MATH. SCAND. 34 (1974), 37–43

SETS OF PRIMES WITH INTERMEDIATE DENSITY

H. G. MEIJER

1. Introduction.

Let P denote the sequence of primes. A subsequence $\{q_n\}$ of P satisfying $3 \leqq q_1 < q_2 < \ldots$ and

$$(1) \qquad q_n \not\equiv 1 \pmod{q_i}, \quad 1 \leqq i < n,\ n \geqq 2 ,$$

will be called here a G-sequence. For a sequence $A = \{a_n\}$ we denote by $A(A,x)$ the number of elements of A not exceeding x and we put

$$P(A,x) = \prod\nolimits_{a_n \leqq x} (1-a_n^{-1})^{-1}, \quad S(A,x) = \sum\nolimits_{a_n \leqq x} a_n^{-1} .$$

In [2] S. W. Golomb pointed out the importance of G-sequences. He especially studied their density and he proved that there does not exist a constant $A > 0$ such that

$$A(G,x) > Ax/\log x \quad \text{for all sufficiently large } x .$$

In view of this property he said that the G-sequences are of "intermediate density". A special example is the sequence G_1 defined inductively by $q_1 = 3$ and q_n for $n \geqq 2$ is the smallest prime greater than q_{n-1} for which $q_n \not\equiv 1 \pmod{q_i}$, $1 \leqq i < n$. Erdös [1] proved for the sequence G_1

$$(2) \qquad A(G_1,x) = (1+o(1))x(\log x \log\log x)^{-1} ,$$

$$(3) \qquad \log\log\log x - c_1 < S(G_1,x) < \log\log\log x + c_1 ,$$

for some constant c_1.

Investigation of the equation $k\varphi(M) = M-1$ where φ is Euler's totient function also leads in a natural way to the study of G-sequences; see e.g. Lieuwens [4]. Numerical computations lead Lieuwens to conjecture in [4] that

$$\lim\nolimits_{x\to\infty} P(G,x) < 3$$

for every G-sequence. We remark that Erdös result (3) implies that this conjecture is false. In fact we have

$$(4) \qquad \log P(G,x) = \sum -\log(1-q_i^{-1}) > \sum q_i^{-1} = S(G,x)$$

Received August 16, 1972.

and therefore by (3)

(5) $$P(G_1,x) > c_2 \log\log x \quad \text{for some constant } c_2 > 0 .$$

In section 2 of this paper we give some notes on G-sequences and in section 3 a generalization is stated.

2. G-sequences.

The computations in this section are based on the following lemma.

LEMMA. *For every G-sequence*

$$A(G,x)P(G,x) \leqq cx(\log x)^{-1} \quad \textit{for } x \geqq 3 ,$$

where c is a constant independent of x and the sequence G.

PROOF. Let $N(G,x)$ denote the number of integers $1 \leqq z \leqq x$ satisfying $z \not\equiv 0 \pmod{p_i}$ for all primes $p_i \leqq \sqrt{x}$ not occurring in G and $z \not\equiv 0,1 \pmod{q_i}$ for all primes $q_i \leqq \sqrt{x}$ occurring in G. It follows from the Brun-Selberg sieve method (compare [1, lemma 2] and [3, p. 214, (7.25)]) that

$$N(G,x) \leqq c_3 x \prod_{p_i \leqq \sqrt{x}} (1 - r_i p_i^{-1}) ,$$

where $r_i = 1$ if p_i does not occur in G, $r_i = 2$ if p_i occurs in G and c_3 is a constant. Therefore

$$\begin{aligned} N(G,x) &\leqq c_3 x \prod_{p_i \leqq \sqrt{x}} (1 - p_i^{-1})^{r_i} \\ &= c_3 x P(P, \sqrt{x})^{-1} P(G, \sqrt{x})^{-1} \\ &\leqq c_3 x P(P, \sqrt{x})^{-1} P(G,x)^{-1} \prod_{\sqrt{x} < p \leqq x} (1 - p^{-1})^{-1} , \end{aligned}$$

where the last product is extended over all primes $\sqrt{x} < p \leqq x$. It is well-known (see e.g. [5, p. 20]) that

