ON A PROBLEM OF G. GOLOMB

P. ERDÖS

(received 3 August 1960)

In his paper on sets of primes with intermediate density Golomb [1] proved the following theorem:

Let $2 < P_1 < P_2 < \cdots$ be any sequence of primes for which

$$P_j \not\equiv 1 \pmod{P_i} \tag{1}$$

for every i and j. Denote by $A(x)$ the number of P's not exceeding x. Then

$$\liminf_{x=\infty} A(x)/x = 0. \tag{2}$$

It is not difficult to see that in some sense (2) is best possible since it is easy to construct a sequence of primes satisfying (1) for which

$$\limsup_{x=\infty} A(x)/x > 0,$$

and in fact the lim sup can be as close to 1 as we wish. Golomb pointed out that in some ways the most natural sequence satisfying (1) can be obtained as follows: $q_1 = 3$, $q_2 = 5$, $q_3 = 17, \cdots q_k$ is the smallest prime greater than q_{k-1} for which

$$q_k \not\equiv 1 \pmod{q_i}, \qquad 1 \leqq i < k.$$

Henceforth we will only consider this special sequence satisfying (1). We shall prove the following (as before $A(x)$ denotes the number of $q_i \leqq x$).

THEOREM.

$$A(x) = (1 + o(1)) \frac{x}{\log x \operatorname{loglog} x}.$$

$\log_k x$ will denote the k times iterated logarithm, $c_1, c_2, \cdots$ will denote positive absolute constants.

Our method will be similar to the one used in our recent joint paper with Jabotinsky [2], but we will also need Brun's method and the results on primes in short arithmetic progressions.

[1] S. Golomb, Math. Scand. 3 (1955), 264—74.

[2] P. Erdös and E. Jabotinsky, Indig. Math. 20 (1958), 115-128.

LEMMA 1. Denote by $\pi(x, k, l)$ the number of primes $p \leqq x$, $p \equiv l \pmod{k}$, $(l, k) = 1$. Then ($\exp z = e^z$)

$$\pi(x, k, l) = \frac{x}{\varphi(k) \log x}\left(1 + O\left(\frac{1}{\log x}\right)\right) \tag{3}$$

uniformly for all $k < \exp(c_1 \log x/\log\log x)$, except possibly for the multiples of a certain $k^* = k^*(x)$ where $k^* > (\log x)^A$ (A is an arbitrary constant, but the constant in $O(1/\log x)$ depends on A).

Lemma 1 is well known [3].

LEMMA 2. Let $2 = p_1 < p_2 < \cdots$ be the sequence of consecutive primes, and let r be a fixed integer, $0 \leqq r_i < r$. Denote by $N_k(x)$ the number of integers $1 \leqq z \leqq x$ for which $z \equiv l \pmod{k}$, $(l, k) = 1$ and

$$z \not\equiv a_i^{(j)} \pmod{p_i}, \qquad 1 \leqq j \leqq r_i$$

where the $a_i^{(j)}$ are arbitrary residues and $p_i \leqq x$. Then

$$N_k(x) < c_2 \frac{x}{k} \prod_{p_i \leqq x/k} (1 - r_i/p_i).$$

The proof follows immediately from Brun's method [4].

LEMMA 3. There exists a constant c_3 so that

$$\log_3 x - c_3 < \sum_{q_i \leqq x} 1/q_i < \log_3 x + c_3. \tag{4}$$

First we prove the upper bound. If the upper estimation in (4) would not hold then for every c there would be arbitrarily large values of x so that for every $z < x$

$$\sum_{q_i \leqq x} \frac{1}{q_i} - \log_3 x > \sum_{q_i \leqq z} \frac{1}{q_i} - \log_3 z \tag{5}$$

and

$$\sum_{q_i \leqq x} \frac{1}{q_i} > \log_3 x + c. \tag{6}$$

Let $x^{1/2} < q_i \leqq x$. Clearly by the definition of the q's $q_i \not\equiv 0 \pmod{p}$ for all $p < x^{1/2}$ and $q_i \not\equiv 1 \pmod{q_j}$ for $q_j < x^{1/2}$. Thus by lemma 2 ($k = 1$)

$$A(x) < x^{1/2} + c_2 x \prod_{p_i \leqq x^{1/2}} (1 - r_i/p_i) \tag{7}$$

where $r_i = 2$ if p_i is a q and is 1 otherwise. From (7), (6) and from $\prod_{p < x^{1/2}} (1 - 1/p) < c_4/\log x$

$$A(x) < c_5 \frac{x}{\log x} \prod_{q_i < x^{1/2}} \left(1 - \frac{1}{q_i}\right) < c_6 x \exp(-c)/\log x \log_2 x. \tag{8}$$

[3] This is Theorem 2.3 p. 230 of Prachar's book Primzahlverteilung (Springer 1957) where the literature of this question can be found.

