TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 293, Number 2, February 1986

THE LEAST r-FREE NUMBER IN AN ARITHMETIC PROGRESSION

BY

KEVIN S. McCURLEY

ABSTRACT. Let $n_r(a,q)$ be the least r-free number in the arithmetic progession a modulo q. Several results are proved that give lower bounds for $n_r(a,q)$, improving on previous results due to Erdös and Warlimont. In addition, a heuristic argument is given, leading to two conjectures that would imply that the results of the paper are close to best possible.

1. Introduction. If r is at least 2, define $n_r(a,q)$ as the least positive integer in the arithmetic progression a modulo q that is not divisible by the rth power of a prime, and define

$$n_r^*(q) = \max_{(a,q)=1} n_r(a,q), \qquad n_r(q) = \max_{(a,q)\ r\text{-free}} n_r(a,q).$$

In this paper we give some lower bounds for these functions, and state some conjectures concerning their rate of growth.

There are several known upper bound results. Prachar [8] has shown that

$$n_r^*(q) \ll q^{1+1/r}\exp\left(\frac{r}{r-1}\omega(q)\log r\right),$$

where $\omega(q)$ is the number of distinct prime factors of q. If r is large as a function of q, the work of Cohen and Robinson [1] yields the sharper estimates

$$n_r^*(q) \ll q^{1+1/(r-1)}, \qquad n_r(q) \ll \left(q^2/\varphi(q)\right)^{1+1/(r-1)}.$$

Furthermore any result is trivial if r exceeds $\log q/\log 2$, since $n_r(a,q)$ does not exceed q in this case.

In the case when r is 2, improvements in Prachar's upper bound result have been made by Erdös [2] and more recently by Heath-Brown [3], who proved that

$$n_2(q) \ll q^{13/9}(d(q)\log q)^6,$$

where $d(q)$ is the number of divisors of q. Hooley [4] has also shown that $n_2^*(q) \ll q^{4/3+\varepsilon}$ for a sequence of q's having positive lower density. Further results concerning averages of $n_2(a,q)$ have been given by Warlimont [13].

As for lower bounds, Warlimont [12] proved that for every $C \geqslant 1$ there exists an $\varepsilon = \varepsilon(C)$ such that $n_2(a,q)$ exceeds Cq for infinitely many q and at least $\varepsilon\varphi(q)$ values of a for each q. Erdös [2] stated without proof that

$$n_2^*(q) \neq o\left(q\frac{\log q}{\log\log q}\right),$$

Received by the editors November 16, 1983. The contents of this paper were presented on March 18, 1983 at the 802nd Meeting of the American Mathematical Society in Norman, Oklahoma.

1980 *Mathematics Subject Classification*. Primary 10H15, 10H20.

and Warlimont [12] gave a proof that there exist infinitely many q with

$$n_2^*(q) > \left(\frac{1}{3} - \varepsilon\right) q \frac{\log q}{\log\log q}. \tag{1}$$

The values of q constructed by Warlimont were quite rare, being a product of many small prime factors, and one might be led to believe that (1) occurs only very rarely. In fact, we show that a slightly stronger result is true for *all* q, and for a large number of residue classes for each q.

THEOREM 1. *If $\varepsilon > 0$ and $\log q/(r \log\log q)$ is sufficiently large, then there exist at least*

$$\exp\left(\frac{1-\varepsilon}{r}\left(1 - \frac{\log r}{\log\log q}\right)\log q\right)$$

values of a with $(a, q) = 1$, $0 < a < q$, and

$$n_r(a, q) > \frac{1-\varepsilon}{r} q \frac{\log q}{\log\log q}. \tag{2}$$

For some values of q the size of the constant $(1 - \varepsilon)/r$ in (2) can be improved slightly, but for this we pay a price in the number of residue classes a for which the estimate is known to hold.

THEOREM 2. *If $\varepsilon > 0$, $\log q/(r \log\log q)$ is sufficiently large, and q is not divisible by the first $[\log q/(r \log\log q)]$ primes, then*

$$n_r^*(q) > (1 - \varepsilon)\frac{\zeta(r)}{r} q \frac{\log q}{\log\log q}.$$

If we focus our attention on a particular value of a, then the following result gives positive information.

