

ON THE DIFFERENCE IN VALUES OF THE EULER TOTIENT FUNCTION NEAR PRIME ARGUMENTS

STEPHAN RAMON GARCIA AND FLORIAN LUCA

ABSTRACT. We prove unconditionally that for each $\ell \geq 1$, the difference $\varphi(p - \ell) - \varphi(p + \ell)$ is positive for 50% of odd primes p and negative for 50%.

1. INTRODUCTION

In what follows, p always denotes an odd prime number. The inequality $\varphi(p - 1) \geq \varphi(p + 1)$ appears to hold for an overwhelming majority of twin primes $p, p + 2$, and to be reversed for small, but positive, proportion of the twin primes [4]. To be more specific, if the Bateman–Horn conjecture is true, then the inequality above holds for at least 65.13% of twin prime pairs and is reversed for at least 0.47% of pairs. Numerical evidence suggests, in fact, that the ratio is something like 98% to 2%. In other words, for an overwhelming majority of twin prime pairs $p, p + 2$, it appears that the first prime has more primitive roots than does the second.

Based upon numerical evidence, it was conjectured in [4] that this bias disappears if only p is assumed to be prime. That is, $\varphi(p - 1) > \varphi(p + 1)$ for 50% of primes and the inequality is reversed for 50% of primes. We prove this unconditionally and, moreover, we are able to handle wider spacings as well. If all primality assumptions are dropped, then it is known that $\varphi(n - 1) > \varphi(n + 1)$ asymptotically 50% of the time. This follows from work of Shapiro, who considered the distribution function of $\varphi(n)/\varphi(n - 1)$ [11].

Let $\pi(x)$ denote the number of primes at most x and let \sim denote asymptotic equivalence. The Prime Number Theorem ensures that $\pi(x) \sim x/\log x$. Our main theorem is the following.

Theorem 1.1. *Let ℓ be a positive integer. As $x \rightarrow \infty$ we have the following:*

- (a) $\#\{p \leq x : \varphi(p - \ell) > \varphi(p + \ell)\} \sim \frac{1}{2}\pi(x)$.
- (b) $\#\{p \leq x : \varphi(p - \ell) < \varphi(p + \ell)\} \sim \frac{1}{2}\pi(x)$.
- (c) $\#\{p \leq x : \varphi(p - \ell) = \varphi(p + \ell)\} = o(\pi(x))$.

2010 *Mathematics Subject Classification.* 11A41, 11A07, 11N36, 11N37.

Key words and phrases. Euler totient, prime, twin prime, Bateman–Horn conjecture, Twin Prime Conjecture, Brun Sieve.

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.

F. L. was supported in part by grants CPRR160325161141 and an A-rated researcher award both from the NRF of South Africa and by grant no. 17-02804S of the Czech Granting Agency.

ℓ	#	ℓ	#	ℓ	#	ℓ	#	ℓ	#	ℓ	#	ℓ	#		
1	103	9	359	17	106	25	6	33	338	41	109	49	38	57	295
2	49	10	4	18	219	26	47	34	47	42	322	50	5	58	54
3	201078	11	107	19	104	27	357	35	3	43	121	51	371	59	127
4	58	12	214	20	3	28	17	36	374	44	39	52	38	60	538
5	5	13	98	21	403	29	117	37	97	45	486	53	126	61	126
6	231	14	7	22	52	30	507	38	45	46	47	54	303	62	45
7	43	15	108772	23	136	31	98	39	380	47	124	55	2	63	22654
8	50	16	39	24	301	32	53	40	5	48	236	56	6	64	48

Table 1. The number (#) of primes $p \leq 2,038,074,743$ (the hundred millionth prime) for which $\varphi(p - \ell) = \varphi(p + \ell)$. This number is exceptionally large if $\ell = 4^n - 1$; see Theorem 3.1 for an explanation.

A curious phenomenon occurs in (c), in the sense that the decay rate relative to $\pi(x)$ depends upon ℓ in a peculiar manner. Theorem 3.1 shows that

$$\#\{p \leq x : \varphi(p - \ell) = \varphi(p + \ell)\} \ll \begin{cases} \frac{x}{(\log x)^3} & \text{if } \ell = 4^n - 1, \\ \frac{x}{e^{(\log x)^{1/3}}} & \text{otherwise.} \end{cases}$$

This does not appear to be an artifact of the proof since it is borne out in numerical computations (see Table 1) and is consistent with the Bateman–Horn conjecture.

We first prove Theorem 1.1 in the case $\ell = 1$. This is undertaken in Section 2 and it comprises the bulk of this article. For the sake of readability, we break the proof into several steps which we hope are easy to follow. In Section 3, we outline the modifications necessary to treat the case $\ell \geq 2$. This approach permits us to focus on the main ingredients that are common to both cases, without getting sidetracked by all of the adjustments necessary to handle the general case.

2. PROOF OF THEOREM 1.1 FOR $\ell = 1$

2.1. The case of equality. Our first job is to show that the set of primes p for which $\varphi(p - 1) = \varphi(p + 1)$ has a counting function that is $o(\pi(x))$. We need the following lemma, which generalizes earlier work by Erdős, Pomerance, and Sárközy [3] in the case $k = 1$. The upper bound (b) in the following was strengthened in a preprint of Yamada [14].

Lemma 2.1 (Graham–Holt–Pomerance [6]). *Suppose that j and $j + k$ have the same prime factors. Let $g = \gcd(j, j + k)$ and suppose that*

$$\frac{jt}{g} + 1 \quad \text{and} \quad \frac{(j+k)t}{g} + 1 \tag{2.2}$$

are primes that do not divide j .

(a) *Then $n = j \left(\frac{(j+k)t}{g} + 1 \right)$ satisfies $\varphi(n) = \varphi(n + k)$.*

(b) *For k fixed and sufficiently large x , the number of solutions $n \leq x$ to $\varphi(n) = \varphi(n + k)$ that are not of the form above is less than $x / \exp((\log x)^{1/3})$.*

Consider the case $k = 2$ and $n = p - 1$, in which p is prime. Suppose that j and $j + 2$ have the same prime factors and let $g = \gcd(j, j + 2)$. Let us also suppose that t is a positive integer such that

$$\frac{jt}{g} + 1 \quad \text{and} \quad \frac{(j+2)t}{g} + 1$$

are primes and

$$p - 1 = j \left(\frac{(j+2)t}{g} + 1 \right).$$

Since j and $j + 2$ have the same prime factors, they are both powers of 2. Then $j = 2$ and $j + 2 = 4$, so $g = 2$. Consequently,

$$t + 1, \quad 2t + 1, \quad \text{and} \quad 4t + 3 \tag{2.3}$$

are prime. Reduction modulo 3 reveals that at least one of them is a multiple of 3. The only prime triples produced by (2.3) are $(2, 3, 7)$ and $(3, 5, 11)$, in which $r = 1$ and $r = 2$, respectively. Consequently,

$$\#\{p \leq x : \varphi(p - 1) = \varphi(p + 1)\} < \frac{x}{\exp((\log x)^{1/3})} + 2 = o(\pi(x)). \tag{2.4}$$

This is Theorem 1.1.c in the case $\ell = 1$.

2.2. A comparison lemma. Instead of comparing $\varphi(p - 1)$ and $\varphi(p + 1)$ directly, it is more convenient to compare the related quantities

$$\frac{\varphi(p - 1)}{p - 1} = \prod_{q|(p-1)} \left(1 - \frac{1}{q}\right) \quad \text{and} \quad \frac{\varphi(p + 1)}{p + 1} = \prod_{q|(p+1)} \left(1 - \frac{1}{q}\right), \tag{2.5}$$

in which q is prime. Let

$$S(p) := \frac{\varphi(p - 1)}{p - 1} - \frac{\varphi(p + 1)}{p + 1}, \tag{2.6}$$

which we claim is nonzero for $p \geq 5$. Let $P(n)$ denote the largest prime factor of n . Since

$$\frac{\varphi(n)}{n} = \prod_{q|n} \left(\frac{q - 1}{q}\right), \tag{2.7}$$

it follows that $P(n)$ is the largest prime factor of the denominator of $\varphi(n)/n$. Since $\gcd(p - 1, p + 1) = 2$, the condition $S(p) = 0$ implies that $p - 1$ and $p + 1$ are both powers of 2. Thus, $S(p) = 0$ holds only for $p = 3$.

Something similar to the following lemma is in [4], although there it was assumed that $p + 2$ is also prime. The adjustment for $\ell \geq 2$ is discussed in Section 3.

