ON THE NUMBER OF PRIMITIVE $\lambda$-ROOTS

THOMAS W. MÜLLER and JAN-CHRISTOPH SCHLAGE-PUCHTA

1. Introduction and results

For an integer $n$, denote by $U(n)$ the multiplicative group of residue classes modulo $n$. The structure of $U(n)$ is well known:

(i) If $n = \prod_{i=1}^{k} p_i^{a_i}$, then

$$U(n) \cong U(p_1^{a_1}) \times U(p_2^{a_2}) \times \cdots \times U(p_k^{a_k}).$$

(ii) If $p$ is an odd prime, then $U(p^a) \cong C_{p^a-1(p-1)}$.

(iii) $U(2)$ is trivial, $U(4) \cong C_2$, and $U(2^a) \cong C_2 \times C_{2^{a-2}}$ for $a \geq 3$.

The exponent of $U(n)$, that is, the least integer $\nu$ such that $a^\nu \equiv 1 \pmod{n}$ for all integers $a$ prime to $n$, is denoted by $\lambda(n)$. This function was introduced around 1910 by Carmichael; cf. [2] and [3]. By a primitive $\lambda$-root of $n$, we mean any element of maximal order $\lambda(n)$ in $U(n)$. This concept, which was introduced by Carmichael in [2], is a natural generalization of primitive roots. Let $r(n)$ be the number of primitive $\lambda$-roots of $n$. It is not difficult to see that

$$r(n) = \varphi(n) \prod_{p|\varphi(n)} \left(1 - p^{-m(p)}\right),$$

where $\varphi(n)$ is Euler’s totient function, and $m(p)$ is the number of elementary divisors of $U(n)$ whose $p$-part is maximal. We see that $r(n) \geq \varphi(\varphi(n))$ with equality if and only
if \( m(p) = 1 \) for all prime numbers \( p \). In [1], Cameron and Preece raise the problem to determine the density of the set

\[ \mathcal{R} = \{ n : r(n) = \varphi(\varphi(n)) \}. \] (2)

They note that a computer search reveals almost 60% of all numbers below \( 10^5 \) to have this property and wonder whether the set \( \mathcal{R} \) might have positive density. Integers \( n \in \mathcal{R} \) have another interesting property. Define an equivalence relation \( \sim \) on the set of primitive \( \lambda \)-roots by \( a \sim b \) if and only if \( \langle a \rangle = \langle b \rangle \). Then the number of equivalence classes is at least \( \varphi(n)/\lambda(n) \), with equality occurring in the latter inequality if and only if \( n \in \mathcal{R} \).

For a positive integer \( n \), define \( f(n) \) to be the number of primes \( p \) such that \( m(p) \geq 2 \), where \( m(p) \) is defined as in (1). Our main results are as follows.

**Theorem 1.** The function \( f(n) \) has a normal distribution with mean \( \frac{\log_2 n}{\log_3 n} \) and variance \( \frac{\log_2 n}{2 \log_3 n} \).

**Theorem 2.** For any constant \( A > 0 \), we have

\[ \sum_{\substack{n \in \mathcal{R} \\ n \leq x}} 1 \ll \frac{x}{(\log_2 x)^A}; \]

in particular, \( \mathcal{R} \) has density 0.

Here, \( \log_k x \) denotes the \( k \)-fold iterated logarithm.

2. **Proof of theorem 1**

We will repeatedly use the following result.
Lemma 1. Let $q \geq 3$ be an integer. Then we have uniformly in $x > e^q$ the estimate

$$
\sum_{\substack{p \leq x \\ p \equiv 1 (q)}} \frac{1}{p} \sim \frac{\log_2 x}{\varphi(q)}.
$$

Proof. Let $\varepsilon > 0$ be given, and set $y = \exp \left( (\log x)^\varepsilon \right)$. Using the Siegel-Walfisz-Theorem (see [7]), we find that

$$
\sum_{\substack{p \leq y \\ p \equiv 1 (q)}} \frac{1}{p} = \frac{\log_2 x - \log_2 y}{\varphi(q)} + O(1),
$$

whereas the Brun-Titchmarsh-inequality (cf. [5, Theorem 3.8] or [6]) implies

$$
\sum_{\substack{q \leq p < y \\ p \equiv 1 (q)}} \frac{1}{p} \leq \frac{(4 + o(1)) \log_2 y}{\varphi(q)}.
$$