THEOREM 3. *Let r be fixed, $\varepsilon > 0$, and $a > 0$. If a is not r-free, then there exist infinitely many q with $(a, q) = 1$ and*

$$n_r(a, q) > \frac{1-\varepsilon}{r} q \frac{\log q}{\log\log q}.$$

The previous results have been concerned with the residue classes a with $(a, q) = 1$. If we relax this condition, then we can obtain the following improvement.

THEOREM 4. *For each r and $\varepsilon > 0$ there exist infinitely many q such that*

$$n_r(q) > \frac{e^\gamma - \varepsilon}{r} q \log q \log_2 q \frac{\log_4 q}{(\log_3 q)^2},$$

where $\log_n q$ is the n-fold iterated logarithm and γ is Euler's constant.

This result is very similar to the best known lower bounds of Prachar [9] and Pomerance [7] for the least prime in an arithmetic progression.

There may still be improvements that can be made here, particularly in the distribution of $n_r(a, q)$ about its mean value.

2. Some conjectures. There is still a large gap between the lower bounds presented here and the upper bound results of Prachar and Heath-Brown. In this section we give heuristic arguments for two conjectures concerning the order of magnitude of $n_r^*(q)$ and $n_r(q)$. For simplicity we shall assume that r is fixed in this section.

Previously Erdös [2] conjectured that $n_r^*(q) \ll q^{1+\varepsilon}$, and to this we add the following conjectures.

CONJECTURE 1. Let $C_r(q) = \prod_{p \nmid q}(1 - p^{-r})$. Then

$$\overline{\lim_{q\to\infty}} \frac{n_r^*(q)}{q \log q/-\log(1 - C_r(q))} = 2.$$

Furthermore there exists a sequence S with asymptotic density zero such that

$$\lim_{\substack{q\to\infty \\ q \notin S}} \frac{n_r^*(q)}{q \log q/-\log(1 - C_r(q))} = 1.$$

CONJECTURE 2. Let

$$D_r(q) = \prod_{p^r \nmid q} (1 - (p^r, q)p^{-r}).$$

If $\varepsilon > 0$, then for q sufficiently large we have

$$n_r(q) < (2 + \varepsilon)q \frac{\log q}{-\log(1 - D_r(q))}.$$

It is interesting to compare the conjectures with the results stated in §1. If q is the product of all primes not exceeding z, with z large, then

$$C_r(q) = \exp\Big(\sum_{p>z} \log(1 - p^{-r})\Big) = \exp\Big(-\sum_{p>z} p^{-r} + O(z^{1-2r})\Big)$$
$$= 1 - \sum_{p>z} p^{-r} + O(z^{2-2r}).$$

It follows from the Prime Number Theorem and partial summation that

$$-\log(1 - C_r(q)) \sim (r - 1)\log z \sim (r - 1)\log\log q.$$

It follows from Conjecture 1 that for these q's we have

$$\frac{n_r^*(q)}{q \log q/\log\log q} < \frac{2 + \varepsilon}{r - 1}. \tag{3}$$

Note that Theorem 1 shows that the quantity on the left of (3) is at least $(1 - \varepsilon)/r$ for all large q, so Theorem 1 may be close to best possible for values of q with many small prime factors. On the other hand, if q has no small prime factors, then Conjecture 1 suggests that $n_r^*(q)$ is the same order of magnitude as $q \log q$.

Note that $D_r(q)$ is smallest when q has the form $q = \prod_{p \leq z} p^{r-1}$. When z tends to infinity we have

$$D_r(q) = \prod_{p>z} (1 - p^r) \prod_{p\leq z} (1 - p^{-1}) \sim e^{-\gamma}/\log\log q.$$

Conjecture 2 then implies that

$$n_r(q) < (2e^{\gamma} + \varepsilon) q \log q \log\log q$$

for all sufficiently large q. This suggests that there may be very little room for improvement in Theorem 4.