Lemma 2.8 (Comparison Lemma). *The set of primes p for which $\varphi(p - 1) - \varphi(p + 1)$ and $S(p)$ have the same sign has counting function asymptotic to $\pi(x)$ as $x \rightarrow \infty$.*

Proof. In light of (2.4), it suffices to show that

$$\varphi(p - 1) > \varphi(p + 1) \quad \iff \quad \frac{\varphi(p - 1)}{p - 1} > \frac{\varphi(p + 1)}{p + 1} \tag{2.9}$$

on a set of full density in the primes. The forward direction is clear, so we focus on the reverse. If the inequality on the right-hand side of (2.9) holds, then

$$\begin{aligned} 0 &< p(\varphi(p - 1) - \varphi(p + 1)) + \varphi(p - 1) + \varphi(p + 1) \\ &\leq p(\varphi(p - 1) - \varphi(p + 1)) + \frac{1}{2}(p - 1) + \frac{1}{2}(p + 1) \end{aligned}$$

$$= p(\varphi(p-1) - \varphi(p+1) + 1) \quad (2.10)$$

because $p-1$ and $p+1$ are even. Since $\varphi(n)$ is even for $n \geq 3$, it follows that $\varphi(p-1) - \varphi(p+1) \geq 0$. Now appeal to (2.4) to see that strict inequality holds on a set of full density in the primes. \square

2.3. Some preliminaries. In our later study of the quantity $S(p)$, we need to avoid four classes of inconvenient primes. To make the required estimates, we need some notation. Let x be large, let $y := \log \log x$, and define

$$L_y := \text{lcm}\{m : m \leq y\}. \quad (2.11)$$

Then $L_y = e^{\psi(y)}$, in which

$$\psi(y) := \sum_{p^k \leq y} \log p$$

is Chebyshev's function. Since the Prime Number Theorem asserts that $\psi(y) = y + o(y)$ as $y \rightarrow \infty$, we obtain

$$L_y = e^{y+o(y)} < e^{2y} = (\log x)^2 \quad (2.12)$$

for sufficiently large x . For a positive integer n , let $D_y(n)$ denote the largest divisor of n that is y -smooth:

$$D_y(n) := \max\{d : d|n \text{ and } P(d) \leq y\}. \quad (2.13)$$

On occasion, we will need the Brun sieve. Let f_1, f_2, \dots, f_m be a collection of distinct irreducible polynomials with positive leading coefficients. An integer n is *prime generating* for this collection if each $f_1(n), f_2(n), \dots, f_m(n)$ is prime. Let $G(x)$ 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. As $x \rightarrow \infty$,

$$G(x) \ll \frac{x}{(\log x)^m},$$

in which the implied constant depends only upon m and $\prod_{i=1}^m \deg f_i$ [13, Thm. 3, Sect. I.4.2]. The upper bound obtained in this manner has the same order of magnitude as the prediction furnished by the Bateman–Horn conjecture [1].

2.4. Inconvenient primes of Type 1. Let

$$\mathcal{E}_1(x) := \{p \leq x : D_y(p^2 - 1) \nmid L_y\}.$$

We will prove that

$$\#\mathcal{E}_1(x) \ll \frac{x}{y^{1/2} \log x} = o(\pi(x)). \quad (2.14)$$

Suppose that $p \in \mathcal{E}_1(x)$. Then (2.13) and the definition of $\mathcal{E}_1(x)$ yield a divisor d of $p^2 - 1$ such that $P(d) \leq y$ and $d \nmid L_y$. These conditions provide a prime power q^b with least exponent b such that

$$q^b | d, \quad q = P(q^b) \leq y, \quad \text{and} \quad y < q^b. \quad (2.15)$$

Indeed, if every y -smooth prime power q^b that divides d satisfies $q^b \leq y$, then (2.11) would imply that $d | L_y$, a contradiction. We also observe that the second two conditions in (2.15) ensure that $b \geq 2$.

We claim that for large x either $p-1$ or $p+1$ has a prime power divisor q^c with $c \geq 2$ in the interval $[y/2, y^2]$. Since

$$p^2 - 1 = (p-1)(p+1) \quad \text{and} \quad \gcd(p-1, p+1) = 2,$$

it follows that $d/(2, d)$ divides one of $p + 1, p - 1$. There are two cases to consider.

- If $q = 2$, then 2^{b-1} divides one of $p - 1, p + 1$. Since we aim to prove (2.14) as $x \rightarrow \infty$, we may assume that $b \geq 3$ since for $b = 2$, the third statement in (2.15) implies $\log \log x < 4$. Next observe that (2.15) implies that $y/2 < 2^{b-1}$. The minimality of b in (2.15) ensures that $2^{b-1} \leq y < y^2$. Thus, $2^{b-1} \in [y/2, y^2]$ has $b - 1 \geq 2$ (since $b \geq 3$), and divides one of $p - 1, p + 1$.
- If q is odd, then q^b divides $p - 1$ or $p + 1$. The minimality of b ensures that $q^{b-1} \leq y$ and the second statement in (2.15) yields

$$\frac{y}{2} < y < q^b = q^{b-1}q \leq y^2, \quad \text{and hence } q^b \in [\frac{y}{2}, y^2] \quad \text{with } b \geq 2.$$

For large x , we conclude that one of $p - 1, p + 1$ has a prime power divisor q^c with $c \geq 2$ in $[y/2, y^2]$.

Let $\pi_s(x)$ denote the number of prime powers p^a with $a \geq 2$ that are at most x . Since $p^a \leq x$ with $a \geq 2$ implies either $a = 2$ and $p \leq x^{1/2}$, or $a \in [3, (\log x / \log 2)]$ and $p \leq x^{1/3}$, the Prime Number Theorem implies that

$$\pi_s(x) = \pi(\sqrt{x}) + O(\pi(x^{1/3}) \log x) = (2 + o(1)) \frac{x^{1/2}}{\log x}$$

as $x \rightarrow \infty$. Let $\pi(x; m, k)$ denote the number of primes at most x that are congruent to k modulo m . Then the Brun sieve implies

$$\begin{aligned} \#\mathcal{E}_1(x) &\leq \sum_{\substack{q^b \in [y/2, y^2] \\ b \geq 2}} \pi(x; q^b, 1) + \sum_{\substack{q^b \in [y/2, y^2] \\ b \geq 2}} \pi(x; q^b, -1) \\ &\ll \sum_{\substack{q^b \in [y/2, y^2] \\ b \geq 2}} \frac{x}{\varphi(q^b) \log x} \ll \sum_{\substack{q^b \in [y/2, y^2] \\ b \geq 2}} \frac{2x}{q^b \log x} \\ &\ll \frac{x}{\log x} \sum_{\substack{q^b \in [y/2, y^2] \\ b \geq 2}} \frac{1}{q^b} \\ &\ll \frac{x}{\log x} \int_{y/2}^{y^2} \frac{d\pi_s(t)}{t} \leq \frac{x}{\log x} \int_{y/2}^{\infty} \frac{d\pi_s(t)}{t} \\ &\ll \frac{x}{\log x} \left(\frac{\pi_s(t)}{t} \Big|_{y/2}^{\infty} + \int_{y/2}^{\infty} \frac{\pi_s(t) dt}{t^2} \right) \\ &\ll \frac{x}{\log x} \left(\frac{(y/2)^{1/2}}{(y/2) \log(y/2)} + \int_{y/2}^{\infty} t^{-3/2} (\log t)^{-1/2} dt \right) \\ &\ll \frac{x}{y^{1/2} (\log y) (\log x)} = o(\pi(x)) \end{aligned}$$

as $x \rightarrow \infty$. This is the desired estimate (2.14).

2.5. Inconvenient primes of Type 2. Fix a large x and define the function

$$h_y(n) := \sum_{\substack{r|n \\ r > y}} \frac{1}{r},$$

in which r is prime. Let

$$\mathcal{E}_2(x) := \left\{ p \leq x : h_y(p-1) > \frac{1}{y\sqrt{\log y}} \text{ or } h_y(p+1) > \frac{1}{y\sqrt{\log y}} \right\}.$$

We claim that

$$\#\mathcal{E}_2(x) \ll \frac{\pi(x)}{\sqrt{\log y}} = o(\pi(x)) \quad (2.16)$$

as $x \rightarrow \infty$. This will follow from an averaging argument similar to [7, Lem. 3].

