Canad. Math. Bull. Vol. **33** (2), 1990

ON NUMBER OF INTEGERS REPRESENTABLE AS SUMS OF UNIT FRACTIONS

BY
HISASHI YOKOTA

ABSTRACT. Let $N(n)$ be the set of all integers that can be written in the form $\sum_{i=1}^{n} \epsilon_i/i$, where $\epsilon_i = 0$ or 1. Then $|N(n)| \geqq (1/2 - \epsilon(n)) \log n$, where $\epsilon(n) \longrightarrow 0$ as $n \longrightarrow \infty$, answering a question of P. Erdös and R. L. Graham.

1. **Introduction.** Let $N(n)$ be the set of all integers that can be written in the form $\sum_{i=1}^{n} \epsilon_i/i$, where $\epsilon_i = 0$ or 1. It is easy to see that $|N(n)| \leqq \log n + 1$, where $|A|$ denotes the cardinality of the set A. We are interested in a question of the lower bound of $|N(n)|$, i.e., the maximum number of positive integers in $N(n)$. This question was raised by P. Erdös and R. L. Graham [2]. It is known that $|N(n)| \geqq \log\log n$. But it is not known whether $|N(n)| = 0(\log n)$. In this paper, we show that $|N(n)| = c \log n$, where $c \geqq (1/2) - \epsilon(n)$, $\epsilon(n) = \log_2 n/\log n \longrightarrow 0$ as $n \longrightarrow \infty$. Here and in the sequel, we let $\log_2 n$ denote $\log\log n$.

2. **Main theorems.** To improve the lower bound, we need the following theorems. Let S be the increasing sequence of positive integers of the form p^{2^k}, where $k \geqq 0$ and p a prime. Let s_i be the ith element of S. Let $\{p_{t_i}\}$ be the increasing sequence of primes such that $p_{t_0} \leqq s_t < p_{t_1}$. Then we have

THEOREM 1. *If* $\sum_{d \leqq p_{t_k}} 1/d < a < \sum_{d \leqq p_{t_{k+1}}} 1/d$, *where* d*'s are divisors of* $\prod_1^t s_i \prod_1^k p_{t_i}$, *then* $p_{t_k} \leqq e^{(a-1)(1-1/\log s_t - 3/\log^2 s_t)^{-1}}$.

Now let t and k be chosen so that $s_t/2 < \sqrt{p_{t_k}} < 2s_t$ and denote $\log\log n = \log_2 n$. Then

THEOREM 2. *If* $(1 + 1/(\log_2 p_{t_k}) - 2/\sqrt{s_t}) \prod_1^t s_i \prod_1^k p_{t_i} < r < 2 \prod_1^t s_i \prod_1^k p_{t_i}$, *then* $r = \sum d_i$, *where* d_i*'s are distinct divisors of* $\prod_1^t s_i \prod_1^k p_{t_i}$ *such that* $d_i \geqq \prod_1^t s_i \prod_1^k p_{t_i}/9p_{t_k}(\log p_{t_k})^2(\log_2 p_{t_k})^3$.

Assuming Theorems 1 and 2, we show that every positive integer a is in $N(n)$ if $a \leqq (1/2 - \log_2 n/\log n)\log n$ for n sufficiently large. Let a be a large integer and

Received by the editors November 16, 1988 and, in revised form, February 6, 1989.
AMS (1980) Subject Classification: 11D85, 11D68.

choose t so that $\sqrt{e^a} < s_t < 2\sqrt{e^a}$. Then

$$\sum_{d \leqq s_t} \frac{1}{d} < \log s_t + 1 < a.$$

Thus we can choose k so that

$$\sum_{d \leqq p_{t_k}} \frac{1}{d} < a < \sum_{d \leqq p_{t_{k+1}}} \frac{1}{d},$$

