The number of non-cyclic Sylow subgroups of the multiplicative group modulo \( n \)

Paul Pollack

Abstract. For each positive integer \( n \), let \( U(\mathbb{Z}/n\mathbb{Z}) \) denote the group of units modulo \( n \), which has order \( \varphi(n) \) (Euler’s function) and exponent \( \lambda(n) \) (Carmichael’s function). The ratio \( \varphi(n)/\lambda(n) \) is always an integer, and a prime \( p \) divides this ratio precisely when the (unique) Sylow \( p \)-subgroup of \( U(\mathbb{Z}/n\mathbb{Z}) \) is noncyclic. Write \( W(n) \) for the number of such primes \( p \). Banks, Luca, and Shparlinski showed that for certain constants \( C_1, C_2 > 0 \),

\[
C_1 \frac{\log \log n}{(\log \log \log n)^2} \leq W(n) \leq C_2 \log \log n
\]

for all \( n \) from a sequence of asymptotic density 1. We sharpen their result by showing that \( W(n) \) has normal order \( \log \log n / \log \log \log n \).

1. Introduction

For a finite abelian group \( G \), we write \( \lambda(G) \) for the exponent of \( G \), meaning the order of the largest cyclic subgroup of \( G \). Then \( \lambda(G) \) divides \( \#G \), and the primes \( p \) dividing the ratio \( \frac{\#G}{\lambda(G)} \) are precisely those for which the (unique) Sylow \( p \)-subgroup of \( G \) fails to be cyclic. In this note, we are concerned with the function \( W(n) \) counting the number of these primes \( p \) when \( G \) is the group of units modulo a positive integer \( n \). That is, \( W(n) \) is the number of distinct prime factors of \( \frac{\varphi(n)}{\lambda(n)} \), where \( \varphi(n) \) and \( \lambda(n) \) are the usual Euler and Carmichael functions.

The study of \( W(n) \) was initiated by Banks, Luca, and Shparlinski in [1]. Clearly, \( W(n) = 0 \) for infinitely many \( n \) — namely, those \( n \) for which the unit group \( U(\mathbb{Z}/n\mathbb{Z}) \) is cyclic. At the opposite extreme, Banks, Luca, and Shparlinski prove (see their Theorem 6) that \( W(n) \gg \log n / \log \log n \) for infinitely many \( n \). Since we always have \( \frac{\varphi(n)}{\lambda(n)} \leq n \), and \( \omega(m) \leq (1 + o(1)) \log m / \log \log m \) (as \( m \to \infty \)), this latter result is best possible up to the value of the implied constant.

Concerning the typical size of \( W(n) \), Banks, Luca, and Shparlinski show that on a set of \( n \) of asymptotic density 1,

\[
\frac{\log \log n}{(\log \log \log n)^2} \ll W(n) \ll \log \log n.
\]

We leverage ideas from recent joint work with Pomerance [13] to establish the following improvement.

Theorem 1. \( W(n) \) has normal order \( \log \log n / \log \log \log n \). That is, for each fixed \( \epsilon > 0 \), the set of \( n \) with

\[
|W(n) - \frac{\log \log n}{\log \log \log n}| < \epsilon \frac{\log \log n}{\log \log \log n}
\]

has asymptotic density 1.
One consequence of Theorem 1 is that $\frac{\phi(n)}{\lambda(n)}$ typically has many more distinct prime factors than a number of comparable size, although not quite as many as allowed by the maximal order of $\omega(m)$. Indeed, from work of Erdős, Pomerance, and Schmutz (see [6, Theorem 2]), there is a constant $A \approx 0.227$ such that

$$\frac{\phi(n)}{\lambda(n)} = \exp(\log \log \log n \cdot \log \log n + (A + o(1)) \log \log n),$$

as $n \to \infty$ along a set of density 1. So the typical size of $\omega(m)$, for a number $m$ near $\phi(n)/\lambda(n)$, is $\sim \log m \sim \log \log \log n$, while the maximal size of $\omega(m)$ is $\sim \frac{\log m}{\log \log m} \sim \log \log n$. In comparison, Theorem 1 implies that $m = \frac{\phi(n)}{\lambda(n)}$ itself has $\omega(m) \sim \frac{\log m}{(\log \log m)^2}$, as $n \to \infty$ through a set of density 1.