The heuristic arguments for Conjectures 1 and 2 are probabalistic in nature, and are similar to that used by Wagstaff [**11**] for primes in arithmetic progessions. The author wishes to thank D. R. Heath-Brown for correcting an error in the author's original heuristic argument, and also the anonymous referee for suggesting the use of the Borel-Cantelli Lemma and greatly strengthening the conjectures.

Assume first that $(a, q) = 1$. The probability that a randomly selected integer in the arithmetic progession a modulo q will be r-free is $C_r(q)$ (see Cohen and Robinson [**1**]). Hence the probability that none of the numbers $a, a + q, \ldots, a + (w - 1)q$ is r-free is $(1 - C_r(q))^w$. Assuming independence among the residue classes, the probability that every residue class a modulo q with $(a, q) = 1$ contains an r-free number among the first w positive integers in the class is

$$P_q = \left(1 - (1 - C_r(q))^w\right)^{\varphi(q)}.$$

If $w = z \log q / (-\log(1 - C_r(q)))$, then for fixed z we have

$$\lim_{q \to \infty} P_q = \begin{cases} 0 & \text{if } z < 1, \\ 1 & \text{if } z > 1. \end{cases}$$

This is the argument that leads to the second part of Conjecture 1.

If A_q is the event that $n_r^*(q) < wq$, where w is defined above, then $P(A_q) = 1 - P_q$. If $z > 1$, then

$$\begin{aligned} P(A_q) &= 1 - \exp\left(\varphi(q) \log\left(1 - \varphi(q)^{-z}\right)\right) \\ &= 1 - \exp\left(-\varphi(q)^{1-z} + O\left(\varphi(q)^{1-2z}\right)\right) \\ &= \varphi(q)^{1-z} + O\left(\varphi(q)^{2-2z}\right). \end{aligned}$$

If $z > 2$, then $\sum_{q=1}^{\infty} P(A_q) < \infty$, and the Borel-Cantelli Lemma suggests that A_q occurs only finitely many times. If $1 < z < 2$, then $\sum_{q=1}^{\infty} P(A_q) = \infty$, and we expect that A_q occurs infinitely often. This leads to the first part of Conjecture 1.

The argument for Conjecture 2 is similar. If (a, q) is r-free, then by a result of Cohen and Robinson [**1**] the probability that a random integer in the residue class a modulo q will be r-free is

$$D_r(a, q) = \prod_{\substack{p \\ (p^r, q) | a}} \left(1 - (p^r, q) p^{-r}\right).$$

Hence the probability that every residue class a modulo q with (a, q) r-free contains an r-free number among the first w positive integers in the class is

$$P_q = \prod_{\substack{a=1 \\ (a, q)\ r\text{-free}}}^{q} \left(1 - (1 - D_r(a, q))^w\right).$$

Note that $D_r(a,q) \geqslant D_r(q)$, so that

$$P_q \geqslant \left(1 - (1 - D_r(q))^w\right)^q.$$

If $w = z\log q/(-\log(1 - D_r(q)))$ with $z > 2$ fixed, it follows that $\sum_{q=1}^{\infty}(1 - P_q) < \infty$. This leads to Conjecture 2.

Let $p(a,q)$ be the least prime exceeding a in the arithmetic progression a modulo q, and let $P(q) = \max_{(a,q)=1} p(a,q)$. It is interesting to note that when the heuristic argument for the first part of Conjecture 1 is adapted to the case of primes in arithmetic progressions, we arrive at the conjecture that

$$\overline{\lim_{q\to\infty}} \frac{P(q)}{\varphi(q)\log^2 q} = 2.$$

This appears to be in agreement with the numerical data computed by Wagstaff [**11**]. Wagstaff conjectured that the ratio $P(q)/(\varphi(q)\log^2 q)$ is usually near 1, but there were a number of numerical examples where the ratio is closer to 2.