The Brun sieve with $f(t) = rt \pm 1$ provides

$$\pi(x; r, \pm 1) \ll \frac{\pi(x)}{r} \quad \text{for } y \leq r \leq (\log x)^3$$

uniformly for r in the specified range [13, Thm. 3, Sect. I.4.2]. We use the trivial estimate

$$\pi(x; r, \pm 1) \leq \frac{x}{r} \quad \text{for } (\log x)^3 \leq r \leq x.$$

We also require the upper bound

$$\sum_{r \geq y} \frac{1}{r^2} = \int_y^\infty \frac{d\pi(t)}{t^2} = \frac{\pi(t)}{t^2} \Big|_y^\infty + 2 \int_y^\infty \frac{\pi(t) dt}{t^3} \ll \frac{\pi(y)}{y^2} \ll \frac{1}{y \log y}, \quad (2.17)$$

which is afforded by the Prime Number Theorem. As $x \rightarrow \infty$, we have

$$\begin{aligned} \sum_{p \leq x} h_y(p \pm 1) &= \sum_{p \leq x} \sum_{\substack{r|(p \pm 1) \\ r > y}} \frac{1}{r} = \sum_{y \leq r \leq x} \frac{1}{r} \sum_{\substack{p \leq x \\ r|(p \pm 1)}} 1 \\ &= \sum_{y \leq r \leq x} \frac{\pi(x; r, \mp 1)}{r} \\ &\ll \sum_{y \leq r \leq (\log x)^3} \frac{\pi(x)}{r^2} + \sum_{(\log x)^3 < r \leq x} \frac{x}{r^2} \\ &\ll \pi(x) \sum_{r \geq y} \frac{1}{r^2} + x \sum_{r \geq (\log x)^3} \frac{1}{r^2} \\ &\ll \frac{\pi(x)}{y \log y} + \frac{x}{(\log x)^3} \ll \frac{\pi(x)}{y \log y} + \frac{\pi(x)}{(\log x)^2} \\ &\ll \frac{\pi(x)}{y \log y}. \end{aligned}$$

Consequently,

$$\begin{aligned} \frac{\#\mathcal{E}_2(x)}{y\sqrt{\log y}} &\leq \sum_{\substack{p \leq x \\ h_y(p-1) > \frac{1}{y\sqrt{\log y}}}} h_y(p-1) + \sum_{\substack{p \leq x \\ h_y(p+1) > \frac{1}{y\sqrt{\log y}}}} h_y(p+1) \\ &\leq \sum_{p \leq x} h_y(p-1) + \sum_{p \leq x} h_y(p+1) \\ &\ll \frac{\pi(x)}{y \log y}, \end{aligned}$$

which implies (2.16).

2.6. Inconvenient primes of Type 3. Let $\omega(n)$ denote the number of distinct prime factors of n and $\omega_y(n)$ the number of distinct prime factors $q \leq y$ of n . Define

$$\mathcal{E}_3(x) = \{p \leq x : \omega_y(p^2 - 1) \notin [1.5 \log \log y, 2.5 \log \log y]\}.$$

We claim that

$$\#\mathcal{E}_3(x) \ll \frac{\pi(x)}{\log \log y} = o(\pi(x)) \quad (2.18)$$

as $x \rightarrow \infty$. Although this is essentially a result of Erdős [2], we sketch a simpler proof that is easily generalized since we later need to handle $p^2 - \ell^2$ instead of $p^2 - 1$.

If $p \in \mathcal{E}_3(x)$, then for large x we have

$$\omega_y(p^2 - 1) + 1 = \omega_y(p - 1) + \omega_y(p + 1)$$

because $\gcd(p - 1, p + 1) = 2$. Thus, either

$$\min\{\omega_y(p - 1), \omega_y(p + 1)\} \leq 0.75 \log \log y + 1 \leq 0.8 \log \log y$$

or

$$\max\{\omega_y(p - 1), \omega_y(p + 1)\} > 1.25 \log \log y > 1.2 \log \log y$$

for sufficiently large x . Without loss of generality, we may suppose that

$$\omega_y(p - 1) \leq 0.8 \log \log y \quad \text{or} \quad \omega_y(p + 1) \geq 1.2 \log \log y. \quad (2.19)$$

Then

$$0.04(\log \log y)^2 \leq (\omega_y(p - 1) - \log \log y)^2 \quad (2.20)$$

and similarly if $p + 1$ occurs in (2.19).

We next require the following ‘‘Turán–Kubilius’’-type result; see [8, Lem. 2], [10, §V.5, 1, p. 159]. To study $\varphi(p \pm \ell)$ for $\ell \neq 1$ requires a slight generalization; see Lemma 3.2 in Section 3 for a statement and sketch of the proof.

Lemma 2.21 (Motohashi). $\sum_{p \leq x} (\omega_y(p \pm 1) - \log \log y)^2 = O(\pi(x) \log \log y)$.

Now return to (2.20), apply Lemma 2.21, and conclude that

$$\begin{aligned} 0.04(\log \log y)^2 \#\mathcal{E}_3(x) &\leq \sum_{p \in \mathcal{E}_3(x)} (\omega_y(p - 1) - \log \log y)^2 + (\omega_y(p + 1) - \log \log y)^2 \\ &\leq \sum_{p \leq x} (\omega_y(p - 1) - \log \log y)^2 + (\omega_y(p + 1) - \log \log y)^2 \\ &= O(\pi(x) \log \log y). \end{aligned}$$

This yields the desired estimate (2.18).

2.7. Inconvenient primes of Type 4. Let

$$\mathcal{E}_4(x) = \left\{ p \leq x : \frac{p^2 - 1}{\varphi(p^2 - 1)} > (\log y)^{1/3} \right\}. \quad (2.22)$$

We claim that

$$\#\mathcal{E}_4(x) \ll \frac{\pi(x)}{(\log y)^{1/6}} = o(\pi(x)) \quad (2.23)$$

as $x \rightarrow \infty$. Since $\varphi(p - 1)\varphi(p + 1) \leq \varphi(p^2 - 1)$, the condition

$$\frac{p^2 - 1}{\varphi(p^2 - 1)} > (\log y)^{1/3}$$

implies that

$$\frac{p+1}{\varphi(p+1)} > (\log y)^{1/6} \quad \text{or} \quad \frac{p-1}{\varphi(p-1)} > (\log y)^{1/6}.$$

A standard application of the Siegel–Walfisz theorem yields

$$\sum_{p \leq x} \frac{p-1}{\varphi(p-1)} \ll \pi(x); \quad (2.24)$$

see [10, §I.28, 1b, p 30] or [9]. The same holds with $p-1$ replaced by $p+1$; the adjustments necessary to handle $p \pm \ell$ are discussed in Section 3. Thus,

$$\begin{aligned} \mathcal{E}_4(x)(\log y)^{1/6} &\leq \sum_{p \in \mathcal{E}_4(x)} \left(\frac{p-1}{\varphi(p-1)} + \frac{p+1}{\varphi(p+1)} \right) \\ &\leq \sum_{p \leq x} \left(\frac{p-1}{\varphi(p-1)} + \frac{p+1}{\varphi(p+1)} \right) \\ &\ll \pi(x), \end{aligned}$$

which yields (2.23).

2.8. Convenient primes. Throughout the remainder of the proof, we let $5 \leq p \leq x$, in which x is large, and we suppose that

$$p \notin \mathcal{E}_1(x) \cup \mathcal{E}_2(x) \cup \mathcal{E}_3(x) \cup \mathcal{E}_4(x).$$

We say that such a prime is *convenient*. Because $\gcd(p-1, p+1) = 2$, we have

$$D_y(p^2 - 1) = m_1 m_2, \quad (2.25)$$

in which

$$p-1 = m_1 n_1, \quad p+1 = m_2 n_2, \quad \gcd(m_1, m_2) = 2, \quad (2.26)$$

every prime factor of $m_1 m_2$ is at most y , and every prime factor of $n_1 n_2$ is greater than y . In particular, $\gcd(m_1, n_1) = \gcd(m_2, n_2) = 1$.

We claim that

$$\frac{\varphi(m_1)}{m_1} \neq \frac{\varphi(m_2)}{m_2} \quad (2.27)$$

for sufficiently large x . In light of (2.7), it follows that $P(n)$ is the largest prime factor of the denominator of $\varphi(n)/n$. If $\varphi(m_1)/m_1 = \varphi(m_2)/m_2$, then $P(m_1) = P(m_2) = 2$ since $\gcd(m_1, m_2) = 2$. Thus, m_1 and m_2 are powers of 2 and

$$\begin{aligned} 1 &= \omega(m_1) + \omega(m_2) - 1 \\ &= \omega(m_1 m_2) = \omega(D_y(p^2 - 1)) \\ &= \omega_y(p^2 - 1) \in [1.5 \log \log y, 2.5 \log \log y] \end{aligned}$$

because $p \notin \mathcal{E}_3(x)$. This is a contradiction for $x \geq 10^{483}$.

For convenient $p \leq x$, we have

$$S(p) = \frac{\varphi(p-1)}{p-1} - \frac{\varphi(p+1)}{p+1} = \frac{\varphi(m_1)\varphi(n_1)}{m_1 n_1} - \frac{\varphi(m_2)\varphi(n_2)}{m_2 n_2}.$$

We note that $S(p) \neq 0$ because otherwise $P(p-1) = P(p+1)$ by (2.7). Since $\gcd(p-1, p+1) = 2$, it would follow that $p-1$ and $p+1$ are powers of 2, which occurs only for $p = 3$. Lemma 2.8 ensures that $S(p)$ has the same sign as $\varphi(p-1) - \varphi(p+1)$ on a set of full density in the primes.