[4] See e.g. P. Erdös, Proc. Cambridge Phil. Soc. 34 (1957), 8.

The last inequality in (8) follows from $\prod_{q_i<z}(1-1/q_i) < c_7 \exp(-\sum_{q_i<z} 1/q_i)$ and from (using (6))

$$\sum_{q_i<x^{1/2}} \frac{1}{q_i} > \sum_{q_i\leq x}\frac{1}{q_i} - \sum_{x^{1/2}\leq p\leq x}\frac{1}{p} > \log_3 x + c - c_8 .$$

From (8) we have

$$\sum_{x/2<q_i\leq x} \frac{1}{q_i} < \frac{2A(x)}{x} < 2c_6 \exp(-c)/\log x \log\log x. \tag{9}$$

But from (5) we have for $z = x/2$

$$\sum_{x/2<q_i\leq x} \frac{1}{q_i} > \log_3 x - \log_3 \frac{x}{2} > c_9/\log x \log_2 x,$$

which contradicts (9) for sufficiently large c. Thus the upper bound in (4) is proved.

The proof of the lower bound will be more complicated. Put $y = \exp(\log x/(\log\log x)^{10})$ and denote by $A_y(x)$ the number of primes $p \leq x$ satisfying

$$p \not\equiv 1 \pmod{q_i}, \qquad 3 \leqq q_i \leqq y. \tag{10}$$

We evidently have

$$A_y(x) - \sum_{y<q_j<x} B(x, q_j) < A(x) < A_y(x) + y \tag{11}$$

where $B(x, q_j)$ denotes the number of primes $p \leqq x$ satisfying

$$p \equiv 1 \pmod{q_j}, \qquad p \not\equiv 1 \pmod{q_i}, \qquad 3 \leqq q_i \leqq y.$$

Now we estimate $A_y(x)$ by Brun's method.

LEMMA 4.

$$A_y(x) = (1+o(1))\frac{x}{\log x}\prod_{q_i\leq y}\left(1-\frac{1}{q_i-1}\right).$$

By the sieve of Eratosthenes we have

$$A_y(x) = \pi(x) - \sum \pi(x, q_i, 1) + \sum \pi(x, q_{i_1} q_{i_2}, 1) - \cdots$$

where $3 \leqq q_i \leqq y$ and i's are distinct. By the well known idea of Brun[5] we have ($\Sigma_r = \sum \pi(x, q_{i_1}\cdot q_{i_2}\cdots q_{i_r}, 1)$)

$$\pi(x)-\Sigma_1+\Sigma_2-\Sigma_3+\cdots-\Sigma_{2k-1} < A_y(x) < \pi(x)-\Sigma_1+\Sigma_2-\cdots+\Sigma_{2k}. \tag{12}$$

We now choose $k = [10 \log_3 x]$. We distinguish two cases. In the first case none of the numbers $q_{i_1}\cdots q_{i_r}$, $1 \leqq r \leqq 2k$ are exceptional from the point

[5] See e.g. E. Landau, Zahlentheorie Vol. 1.

of view of Lemma 1. In this case we can estimate Σ_r by Lemma 1 and following say Landau's treatment of Brun's method[5] we obtain from (12) by a simple computation

$$(13)\quad A_y(x) = \frac{x}{\log x} \prod_{3 \leq q_i \leq y}\left(1 - \frac{1}{q_i - 1}\right) + O\left(\frac{x}{(\log x)^2}\right) \prod_{3 \leq q_i \leq y}\left(1 + \frac{1}{q_i - 1}\right) + o\left(\frac{x}{(\log x)^2}\right).$$

By the upper bound of (4) we have

$$\prod_{q_i \leq y}\left(1 + \frac{1}{q_i - 1}\right) < c_{10} \log_2 x \text{ and } \prod_{q_i \leq y}\left(1 - \frac{1}{q_i - 1}\right) > c_{10}/\log_2 x,$$

thus from (13) we obtain Lemma 4 in the first case.