where $d | \prod_1^t s_i \prod_1^k p_{t_i}$. Then by Theorem 1, we have $p_{t_k} \leqq e^{a+3/2}$. Now let d_0 be the largest divisors of $\prod_1^t s_i \prod_1^k p_{t_i}$ so that $\sum_{d \leqq d_0} 1/d \leqq a$. Then we have $p_{t_k} \leqq d_0 < p_{t_{k+1}}$. Let $d^-(d_0)$, $d^+(d_0)$ denote the largest divisor of $\prod_1^t s_i \prod_1^k p_{t_i}$ less than d_0 and the smallest divisor of $\prod_1^t s_i \prod_1^k p_{t_i}$ greater than d_0, respectively. Then

$$\sum_{d \leqq d^-(d_0)} \frac{1}{d} < \sum_{d \leqq d_0} \frac{1}{d} < a < \sum_{d \leqq d^+(d_0)} \frac{1}{d}.$$

Thus we have

$$\frac{1}{p_{t_{k+1}}} < \frac{1}{d_0} < a - \sum_{d \leqq d_0} \frac{1}{d} < \frac{1}{d_0} + \frac{1}{d^+(d_0)} < \frac{2}{p_{t_k}}.$$

Now we write

$$a - \sum_{d \leqq d^-(d_0)} \frac{1}{d} = \frac{r}{\prod_1^t s_i \prod_1^k p_{t_i}}.$$

Then we have

$$\frac{1}{p_{t_{k+1}}} < \frac{r}{\prod_1^t s_i \prod_1^k p_{t_i}} < \frac{2}{p_{t_k}}.$$

Let $q = \left[(2 + 1/\log_2 p_{t_k}) \prod_1^t s_i \prod_1^k p_{t_i}/r\right]^*$, where $[x]^*$ denotes the greatest integer less than x. Then

$$a - \sum_{d \leqq d^-(d_0)} \frac{1}{d} - \frac{1}{q} = \frac{r}{\prod_1^t s_i \prod_1^k p_{t_i}} - \frac{1}{q}$$

$$= \frac{r^*}{q \prod_1^t s_i \prod_1^k p_{t_i}},$$

where

$$\left(1 + \frac{1}{(\log_2 p_{t_k})} - \frac{r}{\prod_1^t s_i \prod_1^k p_{t_i}}\right) \prod_1^t s_i \prod_1^k p_{t_i} < r^* < 2 \prod_1^t s_i \prod_1^k p_{t_i}.$$

https://doi.org/10.4153/CMB-1990-037-0 Published online by Cambridge University Press

Thus

$$\left(1+\frac{1}{(\log_2 p_{t_k})}-\frac{2}{\sqrt{s_t}}\right)\prod_1^t s_i \prod_1^k p_{t_i} < r^* < 2\prod_1^t s_i \prod_1^k p_{t_i}.$$

Therefore by Theorem 2, $r^* = \sum d_i$, where $d_i \geqq \prod_1^t s_i \prod_1^k p_{t_i}/9p_{t_k}(\log p_{t_k})^2(\log_2 p_{t_k})^3$. Hence

$$a = \sum_{d \leqq d^-(d_0)} \frac{1}{d}+\frac{1}{q}+\frac{1}{q}\left(\sum \frac{1}{d_i^*}\right),$$

where $d_i^* \leqq 9p_{t_k}(\log p_{t_k})^2(\log_2 p_{t_k})^3$. Thus the largest denominator in this expansion of a is less than $9qp_{t_k}(\log p_{t_k})^2(\log_2 p_{t_k})^3$. Since $q \leqq (2+1/\log_2 p_{t_k})\prod_1^t s_i \prod_1^k p_{t_i}/r < 5p_{t_k}$ and $p_{t_k} \leqq e^{a+3/2}$, we have

$$\begin{aligned} qd_i^* &\leqq 45(e^{a+3/2})^2 \log(e^{a+3/2})^3 \\ &\leqq ae^{2a+8}. \end{aligned}$$

Hence $a \in N(n)$ provided $ae^{2a+8} \leqq n$. But this implies that $a \leqq (1/2 - \log_2 n/\log n)$. Thus $|N(n)| \geqq c \log n$ with $c \geqq 1/2-\epsilon(n)$, where $\epsilon(n) = \log_2 n/\log n \longrightarrow 0$ as $n \longrightarrow \infty$.