Theorem 1 might be compared with existing counting results for subgroups of $U(\mathbb{Z}/n\mathbb{Z})$. Erdős and Pomerance [5] (see §6 of [4] for minor corrections) and Murty and Murty [11], independently, considered the total number of Sylow subgroups of $U(\mathbb{Z}/n\mathbb{Z})$ (equivalently, the number of distinct prime factors of $\phi(n)$). They showed that this quantity has normal order $\frac{1}{2}(\log \log n)^2$. Very recently, Martin and Troupe [9] considered the total number of subgroups of $U(\mathbb{Z}/n\mathbb{Z})$, both up to isomorphism and otherwise (i.e., as sets). They proved that the log of the first count has normal order $\frac{\log 2}{2}(\log \log n)^2$, and that the log of the second quantity has normal order $A(\log \log n)^2$ for an explicitly described constant $A \approx 0.721$. Other statistical questions concerning the structure of the multiplicative groups are taken up in [2, 3, 8].

Notation. The letters $\ell$, $p$, and $q$ (possibly with subscripts or other decorations) are always reserved for primes. We write $\log_k$ for the $k$th iterate of the natural logarithm. We use $1_C$ for the characteristic function of the condition $C$; for example, $1_{d|n}$ takes the value 1 when $d$ divides $n$ and the value 0 otherwise.

2. Lemmata

It will be helpful to have in mind the classical structure theory of the unit group mod $n$, which goes back essentially to Gauss.

**Lemma 2.** Let $n$ be a positive integer, and write $n = \prod_p p^{v_p}$. Then $U(\mathbb{Z}/n\mathbb{Z}) \cong \prod_{p|n} U(\mathbb{Z}/p^{v_p}\mathbb{Z})$. If $p$ is odd, or if $v_p \leq 2$, then $U(\mathbb{Z}/p^{v_p}\mathbb{Z}) \cong \mathbb{Z}/\phi(p^{v_p})\mathbb{Z}$. When $p = 2$ and $v_2 \geq 3$, we have $U(\mathbb{Z}/2^{v_2}\mathbb{Z}) \cong \mathbb{Z}/2^{v_2-2}\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$.

Our chief technical tool in the proof of Theorem 1 will be the fundamental lemma of the sieve. Specifically, we will make repeated use of the following special case of Theorem 7.2 in [7].

**Proposition 3.** Let $x \geq z \geq 2$. If $\mathcal{P}$ is any set of primes not exceeding $z$, then

$$\# \{n \leq x : p \mid n \Rightarrow p \notin \mathcal{P} \} = \left( x \prod_{p \in \mathcal{P}} \left( 1 - \frac{1}{p} \right) \left( 1 + O(e^{-u/z}) \right) \right),$$

where $u = \frac{\log x}{\log z}$.

As input for the sieve, we will need the following estimate, due independently to Pomerance (see Remark 1 of [14]) and Norton (see the Lemma on p. 699 of [12]).

**Lemma 4.** Let $m$ be a positive integer, and let $x \geq 3$. Put

$$S(x;m) = \sum_{\ell \leq x, \ell \equiv 1 \pmod{m}} \frac{1}{\ell}.$$
Then
\[ S(x; m) = \frac{\log_2 x}{\phi(m)} + O\left(\frac{\log (2m)}{\phi(m)}\right). \]

### 3. Proof of Theorem 1

#### 3.1. A preliminary reduction.
Observe that if \( p \) is prime, and \( n \) is divisible by distinct primes \( q, q' \equiv 1 \pmod p \), then \( p \) is counted by \( W(n) \). Indeed, in this case
\[
\mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z} \leq \mathbb{Z}/(q - 1)\mathbb{Z} \oplus \mathbb{Z}/(q' - 1)\mathbb{Z} \leq U(\mathbb{Z}/n\mathbb{Z}).
\]

Thus, the \( p \)-Sylow subgroup of \( U(\mathbb{Z}/n\mathbb{Z}) \) is not cyclic. Conversely, if the prime \( p \) is counted by \( W(n) \), then either

(i) \( n \) is divisible by distinct primes \( q, q' \equiv 1 \pmod p \), or

(ii) \( p^2 | n \).

All of this follows from the decomposition of \( U(\mathbb{Z}/n\mathbb{Z}) \) recalled in Lemma 2.

We now let \( I = (\log_2 x/\log_3 x, \log_2 x \cdot \log_3 x) \) and set
\[
\tilde{W}(n) = \#\{p \in I : \text{there are distinct primes } q, q' | n \text{ with } q, q' \leq x^{1/2}\log_3 x, \text{ and } q, q' \equiv 1 \pmod p\}.
\]

From the last paragraph, \( \tilde{W}(n) \leq W(n) \) for all \( n \). In fact, the two sides are usually close. The next lemma makes this precise, and will allow us to work with \( \tilde{W}(n) \) rather than \( W(n) \) in our proof of Theorem 1.