Since $p \notin \mathcal{E}_2(x)$, for large x we may use the inequality

$$|t + \log(1 - t)| \leq |t|^2, \quad \text{for } |t| \leq \frac{1}{2},$$

to obtain

$$\begin{aligned} \frac{\varphi(n_1)}{n_1} &= \prod_{\substack{r|(p-1) \\ r > y}} \left(1 - \frac{1}{r}\right) = \exp\left(\sum_{\substack{r|(p-1) \\ r > y}} \log\left(1 - \frac{1}{r}\right)\right) \\ &= \exp\left(-\sum_{\substack{r|(p-1) \\ r > y}} \frac{1}{r} + O\left(\left(\sum_{\substack{r|(p-1) \\ r > y}} \frac{1}{r}\right)^2\right)\right) \\ &= \exp\left(-h_y(p-1) + O(h_y(p-1)^2)\right) \\ &= 1 + O\left(\frac{1}{y\sqrt{\log y}}\right), \end{aligned}$$

in which the implied constant in the preceding can be taken to be 2. A similar inequality holds if n_1 is replaced by n_2 . Consequently,

$$\begin{aligned} S(p) &= \frac{\varphi(m_1)}{m_1} \left(1 + O\left(\frac{1}{y\sqrt{\log y}}\right)\right) - \frac{\varphi(m_2)}{m_2} \left(1 + O\left(\frac{1}{y\sqrt{\log y}}\right)\right) \\ &= \frac{\varphi(m_1)}{m_1} - \frac{\varphi(m_2)}{m_2} + O\left(\frac{1}{y\sqrt{\log y}}\right), \end{aligned} \tag{2.28}$$

in which the implied constant can be taken to be 4.

2.9. Weird primes. A convenient prime $p \leq x$ is *weird* if

$$S(p) \left(\frac{\varphi(m_1)}{m_1} - \frac{\varphi(m_2)}{m_2}\right) < 0;$$

that is, if $S(p)$ and $\varphi(m_1)/m_1 - \varphi(m_2)/m_2$ have opposite signs (the second factor is nonzero if x is large; see (2.27)). If this occurs, then (2.28) tells us that

$$\left|\frac{\varphi(m_1)}{m_1} - \frac{\varphi(m_2)}{m_2}\right| < \frac{4}{y\sqrt{\log y}}. \tag{2.29}$$

In general, one expects the sign of $S(p)$ to be determined by small primes; that is, those primes at most y . If p is weird, then the primes $q > y$ that divide $p^2 - 1$ conspire to overthrow the contribution of the small primes.

We say that a pair (m_1, m_2) of positive integers is *weird* if

$$24|m_1m_2, \quad m_1m_2|L_y, \quad \gcd(m_1, m_2) = 2, \quad \text{and} \quad (2.29) \text{ holds.}$$

What is the reason for the appearance of the number 24 in the preceding? If $p \geq 5$, then considering $p^2 - 1$ modulo 3 and 8 reveals that $24|(p^2 - 1)$. If $x \geq 10^9$, then $y \geq 3$ and hence $P(24) = 3$ and $24|D_y(p^2 - 1)$. Consequently, if we are searching for primes p for which $D_y(p^2 - 1) = m_1m_2$, it makes sense for us to insist that m_1m_2 is divisible by 24.

Lemma 2.30. *Let $y \geq \exp(48^6)$ and $D|L_y$ be a multiple of 24.*

(a) *The number of pairs (m_1, m_2) with $D = m_1m_2$ and $\gcd(m_1, m_2) = 2$ is $2^{\omega(D)}$.*

(b) If $\omega(D) \in [1.5 \log \log y, 2.5 \log \log y]$ and $D/\varphi(D) \leq (\log y)^{1/3}$, then the number of weird pairs (m_1, m_2) with $D = m_1 m_2$ is

$$\ll \frac{2^{\omega(D)}}{\sqrt{\log \log y}}.$$

Proof. (a) In what follows, ν_p denotes the p -adic valuation function. Write

$$D = \prod_{q \in S} q^{\nu_q(D)},$$

in which S is a set of primes that contains 2, $\#S = \omega(D)$, and $\nu(2) \geq 3$. Since $\gcd(m_1, m_2) = 2$, it follows that

$$\nu_q(m_1) = \begin{cases} 1 \text{ or } \nu_2(D) - 1 & \text{if } q = 2, \\ 0 \text{ or } \nu_q(D) & \text{if } q \geq 3. \end{cases}$$

For each of the $\omega(D)$ primes in S , there are two possible choices for $\nu_q(m_1)$. Consequently, there are $2^{\omega(D)}$ possible pairs (m_1, m_2) .

(b) Let

$$(m_1, m_2) = (2n_1, 2^{\nu_2(D)-1}n_2) \quad \text{or} \quad (2^{\nu_2(D)-1}n_1, 2n_2)$$

and

$$(m'_1, m'_2) = (2n'_1, 2^{\nu_2(D)-1}n'_2) \quad \text{or} \quad (2^{\nu_2(D)-1}n'_1, 2n'_2)$$

be weird, where n_1, n_2, n'_1, n'_2 are odd, and let $D = m_1 m_2 = m'_1 m'_2$. Suppose toward a contradiction that $n_1 | n'_1$ and $n_1 < n'_1$. Then (2.29) says that

$$\begin{aligned} \left(\frac{\varphi(n_1)}{n_1}\right)^2 &= \frac{4\varphi(m_1)^2}{m_1^2} = \frac{4\varphi(m_1)}{m_1} \left(\frac{\varphi(m_2)}{m_2} + O\left(\frac{1}{y\sqrt{\log y}}\right)\right) \\ &= \frac{4\varphi(m_1)\varphi(m_2)}{m_1 m_2} + O\left(\frac{1}{y\sqrt{\log y}}\right) \end{aligned} \tag{2.31}$$

$$\begin{aligned} &= \frac{2\varphi(m_1 m_2)}{m_1 m_2} + O\left(\frac{1}{y\sqrt{\log y}}\right) \\ &= \frac{2\varphi(D)}{D} + O\left(\frac{1}{y\sqrt{\log y}}\right) \\ &= \frac{2\varphi(D)}{D} \left(1 + O\left(\frac{1}{y(\log y)^{1/6}}\right)\right) \end{aligned} \tag{2.32}$$

since $D/\varphi(D) \leq (\log y)^{1/3}$. The implied constant in (2.32) is 16, in light of (2.29) and the absorption of $4\varphi(m_1)/m_1$ in (2.31). Similar reasoning yields an analogous expression for $\varphi(n'_1)/n'_1$, with the same implied constant.

Let r be the smallest prime divisor of n'_1/n_1 . Use the inequality

$$\frac{1+s}{1+t} \leq 1 + \frac{3}{2}|s-t| \leq 1 + \frac{3}{2}(|s|+|t|), \quad |s|, |t| \leq \frac{1}{3},$$

and the fact that (2.32) holds for n_1 and n'_1 to deduce that

$$\begin{aligned} 1 + \frac{1}{r} &< 1 + \frac{2}{r-1} \leq \left(1 + \frac{1}{r-1}\right)^2 \\ &= \frac{1}{(1-1/r)^2} = \left(\frac{n'_1}{\varphi(n'_1)} \cdot \frac{\varphi(n_1)}{n_1}\right)^2 \end{aligned}$$

$$\leq 1 + \frac{\frac{3}{2}(16+16)}{y(\log y)^{1/6}} = 1 + \frac{48}{y(\log y)^{1/6}}.$$

Since $r|D$ and $D|L_y$, we have $r \leq y$ and hence

$$\frac{y(\log y)^{1/6}}{48} < r \leq y.$$

This is a contradiction if $y \geq \exp(48^6)$.

Hence, the set of odd components n_1 of the parts m_1 as (m_1, m_2) ranges over weird pairs has the property that no two divide each other. Identifying n_1 with the set of its odd prime factors, no two n_1 and n'_1 , as subsets, are contained one in another. Sperner's theorem¹ from combinatorics and Stirling's formula ensure that the number of such n_1 , and hence the number of such pairs (m_1, m_2) , is

$$\leq \binom{\omega(D)}{\lfloor \frac{\omega(D)}{2} \rfloor} \ll \frac{2^{\omega(D)}}{\sqrt{\omega(D)}} \ll \frac{2^{\omega(D)}}{\sqrt{\log \log y}}. \quad \square$$

□

2.10. Conclusion. We have shown that the number of inconvenient primes at most x is $o(\pi(x))$ and hence they can be safely ignored. Each convenient prime p gives rise to a pair (m_1, m_2) as in (2.25).