Together with the trivial estimate

$$
\sum_{\substack{q \leq p < q^2 \\ p \equiv 1 (q)}} \frac{1}{p} \leq \sum_{\substack{q \leq p < q^2}} \frac{1}{p} \ll 1
$$

our claim follows. \qed

We now focus on the proof of Theorem 1. Note that $m(q)$ can also be described as the number of prime power block factors $p^a$ of $n$ such that the $q$-part of $\varphi(p^a)$ is maximal among all such $p$; that is, $f(n)$ is the number prime powers $q^a$ satisfying the following two conditions:

(i) there exist distinct prime divisors $p_1, p_2$ of $n$, such that $p_1, p_2 \equiv 1 \pmod{q^a}$;

(ii) there exists no prime divisor $p$ of $n$ such that $p \equiv 1 \pmod{q^{a+1}}$. 


Fix a parameter $0 < \delta < 1$, and define the auxiliary function $f_\delta(n)$ to be the number of primes $q \in [\delta \log_2 n, \delta^{-1} \log_2 n]$ satisfying conditions (i) and (ii). Our first aim is to show the estimate

$$\sum_{n \leq x} (f(n) - f_\delta(n)) \ll \delta x \frac{\log_2 x}{\log_3 x}.$$  \hspace{1cm} (3)

First note that we may replace the interval $[\delta \log_2 n, \delta^{-1} \log_2 n]$ by $[\delta \log_2 x, \delta^{-1} \log_2 x]$ by increasing the value of $\delta$. Let $q^a$ be a prime power. We bound the number of integers $n \leq x$ such that $q^a$ contributes to $f(n)$ by neglecting condition (ii). This quantity equals

$$\sum_{p_1 < p_2 \atop p_1, p_2 \equiv 1 (q^a)} \left\lfloor \frac{x}{p_1 p_2} \right\rfloor \leq \sum_{p_1 p_2 \leq x \atop p_1, p_2 \equiv 1 (q^a)} \frac{x}{p_1 p_2}$$

$$\leq x \left( \sum_{p \leq x \atop p \equiv 1 (q^a)} \frac{1}{p} \right)^2$$

$$\sim x \left( \frac{\log_2 x}{q^a} \right)^2,$$  \hspace{1cm} (4)

where we have used Lemma 1 for the last step. Summing (4) over prime power values $q^a > \delta^{-1} \log_2 x$, we find that the contribution of such prime powers to the left-hand side of (3) is of acceptable magnitude. Since there are less than $\log_2^{1/2} x$ proper prime powers below $\log_2 x$, we see that the contribution of proper prime powers is altogether negligible. Finally, there are $O(\delta \log_2 x / \log_3 x)$ prime numbers below $\delta \log_2 x$, which is again of acceptable order, and (3) is proved.
Define $\tilde{f}_\delta$ to be the number of primes $q \in [\delta \log_2 x, \delta^{-1} \log_2 x]$ satisfying condition (i).

Then, using Lemma 1, we have

$$\sum_{n \leq x} (\tilde{f}_\delta(n) - f_\delta(n)) \leq \sum_{\delta \log_2 x \leq q \leq \delta^{-1} \log_2 x} \sum_{p \equiv 1 (q^2)} \left\lfloor \frac{n}{p} \right\rfloor \leq x \sum_{\delta \log_2 x \leq q \leq \delta^{-1} \log_2 x} \frac{\log_2 x}{q^2} \ll x \frac{\log_2 x}{\log_3 x + \log \delta}.$$

Now we use the method of moments (see, for instance, [4]) to compute the distribution of $\tilde{f}_\delta$. For an integer $n$, denote by $\tilde{m}(q)$ the number of primes $p_i$ satisfying condition (i). We claim that, for fixed $q \in [\delta \log_2 x, \delta^{-1} \log_2 x]$ and $n \in [1, x]$ chosen at random, the distribution of $\tilde{m}(q)$ converges to a Poisson distribution with mean $\frac{\log_2 x}{q}$, and that for different primes $q_1, \ldots, q_k$ the random variables are asymptotically independent. It follows that the random variables

$$\xi_q = \begin{cases} 1, & \text{if } \tilde{m}(q) \geq 2 \\ 0, & \text{otherwise} \end{cases}$$

are asymptotically independent, have means

$$1 - e^{-(\log_2 x)/q} - \frac{\log_2 x}{q} e^{-(\log_2 x)/q},$$

respectively, and variance

$$\left(1 - e^{-(\log_2 x)/q} - \frac{\log_2 x}{q} e^{-(\log_2 x)/q}\right) \left(e^{-(\log_2 x)/q} + \frac{\log_2 x}{q} e^{-(\log_2 x)/q}\right).$$