In the second case let $d = q_{i_1} \cdot q_{i_2} \cdots q_{i_r}$ be the smallest exceptional number (i.e. for which Lemma 1 does not hold). By Lemma 1 we can assume that $d > (\log x)^A$. We estimate $\pi(x, td, 1)$ from below by 0 and from above by x/td. Since

$$\sum_{t<x} \frac{x}{td} = O\left(\frac{x \log x}{d}\right) = o\left(\frac{x}{(\log x)^2}\right)$$

we can neglect this exceptional d and the proof of Lemma 4 is complete.

Now we complete the proof of Lemma 3. Assume that the lower bound in (4) is false. Then for every c_3 there are infinitely many integers x satisfying for every $z \leqq x$

$$(14)\qquad \sum_{q_i \leqq x} \frac{1}{q_i} - \log_3 x < \sum_{q_i \leqq z} \frac{1}{q_i} - \log_3 z$$

and

$$(15)\qquad \sum_{q_i \leqq x} \frac{1}{q_i} = \log_3 x - c_x, \qquad c_x > c_3.$$

From (14) we have

$$(16)\qquad \sum_{z < q_i < x} \frac{1}{q_i} < \log_3 x - \log_3 z.$$

By Lemma 4 and (16) (since $\log_3 x - \log_3 y = o(1)$)

$$(17)\quad A_y(x) - A_y\left(\frac{x}{2}\right) = (1 + o(1)) \frac{x}{2 \log x} \prod_{q_i \leq y}\left(1 - \frac{1}{q_i - 1}\right) > c_{11} \frac{x \exp c_x}{\log x \log_2 x}.$$

Thus from (11) and (17)

$$(18)\qquad A(x) - A\left(\frac{x}{2}\right) > c_{11} \frac{x \exp c_x}{\log x \log_2 x} - y - \sum_{y < q_j \leqq x} B(x, q_j).$$

Now we estimate $\sum_{y<q_j \leqq x} B(x, q_j)$. Write

[5] See e.g. E. Landau, Zahlentheorie Vol. 1.

$$\sum_{y<q_j\leq x} B(x, q_j) = \Sigma_1 + \Sigma_2 + \Sigma_3 \tag{19}$$

where in Σ_1 $y < q_j \leqq x \exp(-\log x/(\log_2 x)^{1/2})$, in Σ_2 $x \exp(-\log x/(\log_2 x)^{1/2}) < q_j \leqq x \exp(-\log x/(\log_2 x)^{5/4})$ and in Σ_3 $x \exp(-\log x/(\log_2 x)^{5/4}) < q_j \leqq x$. From Lemma 2 we have for the q_j in Σ_1 and Σ_2

$$B(x, q_j) < c_2 \frac{x}{q_j \log \frac{x}{q_j}} \prod{}' \left(1 - \frac{1}{q_i}\right) < c_2 \frac{x}{q_j \log \frac{x}{q_j}} \prod_{q_i<y} \left(1 - \frac{1}{q_i}\right) \tag{20}$$

where in $\prod'$ $q_i < \min(q_j, x/q_j)$. (20) holds since for the q_j in Σ_1 and Σ_2 $\min(q_j, x/q_j) > y$. Now from (16) $\sum_{y<q_i\leq x} 1/q_i < \log_2 x - \log_2 y = o(1)$. Thus from (15)

$$\sum_{q_i<y} \frac{1}{q_i} = \log_3 x - c_x - o(1). \tag{21}$$

From (20) and (21) we have for the q_j in Σ_1

$$B(x, q_j) < c_{12} \frac{x \exp c_x}{q_j \log \frac{x}{q_j} \log_2 x} < c_{12} \frac{x \exp c_x}{q_j \log x (\log_2 x)^{1/2}}. \tag{22}$$

But from (16)

$$\Sigma_1 \frac{1}{q_j} \leqq \sum_{y<q_j\leq x} \frac{1}{q_j} < \log_2 x - \log_2 y < c_{13} \log_3 x/\log_2 x. \tag{23}$$

Thus from (22) and (23)

$$\Sigma_1 < c_{12} \frac{x \exp c_x}{\log x\, (\log_2 x)^{1/2}} \Sigma_1 \frac{1}{q_j} < c_{12} c_{13} \frac{x \log_3 x \exp c_x}{\log x\, (\log_2 x)^{3/2}} = o\left(\frac{x \exp c_x}{\log x \log_2 x}\right). \tag{24}$$