3. **Lemmata.** To prove Theorems 1 and 2, we need a few lemmas.

LEMMA 1. (i) $\prod_1^t s_i^{\epsilon_i}$, $\epsilon_i = 0$ *or* 1, *are all distinct;*
(ii) *if* $1 \leqq a < s_t$, *then* $a = \prod_1^t s_i^{\epsilon_i}$, $\epsilon_i = 0$ *or* 1.

LEMMA 2. *If* $\prod_1^{k-1} p_i < N < \prod_1^k p_i$, *then* $p_k \leqq \log N(1+2/\log_2 N)$ *for* N *large and* $p_k \leqq 2\log N/\log 2$ *for* $N \geqq 2$.

LEMMA 3. *If* $\prod_1^{k-1} s_i < N < \prod_1^k s_i$, *then* $s_k \geqq \log N(1-2/\log_2 N)$ *for* N *large.*

LEMMA 4. *Let* s_t *be a prime such that* $s_t \geqq 5$. *Then* $D = \{d : \sqrt{s_t} < d < 2s_t \log s_t/\log 2, d | \prod_1^{t-1} s_i\} U \{0\}$ *contains all residues modulo* s_t.

LEMMA 5. *If* $(1-2/\sqrt{s_t})\prod_1^t s_i \leqq r \leqq 2\prod_1^t s_i$, $t \geqq 3$, *then there are distinct divisors* d_i *of* $\prod_1^t s_i$ *such that* $r = \sum d_i$, *with* $d_i > \prod_1^t s_i/3s_t^2 \log s_t$.

Proofs of Lemmas 1 and 3 can be found in [4]. A proof of Lemma 2 is in [1]. Proofs of Lemmas 4 and 5 are in [5].

4. **Proof of Theorems.** We start with the proof of Theorem 1. Let d's be divisors of $\prod_1^t s_i \prod_1^k p_{t_i}$. Then we have

$$\begin{aligned} \sum_{d \leqq p_{t_k}} 1/d &= \sum_{i=1}^{p_{t_k}} \frac{1}{i} - \left\{\frac{1}{2^{\alpha_1+1}}\left(1+\frac{1}{2}+\ldots+\frac{1}{m_1}\right)+\ldots+\frac{1}{p_i^{\alpha_i+1}}\left(1+\frac{1}{2}+\ldots+\frac{1}{m_i}\right)\right\} \\ &= \sum_{i=1}^{p_{t_k}} \frac{1}{i} - \sum_{1<p<s_t}\left(\frac{1}{p^{\alpha_p+1}}\right)\left(1+\frac{1}{2}+\ldots+\frac{1}{m_{\alpha_p}}\right), \end{aligned}$$

https://doi.org/10.4153/CMB-1990-037-0 Published online by Cambridge University Press

where m_{α_p} is the largest integer such that $m_{\alpha_p} p^{\alpha_p} \leqq p_{t_k}$ and $p^{\alpha_p} \parallel \prod_1^t s_i$, yielding $p^{\alpha_p+1} \geqq s_t$ by Lemma 1. Thus

$$\sum_{d \leqq p_{t_k}} 1/d > \sum_{i=1}^{p_{t_k}} \frac{1}{i} - \pi(s_t)\left(\frac{1}{s_t}\right)(\log p_{t_k} - \log s_t + 1).$$

Since $\pi(x) \leqq x(1 + 3/2 \log x)/\log x$ for $x > 1$ by [3], we have

$$\sum_{d \leqq p_{t_k}} 1/d > \log p_{t_k}\left(1 - \frac{1}{\log s_t} - \frac{3}{2(\log s_t)^2}\right) + 1.$$

Thus if $p_{t_k} > e^{(a-1)(1-1/\log s_t - 3/2\log^2 s_t)^{-1}}$, Then $\sum_{d \leqq p_{t_k}} 1/d > a$. Hence $p_{t_k} \leqq e^{(a-1)(1-1/\log s_t - 3/2\log^2 s_t)^{-1}}$. □