From this, Theorem 1 follows in view of the facts that

$$\int_0^\infty 1 - e^{-1/t} - \frac{1}{t} e^{-1/t} \, dt = 1.$$
and
\[ \int_0^\infty \left( 1 - e^{-1/t} - \frac{1}{t} e^{-1/t} \right) \left( e^{-1/t} + \frac{1}{t} e^{-1/t} \right) dt = \frac{1}{2}. \]

Hence, it remains to study the higher moments of the variables \( \tilde{m}(q) \) and their correlations. To do so, we compute the expected value of \( \binom{\tilde{m}(q)}{k} \) for fixed \( k \geq 1 \). We find that
\[
E \left( \binom{\tilde{m}(q)}{k} \right) = \sum_{n \leq x} |\{p_1 < p_2 < \cdots < p_k : p_i \equiv 1 \pmod{q}, p_i | n\}|
\]
\[
= \sum_{p_1 < \cdots < p_k \atop p_i \equiv 1 \pmod{q}} \frac{x}{p_1 \cdots p_k} + O\left( \frac{x \log^k x}{\log x} \right)
\]
\[
= \frac{x}{k!} \left( \sum_{p \leq x \atop p \equiv 1 \pmod{q}} \frac{1}{p} + O\left( \frac{1}{q} \right) \right)^k + O\left( \frac{x}{\log_2 x} \right)
\]
\[
= \frac{x}{k!} \left( \frac{\log_2 x}{q} \right)^k + O\left( \frac{x}{\log_2 x} \right).
\]

On the other hand, the \( k \)-th moment of a Poisson distribution with mean \( \frac{\log_2 x}{q} \) is
\[
E(\xi^k) = \sum_{\kappa=0}^k s_{\kappa,k} \left( \frac{\log_2 x}{q} \right)^\kappa,
\]
where the \( s_{\kappa,k} \) are Stirling numbers of the second kind. By the Stirling inversion formula, the last assertion is equivalent to
\[
\sum_{\kappa=0}^k s_{\kappa,k} \left( \frac{\log_2 x}{q} \right)^\kappa = \left( \frac{\log_2 x}{q} \right)^k,
\]
where the $s_{\kappa,k}$ are Stirling numbers of the first kind. Since
\[ \sum_{\kappa=0}^{k} s_{\kappa,k} x^\kappa = x(x-1) \cdots (x-k+1), \]
the variables $\tilde{m}(q)$ converge to a Poisson distribution with mean $(\log_2 x)/q$.

To show that the variables $\tilde{m}(q)$ are asymptotically independent, it suffices to show that for fixed integers $k_1, \ldots, k_l$, we have
\begin{equation}
E\left( \frac{\tilde{m}(q_1)}{k_1} \right) \cdots \left( \frac{\tilde{m}(q_l)}{k_l} \right) \sim \left( E\left( \frac{\tilde{m}(q_1)}{k_1} \right) \right) \left( E\left( \frac{\tilde{m}(q_2)}{k_2} \right) \right) \cdots \left( E\left( \frac{\tilde{m}(q_l)}{k_l} \right) \right). \tag{5} \end{equation}

The left-hand side quantity can be written as
\[ \sum_{n \leq x} \left| \left\{ p_{11} < \cdots < p_{1k_1}, \ldots, p_{l1} < \cdots < p_{lk_1} : \forall i,j : p_{ij} \equiv 1 (q_1), p_{ij} \mid n \right\} \right|. \]
If all primes $p_{ij}$ are different, this can be computed as above and is easily seen to be asymptotically equal to the right-hand side of (5). It suffices to compare the contribution of tuples satisfying $p_{11} = p_{21}$, say, with all tuples. Note that restricting $p_{ij}$ by $x^{1/(2k)}$ does not change the expectations significantly, hence, writing $M$ for the least common multiple of all $p_{ij}$, $(i,j) \neq (1,1), (1,2)$, we have $M \leq \sqrt{x}$. Then we obtain
\[ \sum_{n \leq x} \sum_{\substack{p \mid n \\text{M} \mid n \ p \equiv 1(q_1q_2) \\text{p} \mid n}} 1 \ll \frac{x \log_2 x}{M q_1 q_2} + m \frac{x}{M}, \]
where $m$ denotes the number of primes among $p_{ij}$, $(i,j) \neq (1,1), (1,2)$, which are congruent to 1 modulo $q_1 q_2$. Since
\[ \sum_{n \leq x} \left| \left\{ p_1 \equiv 1 (\text{mod } q_1), p_2 \equiv 1 (\text{mod } q_2), p_1, p_2 \mid n \right\} \right| \gg \frac{x \log_2^2 x}{M q_1 q_2} + m \frac{x}{M}, \]
we see that tuples with repetitions are indeed negligible, proving that the random variables $\tilde{m}(q)$ are asymptotically independent.
Define $f_\delta$ as in the proof of Theorem 1. Since $f(n) \geq f_\delta(n)$, it suffices to consider the set