(6) $$P(P,x) = c_4 \log x + O(1) \quad \text{as } x \to \infty$$

for some constant c_4. Therefore there is a constant c_5 such that

$$N(G,x) \leqq c_5 x (\log x)^{-1} P(G,x)^{-1} .$$

If q_n is a prime of G with $\sqrt{x} < q_n \leqq x$, then $q_n \not\equiv 0 \pmod{p_i}$ for all primes $p_i \leqq \sqrt{x}$ not occurring in G and $q_n \not\equiv 0,1 \pmod{q_i}$ for all primes $q_i \leqq \sqrt{x}$ of G. Hence

$$\begin{aligned} A(G,x) &\leqq \sqrt{x} + N(G,x) \leqq \sqrt{x} + c_5 x (\log x)^{-1} P(G,x)^{-1} \\ &\leqq c x (\log x)^{-1} P(G,x)^{-1} \end{aligned}$$

for a suitable constant c. Since $P(G,x) \leqq P(P,x)$ the constant c may be chosen independent of G.

The lemma enables us to sharpen the above mentioned result of Golomb in the following sense.

THEOREM 1. *Let G be a G-sequence. There does not exist a constant $A > 1$ such that*

$$A(G,x) > \frac{Ax}{\log x \log\log x} \quad \textit{for all sufficiently large } x\,. \tag{7}$$

PROOF. By (4) and the lemma we have

$$S(G,x) \leqq \log P(G,x) \leqq \log\left\{\frac{cx}{\log x A(G,x)}\right\}.$$

On the other hand we have by partial summation

$$S(G,x) = \textstyle\sum_{n\leqq x} n^{-1}(A(G,n)-A(G,n-1)) \geqq \sum_{n\leqq x} A(G,n)(n^{-1}-(n+1)^{-1})\,.$$

Therefore we get

$$\sum_{n\leqq x}\frac{A(G,n)}{n(n+1)} \leqq \log\left\{\frac{cx}{\log x A(G,x)}\right\}. \tag{8}$$

Suppose that (7) holds for some $A > 1$ and $x \geqq x_0$. Then the right-hand side of (8) is less than or equal to $\log\log\log x + \log(cA^{-1})$ for $x \geqq x_0$, while the left-hand side is greater than or equal to $\frac{1}{2}(A+1)\log\log\log x$ if x is sufficiently large. This is a contradiction.

In view of theorem 1 and (2) one might suspect

$$A(G,x) \leqq (1+o(1))x(\log x \log\log x)^{-1}$$

for every G-sequence. This, however, is not true as can be seen from the following theorem.

THEOREM 2. *There exists a G-sequence and a constant $c > 0$ such that*

$$A(G,x) > cx(\log x)^{-1} \tag{9}$$

for infinitely many positive integers x.

PROOF. We will construct a sequence of positive integers $\{x_k\}$ such that $2 = \frac{1}{2}x_1 < x_1 < \frac{1}{2}x_2 < x_2 < \frac{1}{2}x_3 < x_3 < \ldots$ and a G-sequence entirely contained in $\bigcup_{k=1}^{\infty}(\frac{1}{2}x_k, x_k]$ such that (9) holds for this G-sequence and the integers x_k.