PROOF OF THEOREM 2. Let $D_j = \{d : 1 \leqq d \leqq 3(\log p_{t_j})^2 \log_2 p_{t_j},\ d \mid \prod_1^t s_i\}$. Let s_q be an element in S such that $s_q > 3(\log_2 p_{t_k})^2$. Then $s_q \leqq 6(\log_2 p_{t_k})^2 < s_t$. Define for $j = 1, 2, \ldots, k$,

$$D_j^* = \left\{ \frac{\prod_1^t s_i \prod_1^{j-1} p_{t_i}}{s_q \cdot d} : d \in D_j \right\}.$$

Note that if $d_j^* \in D_j^*$, then

$$\frac{\prod_1^t s_i \prod_1^{j-1} p_{t_i}}{3 s_q (\log p_{t_j})^2 \log_2 p_{t_j}} \leqq d_j^* \leqq \frac{\prod_1^t s_i \prod_1^{j-1} p_{t_i}}{s_q}$$

We claim that

$$\left\{ \sum d_j^* \epsilon_j : \epsilon_j = 0 \quad \text{or} \quad 1, d_j^* \in D_j^* \right\} \equiv \{0, 1, 2, 3, \ldots, p_{t_j} - 1\} (\operatorname{mod} p_{t_j}).$$

Let a be a residue modulo p_{t_j}. Let k be such that $\prod_1^{k-1} s_i < p_{t_j} < \prod_1^k s_i$. Then by Lemma 2, $s_k < \log p_{t_j}(1 + 2/\log_2 p_{t_j})$. Now consider

$$\frac{a}{p_{t_k}} = \frac{a \cdot \prod_1^k s_i}{p_{t_j} \cdot \prod_1^k s_i} = \frac{p_{t_j} s + r^*}{p_{t_j} \cdot \prod_1^k s_i},$$

https://doi.org/10.4153/CMB-1990-037-0 Published online by Cambridge University Press

where r^* is chosen so that $(1 - 2/\sqrt{s_k})\prod_1^k s_i \leqq r^* < 2\prod_1^k s_i$. Then by Lemma 5, $r^* = \sum d_i$, where d_i are distinct divisors of $\prod_1^k s_i$ and $d_i \geqq \prod_1^k s_i/3s_k^2 \log s_k$. Thus

$$\begin{aligned} a\prod_1^k s_i &\equiv r^* \\ &\equiv \prod_1^k s_i \left(\frac{r^*}{\prod_1^k s_i}\right) \\ &\equiv \prod_1^k s_i \left(\frac{\sum d_i}{\prod_1^k s_i}\right) \\ &\equiv \prod_1^k s_i \left(\sum_{i=1}^{3s_k^2 \log s_k} \frac{\epsilon_i}{i}\right) \pmod{p_{t_j}}. \end{aligned}$$

Since $\left(\prod_1^k s_i, p_{t_k}\right) = 1$, $a\prod_1^k s_i$ runs through all residues modulo p_{t_j} except 0 as a runs through all residues except 0. Thus

$$\left\{\prod_1^k s_i \left(\sum_{i=1}^{3s_k^2 \log s_k} \frac{\epsilon_i}{i}\right) : \epsilon_i = 0 \text{ or } 1, \quad i \mid \prod_1^k s_i\right\}$$

contains all residues modulo p_{t_j}. Since $\left(\prod_{k+1}^t s_i \prod_1^{j-1} p_{t_i}/s_q, p_{t_j}\right) = 1$, we have $\left\{\sum d_j^* \epsilon_j : \epsilon_j = 0 \text{ or } 1, d_j^* \in D_j^*\right\} \equiv \{0, 1, 2, 3, \ldots, p_{t_j} - 1\} \pmod{p_{t_j}}$. Thus $r \equiv \sum d_k^* \epsilon_k \pmod{p_{t_k}}$ and

$$\begin{aligned} \sum d_k^* \epsilon_k &= \frac{\prod_1^t s_i \prod_1^{k-1} p_{t_i}}{s_q} \left(\sum \frac{\epsilon_i}{i}\right) \\ &\leqq \frac{\prod_1^t s_i \prod_1^{k-1} p_{t_i}}{s_q} [\log(3(\log p_{t_k})^2 \log_2 p_{t_k}) + 1] \\ &\leqq \frac{\prod_1^t s_i \prod_1^{k-1} p_{t_i}}{s_q} [3\log_2 p_{t_k}]. \end{aligned}$$