Suppose that x is large. Let D be a multiple of 24 with

$$D|L_y, \quad \frac{D}{\varphi(D)} \leq (\log y)^{1/3}, \quad \text{and} \quad \omega(D) \in [1.5 \log \log y, 2.5 \log \log y]. \quad (2.33)$$

We wish to count the primes $p \leq x$ for which $D_y(p^2 - 1) = D$. Denote this number by $\pi_D(x)$. To complete the proof of Theorem 1.1 in the case $\ell = 1$, it suffices to show that $(1/2 + o(1))\pi_D(x)$ of primes at most x have $S(p) > 0$ and $(1/2 + o(1))\pi_D(x)$ have $S(p) < 0$, and that the implied constant is uniform for all D as above.

Choose a pair (m_1, m_2) such that $D = m_1 m_2$ and $\gcd(m_1, m_2) = 2$. We want to count the primes $p \leq x$ such that $m_1|(p-1)$ and $m_2|(p+1)$; that is, such that

$$p \equiv 1 \pmod{m_1} \quad \text{and} \quad p \equiv -1 \pmod{m_2}. \quad (2.34)$$

Apply the Chinese Remainder Theorem with moduli $\frac{1}{2}m_1, m_2$ or $m_1, \frac{1}{2}m_2$, depending upon which of m_1, m_2 is exactly divisible by 2, to see that p belongs to an arithmetic progression $a_{m_1, m_2} \pmod{D/2}$, with $\gcd(a_{m_1, m_2}, D/2) = 1$. However, we also want

$$\gcd\left(\frac{p^2 - 1}{m_1 m_2}, L_y\right) = 1.$$

For this, we need to work modulo

$$M_D := (D/2) \prod_{r \leq y} r.$$

To ensure that (2.34) holds, we do the following:

- If $r|D$, then we then want

$$p \equiv \varepsilon + r^{\nu_r(D/2)} \lambda \pmod{r^{\nu_r(D/2)+1}}$$

¹A collection of sets that does not contain X and Y for which $X \subseteq Y$ is a *Sperner family*. If S is a Sperner family whose union has a total of n elements, then $\#S \leq \binom{n}{\lfloor n/2 \rfloor}$ [12].

for some $\lambda \in \{1, 2, \dots, r-1\}$. Here, $\varepsilon = \pm 1$ according to whether $r^{\nu_r(D/2)}$ divides m_1 or m_2 , respectively. For each $r|D$, there are $r-1$ possibilities for λ and hence there are $r-1$ possibilities for p modulo $r^{\nu_r(D/2)+1}$.

- If $r \nmid D$, then we want $p \equiv \lambda \pmod{r}$ for some $\lambda \notin \{0, 1, r-1\}$. For each $r \nmid D$, there are $r-3$ possibilities for p modulo r .

Thus, the number of progressions modulo M_D that can contain a prime p for which (2.34) occurs is

$$\prod_{r|\frac{D}{2}}(r-1) \prod_{\substack{r \leq y \\ r \nmid D}}(r-3). \quad (2.35)$$

By (2.12), the common modulus of all these progressions satisfies

$$M_D \leq L_y^2 \leq (\log x)^4$$

for large x . The Siegel–Walfisz theorem says that the number of primes in each progression is asymptotically

$$\frac{\pi(x)}{\varphi(M_D)} + O\left(xe^{-C\sqrt{\log x}}\right) \quad (2.36)$$

for some $C > 0$. Summing up over the number of progressions (or, more precisely, multiplying the (2.36) by the number of acceptable progressions (2.35)), and using the fact that $\varphi(M_D) = (D/2) \prod_{r \leq y} (r-1)$, we get a count of

$$\frac{2\pi(x)}{D} \prod_{\substack{r \leq y \\ r \nmid D}} \left(\frac{r-3}{r-1}\right) + O\left(xL_y^2 e^{-C\sqrt{\log x}}\right).$$

The count depends on D but not on the pair of divisors (m_1, m_2) of D . We now apply Lemma 2.30 and obtain

$$\pi_D(x) = \frac{2^{\omega(D)+1}\pi(x)}{D} \prod_{\substack{r \leq y \\ r \nmid D}} \left(\frac{r-3}{r-1}\right) + O(2^{\omega(D)} xL_y^2 e^{-C\sqrt{\log x}}). \quad (2.37)$$

Although it is not crucial to our proof, we show in Subsection 2.11 that

$$\sum_D \pi_D(x) = \pi(x) \left(1 + O\left(\frac{1}{\sqrt{\log \log y}}\right)\right), \quad (2.38)$$

where the index D runs over all D for which (2.33) holds, because the computation is of independent interest.

The product in (2.37) is less than 1 and bounded below by

$$\begin{aligned} \prod_{\substack{r \leq y \\ r \nmid D}} \left(\frac{r-3}{r-1}\right) &\gg \prod_{r \leq y} \left(\frac{r-3}{r-1}\right) = \prod_{r \leq y} \left(\frac{r^2(r-3)}{(r-1)^3}\right) \left(1 - \frac{1}{r}\right)^2 \\ &= \prod_{r \leq y} \left(1 - \frac{3r-1}{(r-1)^3}\right) \prod_{r \leq y} \left(1 - \frac{1}{r}\right)^2 \\ &\gg \prod_{r \leq y} \left(1 - \frac{1}{r}\right)^2 \gg (\log y)^{-2} \end{aligned} \quad (2.39)$$

by Mertens' asymptotic formula [10, §VII.29.1b, p. 259]. Since

$$2^{\omega(D)} \geq 2^{1.5 \log \log y} \gg (\log y),$$

we examine the main term in (2.37) and conclude that

$$\pi_D(x) \gg \frac{\pi(x)}{D(\log y)^2}.$$

On the other hand, the error term in (2.37) is

$$\begin{aligned} O(2^{\omega(D)} x L_y^2 e^{-C\sqrt{\log x}}) &= O(2^{2.5 \log \log y} x (\log x)^4 e^{-C\sqrt{\log x}}) \\ &= O((\log x)^4 (\log y)^2 x e^{-C\sqrt{\log x}}) \\ &= o\left(\frac{\pi(x)}{D(\log y)^3}\right) = o\left(\frac{\pi_D(x)}{\log y}\right). \end{aligned}$$

There is a symmetry between non-weird pairs (m_1, m_2) with

$$\frac{\varphi(m_1)}{m_1} - \frac{\varphi(m_2)}{m_2} > 0 \quad \text{and those with} \quad \frac{\varphi(m_1)}{m_1} - \frac{\varphi(m_2)}{m_2} < 0$$

given by the transposition $(m_1, m_2) \mapsto (m_2, m_1)$. Indeed, we could return to (2.26) and insist that $m_1 | (p+1)$ and $m_2 | (p-1)$ instead. The subsequent asymptotic estimates go through in exactly the same manner. Via this transposition, we obtain an asymptotically equal count between the convenient primes $p \leq x$ corresponding to (m_1, m_2) and the convenient primes $p \leq x$ corresponding to (m_2, m_1) . If only non-weird pairs (m_1, m_2) are taken into account, for fixed $D_y(p^2 - 1) = D$ this symmetry gives an asymptotically equal count of convenient primes $p \leq x$ with $S(p) > 0$ and with $S(p) < 0$.

Lemma 2.30 ensures that the number of weird pairs (m_1, m_2) with $D = m_1 m_2$ is $\ll 2^{\omega(D)} / \sqrt{\log \log y}$. As $x \rightarrow \infty$, we see from (2.37) that the number of primes at most x that arise from some weird pair is

$$\ll \frac{2^{\omega(D)} \pi(x)}{D \sqrt{\log \log y}} \prod_{\substack{r \leq y \\ r \nmid D}} \left(\frac{r-3}{r-1} \right) = O\left(\frac{\pi_D(x)}{\sqrt{\log \log y}} \right) = o(\pi_D(x)).$$

Recall that the non-weird, convenient primes have full density in the set of primes. Of such primes $p \leq x$, the argument above shows that an asymptotically equal amount have $S(p) > 0$ versus $S(p) < 0$ (recall that $S(p) = 0$ only for $p = 3$). This completes the proof of Theorem 1.1 in the case $\ell = 1$. \square

2.11. Sanity check. Before extending the preceding proof to the case $\ell \geq 2$, it is helpful to perform a quick sanity check. Our goal here is to prove (2.38). In light of (2.37), it suffices to prove that

$$\sum_D \frac{2^{\omega(D)+1}}{D} \prod_{\substack{r \leq y \\ r \nmid D}} \left(\frac{r-3}{r-1} \right) = 1 + o(1), \quad (2.40)$$

in which the index D runs over all D for which (2.33) holds. In particular, (2.38) holds and the preceding product does not run over $r = 2, 3$. These developments seem remarkably fortuitous. Let us provide an independent derivation of (2.40), which will help corroborate some of the fine details in the preceding proof.