Again from (20), (21) and (16) we obtain as in the estimation

$$\Sigma_2 < c_{14} \frac{x (\log_2 x)^{1/4} \exp c_x}{\log x} \Sigma_2 \frac{1}{q_j} < c_{14} c_{15} \frac{x \exp c_x}{\log x (\log_2 x)^{5/4}} = o\left(\frac{x \exp c_x}{\log x \log_2 x}\right). \tag{25}$$

To estimate Σ_3 denote by $N(a, x)$ the number of primes $p < x/a$, $a < x^{1/2}$, for which $a \cdot p + 1$ is also a prime. A well known consequence of Brun's method implies that

$$N(a, x) < c_{16} \frac{x}{(\log x)^2} \prod_{p/a} \left(1 + \frac{1}{p}\right). \tag{26}$$

(26) easily follows from Lemma 2. From (26) we have by interchanging the order of summation ($\sum'$ denotes that $1 \leqq a < \exp(\log x/(\log_2 x)^{5/4})$)

$$(27)\quad \Sigma_3 \leqq \sum{}' N(a, x) < c_{16} \frac{x}{(\log x)^2} \sum{}' \frac{\prod_{p/a}\left(1 + \frac{1}{p}\right)}{a} < c_{17} \frac{x}{\log x (\log_2 x)^{5/4}}.$$

The last inequality of (27) holds since it is well known that

$$(28)\quad \sum_{a=1}^{z} \frac{\prod_{p/a}\left(1 + \frac{1}{p}\right)}{a} < c_{18} \log z.$$

((28) follows easily from the well known result $\sum_{a=1}^{z} \prod_{p/a} (1 + 1/p) < \sum_{a=1}^{z} \sigma(a)/a = (1 + o(1))\pi^2/6 \log z$ by partial summation), From (24), (25) and (27) we obtain

$$(29)\quad \sum_{v \leqq q_j \leqq x} B(x, q_j) = o\left(\frac{x \exp c_x}{\log x \log_2 x}\right).$$

From (18) and (29) we have

$$(30)\quad A(x) - A\left(\frac{x}{2}\right) > c_{19} \frac{x \exp c_x}{\log x \log_2 x}.$$

(30) implies that

$$(31)\quad \sum_{(x/2) < q_i < x} \frac{1}{q_i} > c_{19} \exp c_x/\log x \log_2 x.$$

On the other hand (16) implies that

$$\sum_{(x/2) < q_i < x} \frac{1}{q_i} < \log_3 x - \log_3 \frac{x}{2} < c_{20}/\log x \log_2 x)$$

an evident contradiction for sufficiently large c_3 $(c_x > c_3)$. Thus the upper bound of (4) is proved and the proof of Lemma 3 is complete.

From the upper bound in (4), (19), (24), (25) and (27) we immediately obtain (we now know that $c_x < c_3$)

$$(32)\quad \sum_{v \leqq q_j \leqq x} B(x, q_j) = o\left(\frac{x}{\log x \log_2 x}\right).$$

From (11), (32) and Lemmas 3 and 4 we obtain

$$(33)\quad \begin{aligned} A(x) &= (1 + o(1)) \frac{x}{\log x} \prod_{q_i \leqq v} \left(1 - \frac{1}{q_i - 1}\right) + o\left(\frac{x}{\log x \log_2 x}\right) \\ &= (1 + o(1)) \frac{x}{\log x} \prod_{q_i \leqq v} \left(1 - \frac{1}{q_i - 1}\right). \end{aligned}$$

The last inequality of (33) follows, since by the lower bound in (4) $\prod_{q_i \leqq v} (1 - 1/(q_i - 1)) > c_{21}/\log_2 x$. From (33) and the lower bound in (4)

(34) $A(x) < c_{22}\, x/\log x \log_2 x$ (since $\prod_{q_i<y}\left(1-\frac{1}{q_i-1}\right) < c_{23}/\log_2 x$).