**Lemma 5.** For all large \( x \),
\[
\sum_{n \leq x} (W(n) - \tilde{W}(n)) = O(x \log_2 x/(\log_3 x)^2).
\]

Thus, if \( \xi(x) \) is any function tending to infinity, then \( W(n) - \tilde{W}(n) < \xi(x) \frac{\log_2 x}{(\log_3 x)^2} \) for all but \( o(x) \) values of \( n \leq x \), as \( x \to \infty \).

**Proof.** The difference \( R(n) := W(n) - \tilde{W}(n) \) counts those primes \( p \) captured by the definition of \( W(n) \) but not by that of \( \tilde{W}(n) \). We decompose
\[
R(n) = R_0(n) + R_1(n),
\]
where the right-hand summands correspond to the conditions \( p \leq \log_2 x/\log_3 x \) and \( p > \log_2 x/\log_3 x \), respectively. For all large \( x \), we have by the prime number theorem that
\[
R_0(n) < 2\log_2 x/(\log_3 x)^2 \quad \text{for every } n \leq x.
\]

If \( p \) is counted by \( R_1(n) \), then either

- \( p^2 | n \),
- \( p \leq \log_2 x \log_3 x \) and \( q | n \) for some prime \( q \equiv 1 \pmod p \), \( q > x^{1/2}\log_3 x \), or
- \( p > \log_2 x \log_3 x \) and \( qq' | n \) for distinct primes \( q, q' \equiv 1 \pmod p \).
Thus,

\[ \sum_{n \leq x} R_1(n) \]

\[
\leq \sum_{n \leq x} \sum_{p \in \mathbb{P}, \log_2 x \leq n} \left( 1_{p^2 \mid n} + 1_{p \leq \log_3 x} \sum_{q \leq x} 1_q | n \right) + 1_{p \geq \log_3 x} \sum_{q \leq x} 1_q | n + 1_{q \geq \log x} \sum_{q \mid n, q \equiv 1 \mod p} \frac{1}{q} \\
\leq \sum_{p \geq \log_2 x} \frac{x}{p^2} + \sum_{p \in \mathcal{I}} \frac{x}{\log_2 x} \sum_{x^{1/2 \log_3 x} < q \leq x} 1_q + \frac{1}{q} \sum_{p \geq \log_2 x} \frac{x}{\log_3 x} \left( \sum_{q \equiv 1 \mod p} \frac{1}{q} \right)^2 \\
\ll \frac{x}{\log_2 x} + x \log_3 x \sum_{p \in \mathcal{I}} \frac{1}{p} + x (\log_2 x)^2 \sum_{p \geq \log_2 x} \frac{1}{p^2} \\
\ll x \frac{\log_2 x}{(\log_3 x)^2}.
\]

(We used Lemma 4 to estimate the various sums on \( q \) appearing above, and Mertens' theorem to estimate the sum of the reciprocals of those primes \( p \in \mathcal{I} \).) Collecting our estimates gives the first claim of the lemma. The second is an immediate consequence. \( \square \)

3.2. The second-moment strategy. Inspired by Turán’s simple proof of the Hardy–Ramanujan normal order theorem, we prove Theorem 1 by estimating a second moment. Specifically, we show that

\[
\sum_{n \leq x} \left( \tilde{W}(n) - \frac{\log_2 x}{\log_3 x} \right)^2 = o(x (\log_2 x / \log_3 x)^2).
\]

Once (1) is proved, it follows immediately that for any fixed \( \epsilon > 0 \), all but \( o(x) \) values of \( n \leq x \) are such that

\[
\left| \tilde{W}(n) - \frac{\log_2 x}{\log_3 x} \right| < \epsilon \frac{\log_2 x}{\log_3 x}.
\]

Lemma 5 then allows us to replace \( \tilde{W}(n) \) here with \( W(n) \). The resulting statement is equivalent to Theorem 1, upon observing that \( \frac{\log_2 n}{\log_3 n} = \frac{\log_2 x}{\log_3 x} + o(1) \) for \( \sqrt{n} < x \leq x \) (as \( x \to \infty \)).

Thus it remains only to establish (1). The following two lemmas suffice.

**Lemma 6.** As \( x \to \infty \),

\[
\frac{1}{x} \sum_{n \leq x} \tilde{W}(n) = (1 + o(1)) \frac{\log_2 x}{\log_3 x}.
\]

**Lemma 7.** As \( x \to \infty \),

\[
\frac{1}{x} \sum_{n \leq x} \tilde{W}(n)^2 = (1 + o(1)) \left( \frac{\log_2 x}{\log_3 x} \right)^2.
\]

We give a detailed proof of Lemma 6, and we sketch the (similar, but somewhat more tedious) proof of Lemma 7.