First write $D = 2^k D_1$, in which $k \geq 3$, and sum to obtain

$$\left(\sum_{k \geq 3} \frac{2^k}{2^k} \right) \left(\sum_{D_1 \text{ odd}} \frac{2^{\omega(D_1)}}{D_1} \right) = \sum_{D_1 \text{ odd}} \frac{2^{\omega(D_1)}}{D_1}.$$

Now write $D_1 = 3^k D_2$ and sum over $k \geq 1$ getting

$$\left(\sum_{k \geq 1} \frac{2}{3^k} \right) \left(\sum_{\gcd(D_2, 6)=1} \frac{2^{\omega(D_2)}}{D_2} \right) = \sum_{\gcd(D_2, 6)=1} \frac{2^{\omega(D_2)}}{D_2}.$$

For the rest, we use multiplicativity to say that the sum in (2.40) is

$$\prod_{r \leq y} \left(\frac{r-3}{r-1} \right) \prod_{5 \leq r \leq y} \left(1 + \frac{2}{(r-3)/(r-1)} \sum_{k \geq 1} \frac{1}{r^k} \right).$$

However, this is not strictly correct since the sum above stops at the largest power $r^b \leq y$. Moreover, the sum runs over all D without restrictions such as $\omega(D) \in [1.5 \log \log y, 2.5 \log \log y]$ or $D/\varphi(D) \leq (\log y)^{1/3}$. We deal with these omissions shortly. For the time being, let us ignore these restrictions. Then the amount inside the Euler factor is

$$1 + \frac{2(r-1)}{r-3} \frac{1}{r-1} = 1 + \frac{2}{r-3} = \frac{r-1}{r-3},$$

which cancels with the outside $(r-3)/(r-1)$.

Now we must examine the errors. There are essentially four types:

- (a) In each Euler factor we only sum up to r^b , in which b is maximal such that $r^b \leq y$. By extending the sum to infinity we incurred an error of

$$\frac{2}{(r-3)/(r-1)} \sum_{k \geq b+1} \frac{1}{r^k} = \frac{2}{r^b(r-3)} = O\left(\frac{1}{y}\right).$$

For $5 \leq r \leq y$, the actual Euler factor is

$$\frac{r-1}{r-3} + O\left(\frac{1}{(r-3)y}\right) = \frac{r-1}{r-3} \left(1 + O\left(\frac{1}{y}\right) \right)$$

Similar considerations apply for $r = 2, 3$. The total multiplicative error is

$$\left(1 + O\left(\frac{1}{y}\right) \right)^{\pi(y)} = 1 + O\left(\frac{\pi(y)}{y}\right) = 1 + O\left(\frac{1}{\log y}\right).$$

- (b) We consider only D such that $D/\varphi(D) \leq (\log y)^{1/3}$. Let the set of remaining D be denoted \mathcal{D} . For $D \in \mathcal{D}$, we have

$$\frac{1}{D} \leq \frac{1}{(\log y)^{1/3}} \cdot \frac{1}{\varphi(D)}.$$

Applying this inequality and extending then the sum over all possible D , the piece of the sum over \mathcal{D} is at most

$$\ll \frac{\pi(x)}{(\log y)^{1/3}} \sum_D \frac{2^{\omega(D)+1}}{\varphi(D)} \prod_{\substack{r \leq y \\ r \nmid D}} \left(\frac{r-3}{r-1} \right).$$

We separate out the power of 2 in D as $D = 2^k D_1$ with $k \geq 3$ getting an Euler factor corresponding to 2 of

$$4 \left(\frac{1}{\varphi(8)} + \frac{1}{\varphi(16)} + \cdots + \frac{1}{\varphi(2^k)} + \cdots \right) = 2.$$

Then we separate out a factor of 3 from D_1 writing it as $D_1 = 3^k D_2$, getting an Euler factor corresponding to 3 of

$$2 \left(\frac{1}{\varphi(3)} + \frac{1}{\varphi(9)} + \cdots + \frac{1}{\varphi(3^k)} + \cdots \right) = \frac{3}{2}.$$

For the remaining primes $r \geq 5$, we form the Euler product getting

$$\frac{\pi(x)}{(\log y)^{2/3}} \prod_{5 \leq r \leq y} \left(\frac{r-3}{r-1} \right) \prod_{r=5}^y \left(1 + \frac{2}{(r-3)/(r-1)} \sum_{b \geq 1} \frac{1}{\varphi(r^b)} \right).$$

The factor inside the parentheses is

$$\begin{aligned} 1 + \frac{2}{(r-3)/(r-1)} \frac{1}{(r-1)(1-1/r)} &= 1 + \frac{2r}{(r-1)(r-3)} \\ &= \frac{r^2 - 2r + 3}{(r-1)(r-3)} \\ &= \frac{(r-1)}{(r-3)} \left(1 + \frac{2}{(r-1)^2} \right). \end{aligned}$$

Multiply this by the outside factor $(r-3)/(r-1)$ and get

$$\left(\frac{r-3}{r-1} \right) \left(1 + \frac{2}{(r-3)/(r-1)} \sum_{b \geq 1} \frac{1}{\varphi(r^b)} \right) = 1 + \frac{2}{(r-1)^2}.$$

Taking the product of the factors above over $r \in [5, y]$, we get a convergent product. Consequently,

$$\pi(x) \sum_{\substack{D|L_y, 24|D \\ D/\varphi(D) > (\log y)^{1/3}}} \frac{2^{\omega(D)+1}}{D} \prod_{\substack{5 \leq r \leq y \\ r \nmid D}} \left(\frac{r-3}{r-1} \right) \ll \frac{\pi(x)}{(\log y)^{1/3}}.$$

- (c) We need to consider D with $\omega(D) \notin [1.5 \log \log y, 2.5 \log \log y]$. From the preceding material and (2.39), we have

$$\prod_{5 \leq r \leq y} \left(1 + \sum_{b \geq 1} \frac{2}{(r-3)/(r-1)} \sum_{b \geq 1} \frac{1}{r^b} \right) \asymp \left(\prod_{5 \leq r \leq y} \left(\frac{r-3}{r-1} \right) \right)^{-1} \asymp (\log y)^2.$$

We first deal with D with many prime factors. Consider the multiplicative function defined for prime powers r^b with $r \geq 5$ by

$$f(r) = \frac{2(r-1)}{(r-3)r^b}.$$

If $K := \lfloor 2.5 \log \log y \rfloor$, then

$$\sum_{\substack{r|D_2 \implies r \in [5, y] \\ \omega(D_2) > K-2}} f(D_2) \leq \sum_{k > K-2} \frac{1}{k!} \left(\sum_{\substack{r \in [5, y] \\ b \geq 1}} f(r^b) \right)^k.$$

We have

$$\begin{aligned}
S &= \sum_{\substack{r \in [5, y] \\ b \geq 1}} f(r^b) = \sum_{\substack{b \geq 1 \\ 5 \leq r \leq y}} \left(\frac{2}{r^b} + \frac{2}{r^b} \left(\frac{r-1}{r-3} - 1 \right) \right) \\
&= \sum_{\substack{r \in [2, y] \\ b \geq 1}} \frac{2}{r^b} + O\left(\sum_{\substack{r \geq 5 \\ b \geq 1}} \frac{1}{r^{b+1}} \right) \\
&= 2 \log \log y + O(1),
\end{aligned}$$

in which we have used Mertens' theorem [10, §VII.28.1b]. In the sum $\sum_{k > K} S^k/k!$, the ratio of two consecutive terms is

$$\frac{S^{k+1}/(k+1)!}{S^k/k!} = \frac{S}{k+1} = \frac{2 \log \log y + O(1)}{2.5 \log \log y + O(1)} < \frac{5}{6}$$

for $k > K - 2$ and large x , so the first term dominates. With $K! > (K/e)^K$, the contribution of D with $\omega(D) > 2.5 \log \log y$ is at most

$$\ll \frac{S^{K-2}}{(K-2)!} \leq \left(\frac{2e \log \log y + O(1)}{2.5 \log \log y + O(1)} \right)^{2.5 \log \log y + O(1)} \ll (\log y)^c,$$

in which $c = 2.5 \log(2e/2.5) < 1.95$. Multiplying this by (see (2.39))

$$\prod_{5 \leq r \leq y} \left(\frac{r-3}{r-1} \right) \ll (\log y)^{-2},$$

we obtain

$$\pi(x) \sum_{\substack{D|L_y \\ 24|D \\ \omega(D) > 2.5 \log \log y}} \frac{2^{\omega(D)+1}}{D} \prod_{\substack{5 \leq r \leq y \\ r \nmid D}} \left(\frac{r-3}{r-1} \right) \ll \frac{\pi(x)}{(\log y)^{0.5}}.$$