Thus by a simple computation

(35) $$\sum_{y\le q_i\le x}\frac{1}{q_i} = o(1).$$

From (33) and (35) we finally obtain

(36) $$A(x) = (1+o(1))\frac{x}{\log x}\prod_{q_i\le x}\left(1-\frac{1}{q_i-1}\right).$$

To complete the proof of our Theorem we only have to show that

(37) $$\prod_{q_i\le x}\left(1-\frac{1}{q_i-1}\right) = \frac{1+o(1)}{\log_2 x}.$$

Assume that (37) does not hold. Assume first that

(38) $$\limsup \log_2 x \prod_{q_i\le x}\left(1-\frac{1}{q_i-1}\right) = c > 1.$$

The limit of the expression in (38) cannot exist. For if it would exist it would equal $c > 1$. But then by (36)

$$\lim \frac{A(x)\log x \log_2 x}{x} = c \quad \text{or} \quad \lim \frac{q_n}{n\log n\log_2 n} = \frac{1}{c} < 1$$

which contradicts (38).

Since the limit in (38) does not exist it follows by a simple argument that there exists a constant c', $1 < c' < c$ and two infinite sequences $x_k < z_k$ so that

(39) $$\lim_{k=\infty}\log_2 x_k \prod_{q_i\le x_k}\left(1-\frac{1}{q_i-1}\right) = c'$$

(40) $$\lim_{k=\infty}\log_2 z_k \prod_{q_i\le z_k}\left(1-\frac{1}{q_i-1}\right) = c$$

and for every $x_k < w < z_k$

(41) $$\log_2 x_k \prod_{q_i\le x_k}\left(1-\frac{1}{q_i-1}\right) < \log_2 w \prod_{q_i\le w}\left(1-\frac{1}{q_i-1}\right).$$

From (34) we have for every $a > 1$

(42) $$\prod_{x<q_i<ax}\left(1-\frac{1}{q_i-1}\right) = 1+o(1).$$

Thus from (39), (40) and (42) $z_k/x_k \to \infty$. Choose now $w = (1+\eta)x_k < z_k$ where $\eta > 0$ is a sufficiently small constant. Put

$$U_k = A[(1+\eta)x_k] - A(x_k).$$

From (41) we have

$$(43)\quad \frac{\log_2 x_k}{\log_2 [x_k(1+\eta)]} < \prod_{x_k < q_i < (1+\eta)x_k} \left(1 - \frac{1}{q_i - 1}\right) < \left(1 - \frac{1}{(1+\eta)x_k}\right)^{U_k}.$$

From (36), (39) and (42) we have

$$(44)\quad U_k = (1+o(1)) \frac{c'(1+\eta)x_k}{\log x_k \cdot \log_2 x_k} - (1+o(1)) \frac{c' x_k}{\log x_k \log_2 x_k} = \frac{(1+o(1))c'\eta x_k}{\log x_k \log_2 x_k}.$$

Now by a simple computation

$$(45)\quad \frac{\log_2 x_k}{\log_2 [x_k(1+\eta)]} = 1 - \frac{\log(1+\eta)}{\log x_k \log_2 x_k} + o\left(\frac{1}{\log x_k \log_2 x_k}\right).$$

From (43), (44) and (45) we have

$$(46)\quad \begin{aligned} 1 - \frac{\log(1+\eta)}{\log x_k \log_2 x_k} + o\left(\frac{1}{\log x_k \log_2 x_k}\right) &< \left(1 - \frac{1}{(1+\eta)x_k}\right)^{U_k} \\ &= 1 - \frac{c'\eta}{(1+\eta)\log x_k \log_2 x_k} + o\left(\frac{1}{\log x_k \log_2 x_k}\right). \end{aligned}$$

But (46) is false for sufficiently small η (since $c' > 1$). This contradiction shows that the $\overline{\lim}$ in (38) equals 1. In the same way we can show that the lim of the expression in (38) is 1. Thus (37) is proved, and (36) implies our Theorem.

I do not know whether for infinitely many i's q_{i+1} is the least prime greater than q_i.

By similar arguments we can prove the following more general result: Let $r \geqq 1$, $Q_1 > r+1$, Q_1 prime. Q_{i+1} is the smallest prime greater than Q_i so that $Q_i \not\equiv t \pmod{Q_j}$, $1 \leqq j \leqq i$, $1 \leqq t \leqq r$.

Denote by $B_{Q_1, r}(x)$ the number of Q's not exceeding x, then

$$(47)\quad B_{Q_1,r}(x) = (1+o(1)) \frac{x}{\log x \log_2 x \cdots \log_{r+1} x}.$$

For $Q_1 = 3$, $r = 1$, $A(x) = B_{Q_1,r}(x)$, (47) is thus a generalisation of our Theorem.

Technion,
Haifa.