**Proof of Lemma 6.** We start by writing

\[
\sum_{n \leq x} \tilde{W}(n) = \sum_{p \in \mathcal{I}} \# \{ n \leq x : qq' | n \text{ for distinct } q, q' \leq x^{1/2 \log_3 x} \text{ with } q, q' \equiv 1 \mod p \} \\
= \sum_{p \in \mathcal{I}} (N_*(p) - N_0(p) - N_1(p)),
\]

\[
\sum_{n \leq x} \tilde{W}(n)^2 = \sum_{p \in \mathcal{I}} (N_*(p)^2 - 2N_0(p)N_1(p) + N_0(p)^2 + N_1(p)^2).
\]

...
where \( N_*(p) \) is the total count of positive integers \( n \leq x \), and for \( i = 0, 1 \),
\[
N_i(p) = \#\{n \leq x : \text{there are exactly } i \text{ primes } q \leq x^{1/2\log_3 x}, \; q \equiv 1 \pmod{p} \text{ dividing } n\}.
\]

We proceed to estimate each of \( \frac{1}{x} \sum_{p \in \mathcal{I}} N_*(p), \frac{1}{x} \sum_{p \in \mathcal{I}} N_0(p), \) and \( \frac{1}{x} \sum_{p \in \mathcal{I}} N_1(p) \).

Since \( N_*(p) = |\mathcal{I}| \), estimating \( \frac{1}{x} \sum_{p \in \mathcal{I}} N_*(p) \) is straightforward. Write \( \pi(t) = \int_0^t \frac{dt}{\log t} + E(t) \), so that \( E(t) \ll t/(\log t)^A \) for any fixed \( A \) and all \( t \geq 2 \). Then for each fixed \( A \),
\[
\frac{1}{x} \sum_{p \in \mathcal{I}} N_*(p) = \int_{\log_2 x}^{\log_3 x} \frac{dt}{\log t} + \int_{\log_2 x}^{\log_3 x} dE(t) + O\left(\frac{\log_2 x}{x}\right).
\]

Next we turn attention to \( \frac{1}{x} N_0(p) \). For each \( p \in \mathcal{I} \), Proposition 3 yields
\[
\frac{1}{x} N_0(p) = \left( \prod_{q \leq x^{1/2\log_3 x}, \; q \equiv 1 \pmod{p}} \left(1 - \frac{1}{q}\right) \right) \left(1 + O(1/\log_2 x)\right)
\]
\[
= \exp\left(- \sum_{q \leq x^{1/2\log_3 x}, \; q \equiv 1 \pmod{p}} \frac{1}{q}\right) \left(1 + O(1/\log_2 x)\right).
\]

We used here that \( \log(1 - \frac{1}{q}) = -\frac{1}{q} + O\left(\frac{1}{q^2}\right) \), and that \( \sum_{q \equiv 1 \pmod{p}} \frac{1}{q^2} < \sum_{q > p} \frac{1}{q} \ll 1/p \log p \ll 1/\log_2 x \). Continuing, we have by Lemma 4 that
\[
\sum_{q \leq x^{1/2\log_3 x}, \; q \equiv 1 \pmod{p}} \frac{1}{q} = \frac{\log_2 x - \log(2\log_3 x)}{p - 1} + O\left(\frac{\log_3 x}{p}\right)
\]
\[
= \frac{\log_2 x}{p} + O\left(\frac{\log_3 x}{p}\right) = \frac{\log_2 x}{p} + O\left(\frac{(\log_3 x)^2}{\log_2 x}\right).
\]

Inserting this above,
\[
(2) \quad \frac{1}{x} N_0(p) = \exp\left(- \frac{\log_2 x}{p}\right) \left(1 + O\left((\log_3 x)^2/\log_2 x\right)\right).
\]

Summing by parts,
\[
\sum_{p \in \mathcal{I}} \exp\left(- \frac{\log_2 x}{p}\right)
\]
\[
= \int_{\log_2 x}^{\log_3 x} \exp\left(- \frac{\log_2 x}{t}\right) \frac{dt}{\log t} + \int_{\log_2 x}^{\log_3 x} \exp\left(- \frac{\log_2 x}{t}\right) dE(t).
\]