3. The proof of Theorem 1. Let $g(q)$ denote Jacobsthal's function, i.e. $g(q)$ is the least positive integer such that every interval of $g(q)$ consecutive integers contains an integer relatively prime to q. Good upper bounds for $g(q)$ have been proved by Iwaniec [**6**] using sophisticated sieve methods, but here we shall require only the estimate $g(q) \ll q^{\varepsilon}$, which follows easily from the sieve of Eratosthenes.

The proof of Theorem 1 is based on the Chinese Remainder Theorem and the following lemma.

LEMMA 1. *Let* $\varepsilon > 0$ *and* $m = 1 + [(1-\varepsilon)\log q/(r\log\log q)]$. *If* m *is sufficiently large, then there exist primes* $q_1 < \cdots < q_m$ *not dividing* q *such that*

$$(4) \qquad (q_1q_2\cdots q_m)^r < \frac{q}{g(q)+1}.$$

PROOF. Since $g(q) \ll q^{\delta}$ for every $\delta > 0$, it suffices to prove that there exist primes $q_1,\ldots,q_m$ not dividing q such that

$$(q_1q_2\cdots q_m)^r < q^{1-\delta}.$$

It also suffices to treat the case when q is the product of all primes less than z. In this case we have $z \sim \log q$, and we can simply choose m primes between z and $2z$. This is possible because the number of primes between z and $2z$ is asymptotically $z/\log z$, which exceeds m for q sufficiently large. Since $\log 2z < (1 + \varepsilon/2)\log\log q$ and

$$mr < \left(1 - \frac{\varepsilon}{2}\right)\frac{\log q}{\log\log q}$$

for m sufficiently large, it follows that

$$(q_1q_2\cdots q_m)^r < (2z)^{mr} < q^{1-\varepsilon^2/4}.$$

This completes the proof of the lemma.

We now complete the proof of Theorem 1. Let σ be a permutation of the integers $0, 1, \dots, m-1$ such that

$$(5) \qquad \sigma(i) < q_{i+1}^r, \qquad i = 0, 1, \dots, m-1.$$

Define b_σ as the least positive solution of the system of congruences

$$b_\sigma + \sigma(i) \equiv 0 \pmod{q_{i+1}^r}, \qquad i = 0, 1, \dots, m-1.$$

Choose k to be an integer with $(q, k) = 1$ and

$$\frac{q(b_\sigma - 1)}{(q_1 q_2 \cdots q_m)^r} < k < \frac{q b_\sigma}{(q_1 q_2 \cdots q_m)^r}.$$

Now let $a_\sigma = q b_\sigma - k(q_1 q_2 \cdots q_m)^r$. Note that $(a_\sigma, q) = 1$, $0 < a_\sigma < q$, and

$$a_\sigma + \sigma(i) q \equiv 0 \pmod{q_{i+1}^r}, \qquad i = 0, 1, \dots, m-1.$$

It follows that $n_r(a_\sigma, q) \geqslant mq$. It remains to show that each σ gives rise to a different a_σ, and that the number of permutations σ is at least

$$\exp\left(\frac{1-\varepsilon}{r}\left(1 - \frac{\log r}{\log\log q}\right)\log q\right).$$

Let σ and δ be two permutations satisfying (5), and suppose that $a_\sigma = a_\delta$. Let n be an integer such that $\sigma(n) \neq \delta(n)$. Since $a_\sigma = a_\delta$, it follows that $b_\sigma \equiv b_\delta \pmod{q_{n+1}^r}$. Hence $\sigma(n) \equiv \delta(n) \pmod{q_{n+1}^r}$, but this is a contradiction since $0 \leqslant \sigma(n), \delta(n) < q_{n+1}^r$. Therefore the a_σ's are distinct.