Suppose that we have already chosen $x_1,\dots,x_{k-1}$ $(k \geqq 2)$ and the primes $3=q_1,\dots,q_n$ of the sequence G contained in $\bigcup_{i=1}^{k-1}(\frac{1}{2}x_i,x_i]$. The primes p with

(10) $$p \not\equiv 1 \pmod{q_i} \quad 1 \leqq i \leqq n$$

lie in $(q_1-2)\dots(q_n-2)$ residue classes $\bmod q_1\dots q_n$. Therefore the number of primes $p \leqq x$ satisfying (10) is asymptotically equal to (see e.g. [5, p. 138])

$$\frac{x}{\log x}\prod_{i=1}^{n}\frac{q_i-2}{q_i-1}.$$

Then we can choose an integer $x_k > 2x_{k-1}$ such that the number of primes p with $\frac{1}{2}x_k < p \leqq x_k$ satisfying (10) is greater than

$$\tfrac{1}{4}\frac{x_k}{\log x_k}\prod_{i=1}^{n}\frac{q_i-2}{q_i-1}.$$

By (6) we have $\lim_{x\to\infty}\prod_{\frac{1}{2}x<p\leqq x}(1-(p-1)^{-1})=1$, where the product is extended over all primes in $(\frac{1}{2}x,x]$. Hence we may choose x_k so large that

(11) $$\prod_{\frac{1}{2}x_k<p\leqq x_k}(1-(p-1)^{-1}) = 1-\varepsilon_k \quad \text{with } \varepsilon_k < k^{-2}.$$

Now we choose x_k so large that both conditions are satisfied and we continue the sequence G with the primes in $(\frac{1}{2}x_k,x_k]$ satisfying (10). Then

$$A(G,x_k) > \tfrac{1}{4}\frac{x_k}{\log x_k}\prod_{i=1}^{n}\frac{q_i-2}{q_i-1}$$

and since, by (11), the product is convergent, the theorem follows.

REMARK. As a consequence of theorem 2 and (2) we conclude that there exist G-sequences such that $A(G,x) > A(G_1,x)$ for infinitely many positive integers x.

In view of the difference between (2) and theorem 2 it is interesting to remark that Erdös proof of the upper bound in (3) still holds for an arbitrary G-sequence. We will give here, however, another proof following the method of theorem 1.

THEOREM 3. *There exist constants a and b such that for every G-sequence the following inequalities holds for $x \geqq 3$*

$$P(G,x) \leqq a \log\log x$$
$$S(G,x) \leqq \log\log\log x + b.$$

PROOF. By (4) we only have to prove the inequality for $P(G,x)$. First we remark that for $0 \leqq t \leqq \frac{1}{2}$ we have $-\log(1-t) \leqq 2t$. Therefore we get if y and z are integers satisfying $3 \leqq y < z$,

$$\begin{aligned}\log P(G,z) - \log P(G,y) &= \sum_{y<q_i\leqq z} -\log(1-q_i^{-1}) \\ &\leqq 2\sum_{y<q_i\leqq z} q_i^{-1} \\ &= 2\sum_{y<n\leqq z} n^{-1}\big(A(G,n)-A(G,n-1)\big)\,.\end{aligned}$$

By partial summation we find

$$\begin{aligned}&\log P(G,z) - \log P(G,y) \leqq 2\sum_{y<n\leqq z-1} A(G,n)\big(n^{-1}-(n+1)^{-1}\big) - \\ &\quad -2(y+1)^{-1}A(G,y) + 2z^{-1}A(G,z) \leqq 2\sum_{y<n\leqq z} n^{-2}A(G,n) + 2z^{-1}A(G,z)\,.\end{aligned}$$

Then the lemma gives the following inequality for $P(G,x)$

$$\begin{aligned}\log P(G,z) - \log P(G,y) &\leqq 2c\sum_{y<n\leqq z}\big(n\log nP(G,n)\big)^{-1} + \\ &\quad + 2c\big(\log zP(G,z)\big)^{-1}\,.\end{aligned} \tag{12}$$

Put

$$c_6 = \max\big(2c, 3(2\log\log 3)^{-1}\big)\,. \tag{13}$$

Since $P(G,3) = 1$ or $\frac{3}{2}$ we have

$$P(G,3) \leqq 3(2\log\log 3)^{-1}\log\log 3 \leqq c_6\log\log 3\,. \tag{14}$$

Choose a real number $u > 1$ such that

$$\log u > (\log 3)^{-1}(\log\log 3)^{-1}\,. \tag{15}$$

We shall prove

$$P(G,x) \leqq c_6 u\log\log x \quad \text{for all integers } x \geqq 3\,.$$

Obviously this will prove the theorem.