We treat the second integral as an error term, noting that for any fixed \( A \),
\[
\int_{\log_2 x}^{\log_3 x} \exp\left(- \frac{\log_2 x}{t}\right) dE(t)
\]
\[
= E(t) \exp\left(- \frac{\log_2 x}{t}\right) \bigg|_{t=\log_2 x}^{t=\log_3 x} - \int_{\log_2 x}^{\log_3 x} E(t) \left(\frac{d}{dt} \exp\left(- \frac{\log_2 x}{t}\right)\right) dt
\]
\[
\ll (\sup_{t \in \mathcal{I}} |E(t)|) \left(1 + \int_{\log_2 x}^{\log_3 x} \left| \frac{d}{dt} \exp\left(- \frac{\log_2 x}{t}\right) \right| dt\right) \ll A (\log_2 x)/(\log_3 x)^A.
\]
Thus,
\[
\sum_{p \in \mathcal{I}} \exp \left( -\frac{\log_2 x}{p} \right) = \int_{\log_2 x/\log_3 x}^{\log_2 x \log_3 x} \exp \left( -\frac{\log_2 x}{t} \right) \frac{dt}{\log t} + O \left( \frac{\log_2 x}{(\log_3 x)^{A}} \right).
\]

We now deduce an estimate for \( \frac{1}{x} \sum_{p \in \mathcal{I}} N_0(p) \) by means of (2). The integrand in the last display is \( \ll 1/\log_3 x \) for every \( t \) in the range of integration, and \( \gg 1/\log_3 x \) for those \( t > \log_2 x \). Hence, the integral has size \( \ll \log_2 x \), as does \( \sum_{p \in \mathcal{I}} \exp(-\log_2 x/p) \). Using this along with (2), we conclude that for any fixed value of \( A \),
\[
\sum_{p \in \mathcal{I}} \frac{1}{x} N_0(p) = \int_{\log_2 x/\log_3 x}^{\log_2 x \log_3 x} \exp \left( -\frac{\log_2 x}{t} \right) \frac{dt}{\log t} + O((\log_2 x)/(\log_3 x)^A).
\]

Finally we consider \( \frac{1}{x} N_1(p) \). For each \( p \in \mathcal{I} \), the \( n \) counted by \( N_1(p) \) are precisely those integers expressible as \( Qm \), where \( Q \) is a power of a prime \( q \equiv 1 \pmod{p} \) with \( q \leq x^{1/2\log_3 x} \), and \( m \leq x/Q \) is free of prime factors \( q' \equiv 1 \pmod{p} \) with \( q' \leq x^{1/2\log_3 x} \). Thus,
\[
(3) \quad \frac{1}{x} N_1(p)
= \frac{1}{x} \sum_{q \leq x^{1/2\log_3 x}} \# \{ m \leq \frac{x}{q} : m \text{ not divisible by any } q' \leq x^{1/2\log_3 x}, q' \equiv 1 \pmod{p} \}
+ O \left( \frac{1}{x} \# \{ n \leq x : n \text{ is divisible by } q^2 \text{ for some } q > \log_2 x/\log_3 x \} \right).
\]
The \( O \)-term here has size \( \ll 1/\log_2 x \), and so summing on \( p \in \mathcal{I} \) will introduce an error of size
\[
(4) \quad \ll \frac{1}{\log_2 x} \sum_{p \in \mathcal{I}} 1 \ll 1,
\]
which is negligible for us. So we focus our attention on the main term of (3). For any \( q \leq x^{1/2\log_3 x} \), we have \( \frac{\log(x/q)}{\log(x/2^{1/2\log_3 x})} \geq 2 \log_3 x - 1 \), and so by Proposition 3,
\[
\# \{ m \leq \frac{x}{q} : m \text{ not divisible by any } q' \leq x^{1/2\log_3 x}, q' \equiv 1 \pmod{p} \}
= \frac{x}{q} \left( \prod_{q' \leq x^{1/2\log_3 x}} \left( 1 - \frac{1}{q'} \right) \right) (1 + O(1/\log_2 x))
\]
By an argument already given above,
\[
\prod_{q' \leq x^{1/2\log_3 x}} \left( 1 - \frac{1}{q'} \right) = \exp \left( -\frac{\log_2 x}{p} \right) (1 + O((\log_3 x)^2/\log_2 x))
\]
Thus,
\[
\frac{1}{x} \sum_{q \leq x^{1/2\log_3 x}} \# \{ m \leq \frac{x}{q} : m \text{ not divisible by any } q' \leq x^{1/2\log_3 x}, q' \equiv 1 \pmod{p} \}
= \sum_{q \leq x^{1/2\log_3 x}} \frac{1}{q} \exp \left( -\frac{\log_2 x}{p} \right) (1 + O((\log_3 x)^2/\log_2 x)).
\]
Since