Let N be the number of permutations satisfying (5). An easy counting argument shows that

$$N = \prod_{i=0}^{m-1} \min\{m - i, q_{i+1}^r - i\} \geqslant (m - l)!,$$

where l is the least integer such that $q_{l+1}^r > m$. Since $l \leqslant \pi(m^{1/r}) < m^{1/2}$, Stirling's formula yields

$$\log N \sim \log(m!) > \frac{1-2\varepsilon}{r} \frac{\log q}{\log\log q} \log\left(\frac{\log q}{r \log\log q}\right)$$
$$> \frac{1-3\varepsilon}{r} \log q \left(1 - \frac{\log r}{\log\log q}\right)$$

for q sufficiently large. Since ε is arbitrary, Theorem 1 follows.

4. The proof of Theorem 2. The following lemma is probably due to Erdös, but its proof has apparently never appeared in print.

LEMMA 2. *Let p_n be the nth prime. If $\varepsilon > 0$ and n is sufficiently large, then there exist at least $\zeta(r)(1-\varepsilon)n$ consecutive integers each of which is divisible by at least one of the numbers $p_1^r, p_2^r, \dots, p_n^r$.*

PROOF. Let $N = [\zeta(r)(1-\varepsilon)n]$. By the Chinese Remainder Theorem, it suffices to show that there exist residue classes a_i modulo p_i^r such that each of the integers $1, 2, \dots, N$ lies in at least one of the residue classes a_i modulo p_i^r, $i = 1, 2, \dots, n$.

Choose $M = M(\varepsilon)$ such that

$$\prod_{i=M+1}^{\infty} \left(1 - p_i^{-2}\right)^{-1} < 1 + \varepsilon.$$

It then follows that

$$\prod_{i=1}^{M} (1 - p_i^{-r}) < \frac{1 + \varepsilon}{\zeta(r)}.$$

From the interval $[1, N]$, remove all integers in a residue class a_1 modulo p_1^r, where a_1 is chosen so as to remove as many integers as possible. Then choose a_2 modulo p_2^r so as to remove as many of the remaining integers as possible, and continue in this way for the first M primes. Since at each stage we remove at least the average number of integers in a residue class, the number S of integers remaining after this process satisfies

$$S < N \prod_{i=1}^{M} (1 - p_i^{-r}) < (1 - \varepsilon^2) n.$$

We can then use one prime for each of the remaining S integers, and in this way we can "sieve out" the entire interval using at most $M + (1 - \varepsilon^2)n$ primes. It then suffices to take $n > M\varepsilon^{-2}$.

The proof of Theorem 2 is similar to that of Theorem 1. Let

$$m = \left[\frac{(1 - \varepsilon)\log q}{r \log\log q}\right],$$

and let $q_1, q_2, \ldots, q_m$ be the first m primes. By the prime number theorem,

$$\log(q_1 \cdots q_m) < (1 + \varepsilon) m \log m < \frac{1 - \varepsilon^2}{r} \log q.$$

Since $g(q) \ll q^{\varepsilon}$, it follows that (4) holds. Let $n = [\zeta(r)(1 - \varepsilon)m]$, and choose b in such a way that each of the integers $b + 1, b + 2, \ldots, b + n$ is divisible by at least one of $q_1^r, \ldots, q_m^r$. Then choose k with $(k, q) = 1$ and

$$\frac{q(b - 1)}{(q_1 q_2 \cdots q_m)^r} < k < \frac{qb}{(q_1 q_2 \cdots q_m)^r}.$$

With $a = bq - k(q_1 q_2 \cdots q_m)^r$, it follows that

$$n_r(a, q) \geqslant nq > (1 - 3\varepsilon)\frac{\zeta(r)}{r} q \frac{\log q}{\log\log q}.$$

5. The proof of Theorem 3. Let $q_1, q_2, \ldots$ be the primes that do not divide a. For m large, choose q such that

$$q \equiv 1 \pmod{a},$$
$$nq \equiv -a \pmod{q_n^r}, \qquad n = 1, 2, \ldots, m,$$
$$a(q_1 q_2 \cdots q_m)^r < q \leqslant 2a(q_1 q_2 \cdots q_m)^r.$$

It then follows that $(a, q) = 1$, $a < q$, and $n_r(a, q) > mq$. By the prime number theorem,

$$\frac{1}{r}\log q \sim \log(q_1 q_2 \cdots q_m) \sim m \log m.$$

Hence for m sufficiently large we have

$$m > \frac{1-\varepsilon}{r}\frac{\log q}{\log\log q}.$$

6. The proof of Theorem 4. The idea behind the proof of Theorem 4 is that if (a, q) is r-free but divisible by many small primes to the power $r - 1$, then a number of the form $a + nq$ is "close" to being divisible by an rth power of a prime.

