s theorem from combinatorics. With some nontrivial effort, which involves proving first a Turán–Kubilius estimate for the number of distinct primes \( q \leq 1/\varepsilon \) of \( p-1 \) and \( p+1 \) when \( p \) ranges over twin primes up to \( x \), the same program can be applied to prove that the additional assumption holds under the Bateman–Horn conjectures. We do not give further details here.

Assume that the additional assumption holds. Then the set of twin primes \( p \leq x \) such that
\[ \frac{\varphi(p-1)}{p-1} < \frac{\varphi(p+1)}{p+1} \]

is within a set of cardinality \( O(h(\varepsilon)\pi_2(x)) \) from the set of primes for which
\[ \frac{\varphi(a)}{2a} < \frac{\varphi(b)}{3b} . \quad (5-4) \]

With this assumption, we proceed as in [Garcia and Luca, Section 2.11]. Fix \( \frac{1}{2}, a, b, \) and \( c = c_{a,b} \). We also fix a residue class for \( p \) modulo \( c \) which is not \( \{0, \pm 1, -2\} \). In this case we need to count natural numbers of the form
\[ abct + \kappa , \]

in which \( \kappa \) is fixed such that
- \( abct + \kappa \leq x \),
- \( abct + \kappa \) and \( abct + \kappa + 2 \) are prime,
- \( abct + \kappa - 1 \) are divisible by all primes in \( a \) and coprime to \( cb \),
- \( abct + \kappa + 1 \) is divisible by all primes in \( b \) (and coprime to \( ca \)).

Observe that \( \kappa \) is uniquely determined modulo \( abc \) once it is determined modulo \( c \). By the Bateman–Horn conjecture, this number is
\[ (1 + o(1))\pi_2(x) \prod_{p|abc} \left( 1 - \frac{1}{p-2} \right) . \]

We next sum this over all \( q \leq 4 \) progressions modulo \( q \) for which \( abct + \kappa \) is not congruent modulo \( q \) to some member of \( \{0, \pm 1, -2\} \) and for all \( q \) \( | \) \( c \) getting an amount of
\[ (1 + o(1))\pi_2(x) \prod_{p|abc} \left( 1 - \frac{1}{p-2} \right) \prod_{p|c} \left( \frac{q-4}{q-2} \right) \]
\[ = (1 + o(1)) \prod_{5 \leq q \leq 4} \left( \frac{q-4}{q-2} \right) \prod_{q|abc} \left( \frac{1}{q-4} \right) . \]

We now sum up over all pairs \( a, b \) with
\[ \frac{\varphi(a)}{2a} < \frac{\varphi(b)}{3b} , \]

which yields a proportion of
\[ (1 + o(1)) \prod_{5 \leq q \leq 4} \left( \frac{q-4}{q-2} \right) \sum_{a, b} \mu^2(ab) \prod_{p|abc} \left( \frac{1}{p-4} \right) \]

of \( \pi_2(x) \) with a number of exceptions \( p \leq x \) of counting function \( O(h(\varepsilon)\pi_2(x)) \). This supports Conjecture 2.
6. Comments

We did not need the full strength of the Bateman–Horn conjecture, just the case \( r = 2 \) and \( D = 1 \) for certain specific pairs of linear polynomials \( f_1(t) \) and \( f_2(t) \). Under this conjecture, we have seen that \( \varphi(p − 1) \leq \varphi(p + 1) \) for a substantial majority of twin prime pairs \( p, p + 2 \).

There are a few twin primes \( p, p + 2 \) for which
\[
\varphi(p − 1) = \varphi(p + 1). \tag{6–1}
\]
For only such \( p \leq 100,000,000 \) are
\[
5, 11, 71, 2591, 208,391, 16,692,551, 48,502,931,
92,012,201, 249,206,231, 419,445,251, 496,978,301.
\]
The following result highlights the rarity of these twin primes.

**Theorem 4.** The number of primes \( p \leq x \) with \( p + 2 \) prime and \( \varphi(p − 1) = \varphi(p + 1) \) is \( O(x/\exp((\log x)^{1/3})) \).

**Proof.** Suppose that \( j \) and \( j + k \) have the same prime factors, let \( g = (j, j + k) \), and suppose that
\[
\frac{j}{g} + 1 \quad \text{and} \quad \frac{j + k}{g} + 1 \tag{6–2}
\]
are primes that do not divide \( j \). Then
\[
n = j \left( \frac{j + k}{g} + 1 \right) \tag{6–3}
\]
satisfies \( \varphi(n) = \varphi(n + k) \) [Graham et al. 99, Thm. 1]. For \( k \) fixed, the number of solutions \( n \leq x \) to \( \varphi(n) = \varphi(n + k) \) which are not of the form (6–3) is less than \( x/\exp((\log x)^{1/3}) \) for sufficiently large \( x \) [Graham et al. 99, Thm. 2].

We are interested in the case \( k = 2 \) and \( n = p − 1 \), in which \( p, p + 2 \) are prime. If \( j \) and \( j + k \) have the same prime factors, then they are both powers of \( 2 \). Thus, \( j = 2 \) and \( j + k = 4 \), so \( g = 2 \). From (6–2) we see that \( r \) is such that
\[
r + 1 \quad \text{and} \quad 2r + 1
\]
are prime. Then \( n = 2(2r + 1) = p − 1 \), from which it follows that \( p = 4r + 3 \) and \( p + 2 = 4r + 5 \) are prime. Consequently,
\[
r + 1, \quad 2r + 1, \quad 4r + 3, \quad \text{and} \quad 4r + 5,
\]
are prime. However, this occurs only for \( r = 2 \) since otherwise one of the preceding is a multiple of 3 that is larger than 3. \( \square \)

In particular, the number of primes \( p \leq x \) for which \( p + 2 \) is prime and \( \varphi(p − 1) = \varphi(p + 1) \) is \( o(x/\log x) \).

Assuming the first Hardy–Littlewood conjecture, it follows that the set of such primes has density zero in the twin primes.

**Acknowledgments**

We thank Tomás Silva for independently computing the ratio \( \pi_1(x)/\pi_2(x) \) for large \( x \). We also thank the anonymous referee for suggesting the approach of Section 5.

**Funding**

SRG supported by NSF grant DMS-1265973, a David L. Hirsch III and Susan H. Hirsch Research Initiation Grant, and the Budapest Semesters in Mathematics (BSM) Director's Mathematician in Residence (DMiR) program.

FL was supported in part by grants CPRRI160325161141 and an A-rated researcher award both from the NRF of South Africa and by grant no. 17-028045 of the Czech Granting Agency.

**References**