\[ \sum_{q \leq x^{1/2} \log_3 x \atop q \equiv 1 \pmod{p}} \frac{1}{q} = \frac{\log_2 x}{p} + O \left( \frac{\log_3 x}{p} \right) = \frac{\log_2 x}{p} \left( 1 + O \left( \frac{\log_3 x}{\log_2 x} \right) \right), \]

we deduce that

\[ (5) \quad \frac{1}{x} \sum_{q \leq x^{1/2} \log_3 x \atop q \equiv 1 \pmod{p}} \# \{ m \leq \frac{x}{q} : m \text{ not divisible by any } q' \leq x^{1/2} \log_3 x, q' \equiv 1 \pmod{p} \} \]

\[ = \frac{\log_2 x}{p} \exp \left( - \frac{\log_2 x}{p} \right) \left( 1 + O(\log_3 x)^2 / \log_2 x \right). \]

We now sum on \( p \in \mathcal{I} \). Applying summation by parts in the same manner as before, we find after some calculation that for each fixed \( A \),

\[ (6) \quad \sum_{p \in \mathcal{I}} \frac{\log_2 x}{p} \exp \left( - \frac{\log_2 x}{p} \right) = \int_{\log_2 x / \log_3 x}^{\log_2 x / \log_3 x} \frac{\log_2 x}{t} \exp \left( - \frac{\log_2 x}{t} \right) \frac{dt}{\log t} + O \left( \frac{\log_2 x}{(\log_3 x)^A} \right); \]

moreover, the integral appearing here is of size \( \log_2 x \log_4 x \). (The last estimate may be seen by noting that the integrand is \( \ll \log_2 x / t \) on the entire interval, and \( \gg \log_2 x / t \) for \( t > \log_2 x \).) We now deduce from (3), (4), (5), and (6) that

\[ \sum_{p \in \mathcal{I}} \frac{1}{x} N_1(p) = \int_{\log_2 x / \log_3 x}^{\log_2 x / \log_3 x} \frac{\log_2 x}{t} \exp \left( - \frac{\log_2 x}{t} \right) \frac{dt}{\log t} + O \left( \frac{\log_2 x}{(\log_3 x)^A} \right). \]

Piecing everything together, we find that for any fixed \( A \),

\[ (7) \quad \frac{1}{x} \sum_{n \leq x} \tilde{W}(n) = I_* - I_0 - I_1 + O(\log_2 x / (\log_3 x)^A), \]

where

\[ I_* := \int_{\log_2 x / \log_3 x}^{\log_2 x / \log_3 x} \frac{dt}{\log t}, \]

\[ I_0 := \int_{\log_2 x / \log_3 x}^{\log_2 x / \log_3 x} \frac{1}{t} \exp \left( - \frac{\log_2 x}{t} \right) \frac{dt}{\log t}, \]

\[ I_1 := \int_{\log_2 x / \log_3 x}^{\log_2 x / \log_3 x} \frac{\log_2 x}{t} \exp \left( - \frac{\log_2 x}{t} \right) \frac{dt}{\log t}. \]

Now as \( x \to \infty \),

\[ I_* - I_0 - I_1 = \int_{\log_2 x / \log_3 x}^{\log_2 x / \log_3 x} \left( 1 - \exp \left( - \frac{\log_2 x}{t} \right) - \frac{\log_2 x}{t} \exp \left( - \frac{\log_2 x}{t} \right) \right) \frac{dt}{\log t} \]

\[ = 1 + O(1) \int_{\log_3 x}^{\log_2 x / \log_3 x} \left( 1 - \exp \left( - \frac{\log_2 x}{t} \right) - \frac{\log_2 x}{t} \exp \left( - \frac{\log_2 x}{t} \right) \right) dt. \]

We recognize the integrand as the derivative of \( t - t \exp(-\log_2 x / t) \). Plugging our upper endpoint into this last expression yields

\[ \log_2 x \log_3 x \cdot (1 - \exp(-1 / \log_3 x)) = \log_2 x \log_3 x \cdot (1 / \log_3 x + O(1 / (\log_3 x)^2)) \]

\[ = (1 + o(1)) \log_2 x. \]