$$R_\delta := \{n: f_\delta(n) = 0\}.$$  

Moreover, from the computation of the moments of $\tilde{f}_\delta$ we know that the number of integers $n \leq x$ satisfying $\tilde{f}_\delta(n) \leq \frac{1}{2} \log_2 x$ is bounded above by $O\left(\frac{x}{\log_2 x}\right)$ for every constant $A$, provided that $\delta$ is sufficiently small. Hence, it suffices to consider the set

$$S_\delta := \{n: \tilde{f}_\delta(n) - f_\delta(n) \geq \frac{1}{2} \log_2 x\}.$$  

For an integer $k \geq 1$, we have

$$\sum_{n \leq x} \left(\frac{\tilde{f}_\delta(n) - f_\delta(n)}{k}\right) \leq \sum_{\delta \log_2 x \leq q_1 < q_2 < \cdots < q_k \leq \delta^{-1} \log_2 x} \left|\{(n, p_1, \ldots, p_k): p_i|n, p_i \equiv 1 (q_i^2)\}\right|.$$  

(6)

Restricting the range for $p_i, 1 \leq i \leq k$ to $[1, x^{1/(2k)}]$ introduces an error term of order

$$\sum_{\delta \log_2 x \leq q_1 < q_2 < \cdots < q_k \leq \delta^{-1} \log_2 x} \frac{1}{q_1^2 q_2^2 \cdots q_k^2} \ll \delta^{-k} \log_2^{-k} x.$$  

Now fix $q_1, \ldots, q_k$ as above, and assume that $p_1 = p_2$, say. Fix $p_3, \ldots, p_k$, and let $M$ be the least common multiple of $p_3, \ldots, p_k$. Then the contribution of all possible choices for $p_1$ and $p_2$ is

$$\left|\{(n, p) : pM|n, p \equiv 1 (q_1^2 q_2^2)\}\right| \leq (1 + o(1)) \frac{x \log_2 x}{Mq_1^2 q_2^2},$$  

whereas the number of all triples $(n, p_1, p_2)$ is $(1 + o(1)) \frac{x \log_2 x}{Mq_1^2 q_2^2}$. Hence, the contribution of tuples $(n, p_1, \ldots, p_k)$ with repetitions to the right-hand side of (6) is of lesser order
ON THE NUMBER OF PRIMITIVE $\lambda$-ROOTS

than the contribution of tuples without repetitions. We obtain

$$\sum_{n \leq x} \left( \frac{\tilde{f}_\delta(n) - f_\delta(n)}{k} \right) \leq (1 + o(1)) x \sum_{\delta \log_2 x \leq q_1 < q_2 < \cdots < q_k \leq \delta^{-1} \log_2 x} \prod_{p \leq x} \left( \sum_{p \equiv 1 (q_i^2)} \frac{1}{p} \right) \tag{7}$$

$$\leq (1 + o(1)) x \sum_{\delta \log_2 x < q_1 < q_2 < \cdots < q_k \leq \delta^{-1} \log_2 x} \frac{\log^k x}{q_1^2 q_2^2 \cdots q_k^2}$$

$$\leq \frac{(1 + o(1)) x \pi(\delta^{-1} \log_2 x)}{\delta^{2k} \log_2^k x}$$

$$\leq \frac{(1 + o(1)) x}{\delta^{3k} \log_3^k x}.$$

Since integers $n$ with $\tilde{f}_\delta(n) - f_\delta(n) \geq \frac{1}{2} \log_2 x$ contribute at least $\frac{\log^k x}{3^k k!}$ to the left-hand side of (7), Theorem 2 follows.

REFERENCES