Suppose that there exists an integer $z > 3$ with

$$P(G,z) > c_6 u\log\log z\,. \tag{16}$$

Then, by (14) there exists an integer $y \geqq 3$ such that

$$P(G,y) \leqq c_6\log\log y\,, \tag{17}$$

$$P(G,n) > c_6\log\log n \quad \text{for all integers } n \text{ with } y < n \leqq z\,. \tag{18}$$

From (16) and (17) it follows that the left-hand side of (12) is greater than

$$\log\log\log z - \log\log\log y + \log u\,.$$

On the other hand by (13), and (18) the right-hand side of (12) is less than or equal to

$$\sum_{y<n\leqq z} (n \log n \log\log n)^{-1} + (\log z \log\log z)^{-1}$$
$$\leqq \log\log\log z - \log\log\log y + (\log 3 \log\log 3)^{-1} .$$

By the choice (15) of u this is a contradiction.

REMARK. It follows from (3) and (5) that the upper bounds of theorem 3 are best possible.

3. Generalization.

As a matter of fact the condition $q_n \not\equiv 1 \pmod{q_i}$ in (1) can be replaced by $q_n \not\equiv a_i \pmod{q_i}$ where $\{a_i\}$ is a sequence of integers such that $a_i \not\equiv 0 \pmod{q_i}$ for every positive integer i.

It is also easy to generalize to the case that with every prime from the sequence a set of k sifting classes is associated. Then we get the following situation. Let k be a positive integer. Let $Q = \{q_n\}$ denote a sequence of primes and

$$\{a_{nh} :\ n = 1, 2, \ldots ; h = 1, \ldots, k\}$$

a double sequence of integers such that for every n the integers $0, a_{n1}, \ldots, a_{nk}$ are incongruent $\bmod q_n$. Let moreover

$$k+2 \leqq q_1 < q_2 < \ldots$$

and

$$q_n \not\equiv a_{ih} \pmod{q_i} \quad h = 1, \ldots, k;\ 1 \leqq i < n,\ n \geqq 2 .$$

It is easy to derive that the lemma had to be replaced by

$$A(Q,x) P^k(Q,x) \leqq cx(\log x)^{-1} \quad \text{for } x \geqq 3 .$$

This implies that we get instead of theorem 1 that there does not exist a constant $A > k^{-1}$ such that

$$A(Q,x) > Ax(\log x \log\log x)^{-1} \quad \text{for all sufficiently large } x$$

and the sequence Q is of intermediate density following the definition of Golomb. On the other hand, as in theorem 2, there exists a sequence Q and a constant $c > 0$ such that

$$A(Q,x) > cx(\log x)^{-1}$$

for infinitely many positive integers x.

Finally we get instead of theorem 3 that there exist constants a and b such that for $x \geqq 3$

$$P(Q,x) \leqq a(\log\log x)^{1/k}$$
$$S(Q,x) \leqq k^{-1} \log\log\log x + b\,.$$

REFERENCES

1. P. Erdös, *On a problem of G. Golomb*, J. Austral. Math. Soc. 2 (1961/62), 1–8.
2. S. W. Golomb, *Sets of primes with intermediate density*, Math. Scand. 3 (1955), 264–274.
3. H. Halberstam, K. F. Roth, *Sequences* vol. I, Clarendon Press, Oxford, 1966.
4. E. Lieuwens, *Do there exist composite numbers M for which $k\varphi(M)=M-1$ holds?*, Nieuw Arch. Wisk. (3) 18 (1970), 165–169.
5. K. Prachar, *Primzahlverteilung*, (Grundlehren math. Wissensch. 41) Springer-Verlag, Berlin, 1957.

UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF MATHEMATICS
DELFT, NETHERLANDS