The lower endpoint gives a contribution of smaller order, \( O(\log_2 x / \log_3 x) \). Thus, the integral is asymptotic to \( \log_2 x \), and \( I_* - I_0 - I_1 \) is asymptotic to \( \log_2 x / \log_3 x \). Referring back to (7), the lemma is proved. \( \square \)
Proof of Lemma 7 (sketch). We start by writing

$$\sum_{n \leq x} \tilde{W}(n)^2 =$$

$$\sum_{p_1, p_2 \in \mathcal{I}} \sum_{n \leq x} (1_{\ast p_1, p_2} (n) - 1_{\ast p_1, p_2} (n) - 1_{\ast p_1, p_2} (n))(1_{\ast p_1, p_2} (n) - 1_{\ast p_1, p_2} (n) - 1_{\ast p_1, p_2} (n)),$$

where $1_{\ast p_1, p_2}$ is the indicator function of $n$ having exactly $i_1$ prime factors congruent to $1$ modulo $p_1$ not exceeding $\frac{x^{1/2}}{\log x}$, and exactly $i_2$ prime factors congruent to $1$ modulo $p_2$ not exceeding $\frac{x^{1/2}}{\log x}$, and $\ast$ indicates no restriction. Expanding the product and performing the sum on $n$ gives $\sum_{n \leq x} \tilde{W}(n)^2$ as a signed sum of terms of the form $N_{i_1, i_2}(p_1, p_2)$, where

$$N_{i_1, i_2}(p_1, p_2) = \sum_{n \leq x} 1_{i_1 p_1, i_2 p_2} (n)$$

$$= \# \left\{ n \leq x \mid \exists \text{ exactly } i_1 \text{ primes } q_1 \leq \frac{x^{1/2}}{\log x}, \ q_1 \equiv 1 \mod{p_1} \text{ dividing } n, \ \exists \text{ exactly } i_2 \text{ primes } q_2 \leq \frac{x^{1/2}}{\log x}, \ q_2 \equiv 1 \mod{p_2} \text{ dividing } n \right\}. $$

We claim that for each pair of indices $i_1, i_2 \in \{0, 1, \ast\}$,

$$\frac{1}{x} \sum_{p_1, p_2 \in \mathcal{I}} N_{i_1, i_2}(p_1, p_2) = I_{i_1} I_{i_2} + O((\log x)^2/(\log x)^4),$$

where the $I$s are as in (8), and where as before $A$ is arbitrary but fixed. Retracing our steps shows that

$$\frac{1}{x} \sum_{n \leq x} \tilde{W}(n)^2 = (I_\ast - I_0 - I_1)^2 + O((\log x)^2/(\log x)^4).$$

From the proof of Lemma 6, we know that $(I_\ast - I_0 - I_1) \sim \frac{\log x}{\log x} (as \ x \to \infty)$, and so Lemma 7 follows.

The estimate (9) can be proved by the same method used in the proof of Lemma 6. We say a few words here about $N_{1,1}$; the other cases are similar.

The pairs $p_1 = p_2$ make a contribution to $\frac{1}{x} \sum_{p_1, p_2 \in \mathcal{I}} N_{1,1}(p_1, p_2)$ of size $\frac{1}{x} \sum_{p \in \mathcal{I}} N_1(p)$, which is $O((\log x)^2 \log x)$, and so is negligible for us. Assume now that $p_1 \neq p_2$. In that case the $n$ counted in $N_{1,1}(p_1, p_2)$ include all those that have the form $n = q_1 q_2 m$, where

- $q_1, q_2 \leq \frac{x^{1/2}}{\log x}$,
- $q_1 \equiv 1 \mod{p_1}$ and $q_1 \not\equiv 1 \mod{p_2}$,
- $q_2 \equiv 1 \mod{p_2}$ and $q_2 \not\equiv 1 \mod{p_1}$,
- $m$ is free of prime factors $\equiv 1 \mod{q_1}$ or $\equiv 1 \mod{q_2}$.

Say that these $n$ are of the first kind (with respect to $p_1, p_2$), and that all other $n$ counted by $N_{1,1}(p_1, p_2)$ are of the second kind. If $n$ is of the second kind, then either $n$ is divisible by $q_2$ for some prime $q > \frac{\log x}{\log x}$ or $n$ is divisible by some prime $q \equiv 1 \mod{p_1 p_2}$; the number of these $n$ is

$$\ll x \sum_{q > \frac{\log x}{\log x}} \frac{1}{q^2} + x \sum_{\frac{\log x}{p_1 p_2}} \frac{1}{q} \ll \frac{x}{\log x} + \frac{\log x}{p_1 p_2}.$$