Let p_n be the nth prime, and define $Q_m = p_1 p_2 \cdots p_m$. A result of Rankin **[10]** states that

$$g(Q_m) > (e^\gamma - \varepsilon)\log Q_m \log_2 Q_m \frac{\log_4 Q_m}{(\log_3 Q_m)^2} \tag{6}$$

if m is sufficiently large. Choose b such that $(b, Q_m) = 1$ and each of the integers $b + 1, b + 2, \ldots, b + g(Q_m) - 1$ is divisible by at least one of the first m primes. Determine q' by

$$bq' \equiv p_{m+1}^r \pmod{Q_m}, \qquad q' \equiv 1 \pmod{p_{m+1}},$$
$$p_{m+1}^r < q' \leqslant p_{m+1}^r + p_{m+1}Q_m.$$

It follows that

$$q'n + p_{m+1}^r \equiv q'(n + b) \pmod{Q_m}.$$

Let $a = p_{m+1}^r Q_m^{r-1}$ and $q = q'Q_m^{r-1}$. We now have $(a, q) = Q_m^{r-1}$ is r-free, $a < q$, and $a + nq$ is not r-free for $0 \leqslant n \leqslant g(Q_m) - 1$. Hence $n_r(a, q) \geqslant g(Q_m)q$. Finally, we have $Q_{m+1}^{r-1} < q < Q_{m+1}^r$, so that

$$\log Q_m > \frac{1-\varepsilon}{r}\log q$$

for m sufficiently large.

References

1. E. Cohen and R. L. Robinson, *On the distribution of the k-free integers in residue classes*, Acta Arith. **8** (1962/63), 283–293.

2. P. Erdös, *Über die kleinste quadratfreie Zahl einer arithmetischen Reihe*, Monatsh. Math. **64** (1960), 314–316.

3. D. R. Heath-Brown, *The least square-free number in an arithmetic progression*, J. Reine Angew. Math. **332** (1982), 204–220.

4. C. Hooley, *A note on square-free numbers in arithmetic progressions*, Bull. London Math. Soc. **7** (1975), 133–138.

5. M. Huxley, *The difference between consecutive primes*, Analytic Number Theory, Proc. Sympos. Pure Math., vol. 24, Amer. Math. Soc., Providence, R. I., 1973, pp. 141–145.

6. H. Iwaniec, *On the problem of Jacobsthal*, Demonstratio Math. **11** (1978), 225–231.

7. C. Pomerance, *A note on the least prime in an arithmetic progression*, J. Number Theory **12** (1980), 218–223.

8. K. Prachar, *Über die kleinste quadratfreie Zahl einer arithmetischen Reihe*, Monatsh. Math. **62** (1958), 173–176.

9. ______, *Über die kleinste Primzahl einer arithmetischen Reihe*, J. Reine Angew. Math. **206** (1961), 3–4.

10. R. A. Rankin, *The difference between consecutive prime numbers*. V, Proc. Edinburgh Math. Soc. (2) **13** (1962/63), 331–332.

11. S. S. Wagstaff, *Greatest of the least primes in arithmetic progressions having a given modulus*, Math. Comp. **33** (1979), 1073–1080.

12. R. Warlimont, *On squarefree numbers in arithmetic progressions*, Monatsh. Math. **73** (1969), 433–448.

13. ______, *Über die kleinsten quadratfreien Zahlen in arithmetischen Progessionen*, J. Reine Angew. Math. **250** (1971), 99–106.

Department of Mathematics, Michigan State University, East Lansing, Michigan 48824

Current address: Department of Mathematics, DRB306, University of Southern California, Los Angeles, California 90089-1113