We use a similar argument for D with $\omega(D) < 1.5 \log \log y$. In this case, let $K_1 := \lfloor 1.5 \log \log y \rfloor$. We have to deal with

$$\sum_{\substack{D_2: r|D_2 \implies r \in [5, y] \\ \omega(D_2) < 1.5 \log \log y - 2}} f(D_2) \leq \sum_{k \leq K_1 - 2} \frac{1}{k!} S^k.$$

For $k > K_1 - 2$ and large x , the ratio of any two consecutive terms above is

$$\frac{S^{k+1}/(k+1)!}{S^k/k!} = \frac{S}{k+1} > \frac{2 \log \log y + O(1)}{1.5 \log \log y + O(1)} > \frac{5}{4},$$

it follows that the last term dominates. Thus, this sum is at most

$$\left(\frac{2e \log \log y + O(1)}{K_1 - 2} \right)^{K_1 - 2} = \left(\frac{2e \log \log y + O(1)}{1.5 \log \log y + O(1)} \right)^{1.5 \log \log y + O(1)} \ll (\log y)^{c_1},$$

in which where $c_1 = 1.5 \log(2e/1.5) < 1.95$. Consequently, the contribution of D with $\omega(D) < 1.5 \log \log y$ to the sum defining $\pi_D(x)$ is

$$\ll \frac{\pi_D(x)}{(\log y)^{0.05}}.$$

Putting everything together we obtain (2.40), which is equivalent to (2.38).

3. PROOF OF THEOREM 1.1 FOR $\ell \geq 2$

The proof for $\ell \geq 2$ follows largely on the lines of the case $\ell = 1$, although there are a number of minor adjustments that must be made. For example, in Lemma 2.30 we assumed that D is a multiple of 24. Elementary considerations reveal that the following adjustments are necessary for various values of ℓ :

- (a) D is coprime to all primes that divide ℓ .
- (b) D is odd if ℓ is even.
- (c) D is a multiple of 8 if ℓ is odd.
- (d) D is a multiple of 3 if and only if ℓ is not.

More significant modifications are discussed below.

3.1. The case of equality. We need a variant of the inequality (2.4). The estimate provided by the following theorem involves two special cases. Numerical evidence strongly suggests that this distinction is not simply a byproduct of our proof; see Table 1. If we replace the use of the Brun sieve in what follows with an appeal to the Bateman–Horn conjecture [1], then the larger of the two upper bounds becomes an asymptotic equivalence if the appropriate constant factor is introduced.

Theorem 3.1. *For each $\ell \geq 1$,*

$$\#\{p \leq x : \varphi(p - \ell) = \varphi(p + \ell)\} \ll \begin{cases} \frac{x}{(\log x)^3} & \text{if } \ell = 4^n - 1, \\ \frac{x}{e^{(\log x)^{1/3}}} & \text{otherwise.} \end{cases}$$

Proof. In Lemma 2.1, let $k = 2\ell$ and $n = p - \ell$, in which p is prime. Suppose that j and $j + 2\ell$ have the same prime factors. Since

$$p = j \left(\frac{(j + 2\ell)t}{g} + 1 \right) + \ell, \quad (3.1)$$

it follows that j is not divisible by any prime factor of ℓ and hence $g = \gcd(j, 2\ell) = 2$. Thus, $j = 2^m$ and $j + 2\ell = 2^{m+n}$ for some $m, n \geq 1$. Then $2^{m+n} = 2^m + 2\ell$ and

$$\ell = 2^{m-1}(2^n - 1).$$

If $m \geq 2$, then ℓ is even and (3.1) implies that $2|p$, a contradiction. Thus, the upper bound from Lemma 2.1 applies in this case.

If $m = 1$, then $j = 2$ and $\ell = 2^n - 1$. Then

$$p = 2(\ell + 1)t + (\ell + 2), \quad q = t + 1 \quad \text{and} \quad r = (\ell + 1)t + 1$$

and we count $t \leq \frac{x - (\ell + 2)}{2(\ell + 1)} \sim \frac{x}{2\ell}$ for which p, q, r are simultaneously prime. Let

$$f_1(t) = 2^{n+1}t + (2^n + 1), \quad f_2(t) = t + 1, \quad \text{and} \quad f_3(t) = 2^n t + 1.$$

If n is odd, then

$$f_1(0) \equiv f_2(2) \equiv f_3(1) \equiv 0 \pmod{3}.$$

There are three possibilities:

- (a) If $f_1(0) = 3$, then $f_2(0) = 1$ is not prime and no prime triples are produced.
- (b) If $f_2(2) = 3$, then for each odd n , at most one prime triple is produced.²

²The only odd $n < 99$ for which a prime triple arises in this manner are $n = 1, 3, 7, 15$, from which we obtain the triples $(11, 3, 5)$, $(41, 3, 17)$, $(641, 3, 257)$, and $(163841, 3, 65537)$.

(c) If $f_3(1) = 3$, then $n = 1$ and only the prime triple $(7, 2, 3)$ is produced.

In each case, the upper bound from Lemma 2.1 dominates.

If n is even, then none of f_1, f_2, f_3 vanish identically modulo any prime. The Brun sieve says that the number of $p \leq x$ for which p, q, r are prime is $O(x/(\log x)^3)$, which dominates the estimate from Lemma 2.1. \square

3.2. A more general comparison lemma. The next adjustment that is required is an analogue of the comparison lemma (Lemma 2.8). This turns out to be more involved than expected. In fact, we first need a generalization of the ‘‘Turán–Kubilius’’-type result from Lemma 2.21. Since this is a minor variant of an existing result, we only sketch the proof.

Lemma 3.2. *For each $\ell \geq 1$, $\sum_{p \leq x} (\omega_y(p \pm \ell) - \log \log y)^2 = O(\pi(x) \log \log y)$.*

Proof. Since $y \leq (\log x)^2$, apply the Siegel–Walfisz theorem to obtain

$$\begin{aligned} \sum_{p \leq x} \omega_y(p \pm \ell) &= \pi(x) \log \log y + O(\pi(x)), \\ \sum_{p \leq x} \omega_y(p \pm \ell)^2 &= \pi(x) (\log \log y)^2 + O(\pi(x) \log \log y), \end{aligned}$$

in which the $\log \log y$ term arises from an application of Mertens’ theorem [10, §VII.28.1b]. Now expand $\sum_{p \leq x} (\omega_y(p \pm \ell) - \log \log y)^2$ and apply the preceding. \square

The direct generalization of Lemma 2.8 for $\ell \geq 2$ runs into trouble. If ℓ is sufficiently large, then the $+1$ in (2.10) becomes too large for the same argument to work. The evenness of $\varphi(p - \ell) - \varphi(p + \ell)$ is no longer sufficient to push the argument through. Fortunately, we are able to employ the following lemma instead.

Lemma 3.3. *For $\ell, m \geq 1$,*

$$\#\{p \leq x : \varphi(p \pm \ell) \equiv 0 \pmod{2^m}\} \sim \pi(x).$$

Proof. Fix $\ell \geq 1$. Let $\omega(n)$ denote the number of distinct prime divisors of n . Then $2^{\omega(n)-1} | \varphi(n)$ since $\varphi(n) = \prod_{p^a || n} p^{a-1}(p-1)$. If

$$2 + \log_2 \ell \leq m + 1 \leq \omega(p \pm \ell), \tag{3.4}$$

then

$$2\ell \leq 2^m \leq 2^{\omega(p \pm \ell)-1} \quad \text{and hence} \quad \varphi(p \pm \ell) \equiv 0 \pmod{2^m}.$$

Thus, it suffices to show that the set of primes $p \leq x$ for which (3.4) fails has a counting function that is $o(\pi(x))$. Let x be so large that $y = \log \log x$ satisfies

$$2 + \log_2 \ell \leq \frac{1}{2} \log \log y$$

and let

$$\mathcal{E}(x) = \{p \leq x : \omega(p - \ell) < 2 + \log_2 \ell \quad \text{or} \quad \omega(p + \ell) < 2 + \log_2 \ell\}$$

Let $\omega_y(n)$ denote the number of distinct prime factors $q \leq y$ of n . If $p \in \mathcal{E}(x)$, then

$$\omega_y(p - \ell) < \frac{1}{2} \log \log y \quad \text{or} \quad \omega_y(p + \ell) < \frac{1}{2} \log \log y,$$

and hence

$$\omega_y(p - \ell) \notin \left[\frac{1}{2} \log \log y, \frac{3}{2} \log \log y\right] \quad \text{or} \quad \omega_y(p + \ell) \notin \left[\frac{1}{2} \log \log y, \frac{3}{2} \log \log y\right].$$