Summing on $p_1, p_2 \in \mathcal{I}$, we see that $n$ of the second kind will make a total contribution to $\frac{1}{x} \sum_{p_1, p_2 \in \mathcal{I}} N_{p_1, p_2}$ of size $O(\log x)$. This is of smaller order than $(\log x)^2/(\log x)^2$, and so is negligible for us. So we move our attention over to the $n$ of the first kind. For a given
\[ p_1, p_2 \text{ and a given } q_1, q_2, \text{ the count of these } n, \text{ after dividing by } x, \text{ is} \]
\[ \frac{1}{q_1 q_2} \left( \prod_{q \equiv 1 (\mod p_1) \text{ or } q \equiv 1 (\mod p_2)} \left( 1 - \frac{1}{q} \right)^{1/\log_3 x} \right) (1 + O(1/\log_2 x)) \]
\[ = \frac{1}{q_1 q_2} \exp \left( \sum_{q \equiv 1 (\mod p_1) \text{ or } q \equiv 1 (\mod p_2)} -\frac{1}{q} \right) (1 + O(1/\log_2 x)). \]

By Lemma 4,
\[ \sum_{q \equiv 1 (\mod p_1) \text{ or } q \equiv 1 (\mod p_2)} \frac{1}{q} = \frac{\log_2 x}{p_1 - 1} + \frac{\log_2 x}{p_2 - 1} - \frac{\log_2 x}{(p_1 - 1)(p_2 - 1)} + O \left( \frac{\log_3 x}{p_1} + \frac{\log_3 x}{p_2} \right) \]
\[ = \frac{\log_2 x}{p_1} + \frac{\log_2 x}{p_2} + O \left( \frac{(\log_3 x)^2}{\log_2 x} \right), \]
which shows that our normalized count above is
\[ \frac{1}{q_1 q_2} \exp \left( -\frac{\log_2 x}{p_1} \right) \exp \left( -\frac{\log_2 x}{p_2} \right) (1 + O((\log_3 x)^2/\log_2 x)). \]

Now we sum on \( q_1, q_2 \). We have that
\[ \sum_{q_1} \frac{1}{q_1} = \sum_{\ell \equiv 1 (\mod p_1)} \frac{1}{\ell} - \sum_{\ell \equiv 1 (\mod p_1 p_2)} \frac{1}{\ell} \]
\[ = \frac{\log_2 x}{p_1} + O \left( \frac{(\log_3 x)^2}{\log_2 x} \right) = \frac{\log_2 x}{p_1} \left( 1 + O \left( \frac{(\log_3 x)^3}{\log_2 x} \right) \right), \]
and similarly for the analogous sum on \( q_2 \). Thus, the first kind \( n \) make a contribution to \( \frac{1}{x} N_{1,1}(p_1, p_2) \) of
\[ \frac{\log_2 x}{p_1} \exp \left( -\frac{\log_2 x}{p_1} \right) \frac{\log_2 x}{p_2} \exp \left( -\frac{\log_2 x}{p_2} \right) (1 + O((\log_3 x)^3/\log_2 x)). \]
It remains to sum on \( p_1, p_2 \in \mathcal{I} \) with \( p_1 \neq p_2 \). If we include the terms with \( p_1 = p_2 \), this increases the sum by only \( O(\log_2 x) \), which is negligible for us. So we sum on all pairs \( p_1, p_2 \in \mathcal{I} \), which lets us factor the sum into pieces already estimated in the proof of Lemma 6. Using those results, we see that summing the last displayed expression over \( p_1, p_2 \in \mathcal{I} \) gives \( I_1^2 + O((\log_2 x)^2/(\log_3 x)^3) \), which is our claimed estimate for \( \frac{1}{x} \sum_{p_1, p_2 \in \mathcal{I}} N_{1,1}(p_1, p_2). \)

\[ \square \]

4. Concluding remarks: Beyond normal order

We alluded in the introduction to the result of Erdős–Pomerance and Murty–Murty that the number of primes dividing \( \phi(n) \) has normal order \( \frac{1}{2}(\log_2 n)^2 \). In fact, the main result of [5] (obtained independently in [10]) is quite a bit more precise: The number of primes dividing \( \phi(n) \) is normally distributed with mean \( \frac{1}{2}(\log_2 n)^2 \) and variance \( \frac{1}{2}(\log_2 n)^3 \). The alluded-to results of Martin and Troupe in [9] are also Gaussian laws and not merely normal order theorems. It would seem interesting to investigate whether \( W(n) \) (after an appropriate normalization) also possesses a limiting distribution.
Acknowledgements

The author thanks Greg Martin for his inspiring talk at the 2019 Winter CMS meeting on the statistics of $U(\mathbb{Z}/n\mathbb{Z})$, and for suggesting the remarks in the introduction about $\omega(\lambda(n)/\phi(n))$ versus $\omega(m)$ for $m$ of similar size.

References


Department of Mathematics, University of Georgia, Athens, GA 30602
Email address: pollack@uga.edu