Then

$$\frac{1}{4}(\log \log y)^2 \leq (\omega_y(p-\ell) - \log \log y)^2 \quad \text{or} \quad \frac{1}{4}(\log \log y)^2 \leq (\omega_y(p+\ell) - \log \log y)^2.$$

Lemma 3.2 ensures that

$$\begin{aligned} \frac{1}{4}(\log \log y)^2 \#\mathcal{E}(x) &\leq \sum_{p \in \mathcal{E}(x)} (\omega_y(p-\ell) - \log \log y)^2 + (\omega_y(p+\ell) - \log \log y)^2 \\ &\leq \sum_{p \leq x} (\omega_y(p-\ell) - \log \log y)^2 + (\omega_y(p+\ell) - \log \log y)^2 \\ &= O(\pi(x) \log \log y). \end{aligned}$$

Thus,

$$\#\mathcal{E}(x) \ll \frac{\pi(x)}{\log \log y} = o(\pi(x)). \quad \square$$

□

Our replacement for the comparison lemma is the following. Since the exceptional set is $o(\pi(x))$, it will not affect the proof of Theorem 1.1 in the case $\ell \geq 2$.

Lemma 3.5. *For each $\ell \geq 1$, the set of primes p for which $\varphi(p-\ell) - \varphi(p+\ell)$ and*

$$S(p) := \frac{\varphi(p-\ell)}{p-\ell} - \frac{\varphi(p+\ell)}{p+\ell} \tag{3.6}$$

have the same sign has counting function $\sim \pi(x)$.

Proof. By Theorem 3.1, it suffices to show that the set of primes p for which

$$\varphi(p-\ell) > \varphi(p+\ell) \quad \iff \quad \frac{\varphi(p-\ell)}{p-\ell} > \frac{\varphi(p+\ell)}{p+\ell} \tag{3.7}$$

has counting function $\sim \pi(x)$. The forward implication is straightforward, so we focus on the reverse. If the inequality on the right-hand side of (3.7) holds, then

$$\begin{aligned} 0 &< (p+\ell)\varphi(p-\ell) - (p-\ell)\varphi(p+\ell) \\ &= p[\varphi(p-\ell) - \varphi(p+\ell)] + \ell\varphi(p-\ell) + \ell\varphi(p+\ell) \\ &\leq p[\varphi(p-\ell) - \varphi(p+\ell)] + \ell(p-\ell) + \ell(p+\ell) \\ &= p[\varphi(p-\ell) - \varphi(p+\ell) + 2\ell]. \end{aligned}$$

Let $2^m > 2\ell$ and apply lemma Lemma 3.3 to conclude that $\varphi(p-\ell) - \varphi(p+\ell) > 0$ for p in a set with counting function $\sim \pi(x)$. □

3.3. Final ingredient. The only other ingredient necessary to consider $\ell \geq 2$ is a replacement for the estimate (2.24). We include the proof for completeness.

Lemma 3.8. $\sum_{p \leq x} \frac{p \pm \ell}{\varphi(p \pm \ell)} \ll \pi(x)$.

Proof. Let $\sigma(n)$ denote the sum of the divisors of n . Then

$$\frac{\sigma(n)}{n} = \sum_{d|n} \frac{1}{d} \quad \text{and} \quad \frac{6}{\pi^2} < \frac{\sigma(n)\varphi(n)}{n^2}, \quad \text{for } n \geq 2;$$

see [10, §I.3.5]. The Siegel–Walfisz theorem provides $C > 0$ so that

$$\begin{aligned}
\sum_{p \leq x} \frac{p \pm \ell}{\varphi(p \pm \ell)} &\ll \sum_{p \leq x} \frac{\sigma(p \pm \ell)}{(p \pm \ell)} = \sum_{p \leq x} \sum_{d|(p \pm \ell)} \frac{1}{d} = \sum_{d \leq x} \frac{1}{d} \sum_{\substack{p \leq x \\ p \equiv \mp 1 \pmod{d}}} 1 = \sum_{d \leq x} \frac{\pi(x; \mp 1, d)}{d} \\
&= \sum_{d \leq (\log x)^3} \frac{\pi(x; \mp 1, d)}{d} + \sum_{(\log x)^3 \leq d \leq x} \frac{\pi(x; \mp 1, d)}{d} \\
&\leq \sum_{d \leq (\log x)^3} \left(\frac{\pi(x)}{d\varphi(d)} + O\left(xe^{-C\sqrt{\log x}}\right) \right) + \sum_{(\log x)^3 \leq d \leq x} \frac{x}{d^2} \\
&\leq \pi(x) \sum_{1 \leq d < \infty} \frac{1}{d\varphi(d)} + O\left(x(\log x)^3 e^{-C\sqrt{\log x}}\right) + x \sum_{(\log x)^3 \leq d < \infty} \frac{1}{d^2} \\
&\ll \pi(x) + x(\log x)^3 e^{-C\sqrt{\log x}} + \frac{x}{(\log x)^3} \\
&\ll \pi(x). \quad \square
\end{aligned}$$

□

Acknowledgment. We thank the referee for suggestions which improved the quality of our paper.

REFERENCES

1. Paul T. Bateman and Roger A. Horn, *A heuristic asymptotic formula concerning the distribution of prime numbers*, Math. Comp. **16** (1962), 363–367. MR 0148632
2. P. Erdős, *On the normal number of prime factors of $p - 1$ and some related problems concerning Euler’s φ -function*, Quarterly J. Math. **6** (1935), no. 1, 205–213.
3. P. Erdős, C. Pomerance, and A. Sárközy, *On locally repeated values of certain arithmetic functions. II*, Acta Math. Hungar. **49** (1987), no. 1-2, 251–259. MR 869681
4. Stephan Ramon Garcia, Elvis Kahoro, and Florian Luca, *Primitive root discrepancy for twin primes*, Exp. Math., (accepted). <http://arxiv.org/abs/1705.02485>.
5. S. W. Golomb, *Powerful numbers*, Amer. Math. Monthly **77** (1970), 848–855. MR 0266878
6. S. W. Graham, Jeffrey J. Holt, and Carl Pomerance, *On the solutions to $\varphi(n) = \varphi(n + k)$* , Number theory in progress, Vol. 2 (Zakopane-Kościelisko, 1997), de Gruyter, Berlin, 1999, pp. 867–882. MR 1689549
7. Florian Luca and Carl Pomerance, *On some problems of Mqkowski-Schinzel and Erdős concerning the arithmetical functions φ and σ* , Colloq. Math. **92** (2002), no. 1, 111–130. MR 1899242
8. Yoichi Motohashi, *On a property of $p - 1$* , Proc. Japan Acad. **45** (1969), 639–640. MR 0266880
9. Karl Prachar, *Primzahlverteilung*, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1957. MR 0087685
10. József Sándor, Dragoslav S. Mitrinović, and Borislav Crstici, *Handbook of number theory. I*, Springer, Dordrecht, 2006, Second printing of the 1996 original. MR 2186914
11. Harold N. Shapiro, *Addition of functions in probabilistic number theory*, Comm. Pure Appl. Math. **26** (1973), 55–84. MR 0366861
12. Emanuel Sperner, *Ein Satz über Untermengen einer endlichen Menge*, Math. Z. **27** (1928), no. 1, 544–548. MR 1544925
13. Gérald Tenenbaum, *Introduction to analytic and probabilistic number theory*, third ed., Graduate Studies in Mathematics, vol. 163, American Mathematical Society, Providence, RI, 2015, Translated from the 2008 French edition by Patrick D. F. Ion. MR 3363366
14. T. Yamada, *On equations $\sigma(n) = \sigma(n + k)$ and $\varphi(n) = \varphi(n + k)^*$* , <http://arxiv.org/abs/1001.2511>.

DEPARTMENT OF MATHEMATICS, POMONA COLLEGE, 610 N. COLLEGE AVE., CLAREMONT, CA
91711

E-mail address: `stephan.garcia@pomona.edu`

URL: `http://pages.pomona.edu/~sg064747`

SCHOOL OF MATHEMATICS, UNIVERSITY OF THE WITWATERSRAND, PRIVATE BAG 3, WITS 2050,
JOHANNESBURG, SOUTH AFRICA, MAX PLANCK INSTITUTE FOR MATHEMATICS, VIVATSGASSE 7,
53111 BONN, GERMANY, DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCES, UNIVERSITY OF
OSTRAVA, 30 DUBNA 22, 701 03 OSTRAVA 1, CZECH REPUBLIC

E-mail address: `Florian.Luca@wits.ac.za`