Let $r_1 = (r - \sum d_k^* \epsilon_k)/p_{t_k}$, an integer. Then

$$r_1 \geqq \left(1 + \frac{1}{(\log_2 p_{t_k})} - \frac{2}{\sqrt{s_t}} - \frac{3\log_2 p_{t_k}}{p_{t_k} \cdot s_q}\right) \prod_1^t s_i \prod_1^{k-1} p_{t_i}$$

https://doi.org/10.4153/CMB-1990-037-0 Published online by Cambridge University Press

and

$$r_1 < r/p_{t_k} < 2\prod_1^t s_i \prod_1^{k-1} p_{t_i}.$$

Repeat the same argument $k-1$ times and note that

$$\sum_{p_{t_1}}^{p_{t_k}} \frac{1}{p} \leqq \log_2 p_{t_k} + B_1 + 1/(\log^2 p_{t_k}) - (\log_2 p_{t_1} + B_2 - 1/(2\log^2 p_{t_1}))$$
$$\leqq \log 2 + 3/\log^2 p_{t_k} \quad \text{by [3]}$$

and $s_q \geqq 3(\log_2 p_{t_k})^2$. Then we have

$$r_k \geqq \left(1 + \frac{1}{(\log_2 p_{t_k})} - \frac{2}{\sqrt{s_t}} - \frac{3\log_2 p_{t_k}}{s_q} \sum_{p_{t_1}}^{p_{t_k}} \frac{1}{p}\right) \prod_1^t s_i$$
$$\geqq \left(1 + \frac{1}{(\log_2 p_{t_k})} - \frac{2}{\sqrt{s_t}} - \frac{1}{(\log_2 p_{t_k})}\right) \prod_1^t s_i$$
$$\geqq \left(1 - \frac{2}{\sqrt{s_t}}\right) \prod_1^t s_i .$$

Also $r_k < 2\prod_1^t s_i$. Thus

$$\left(1 - \frac{2}{\sqrt{s_t}}\right) \prod_1^t s_i \leqq r_k < 2\prod_1^t s_i .$$

Note that

$$r = p_{t_k} r_1 + \sum d_k^* \epsilon_k$$
$$= p_{t_k}\left(p_{t_{k-1}} r_2 + \sum d_{k-1}^* \epsilon_{k-1}\right) + \sum d_k^* \epsilon_k$$
$$= \prod_1^k p_{t_i} r_k + \prod_2^k p_{t_i}\left(\sum d_1^* \epsilon_1\right) + \ldots + p_{t_k}\left(\sum d_{k-1}^* \epsilon_{k-1}\right) + \sum d_k^* \epsilon_k,$$

and $\prod_j^k p_{t_i} d_{j-1}^*$ are all distinct for $j = 2, 3, \ldots, k$. Note also that

$$\prod_{j+1}^k p_{t_i} d_j^* > \frac{\prod_1^t s_i \prod_1^k p_{t_i}}{3p_{t_j} s_q (\log p_{t_j})^2 \log_2 p_{t_j}}$$
$$> \frac{\prod_1^t s_i \prod_1^k p_{t_i}}{9p_{t_k} (\log p_{t_k})^2 (\log_2 p_{t_k})^3}$$

https://doi.org/10.4153/CMB-1990-037-0 Published online by Cambridge University Press

for $j = 1, 2, \ldots, k-1$. Thus to show $r = \sum d_i$, where d_i are distinct divisors of $\prod_1^t s_i \prod_1^{j-1} p_{t_i}$ and $d_i \geqq \prod_1^t s_i \prod_1^{j-1} p_{t_i}/9p_{t_k}(\log p_{t_k})^2(\log_2 p_{t_k})^3$, it suffices to show $r_k = \sum d_i'$, where d_i' are distinct divisors of $\prod_1^t s_i$ and $d_i' \geqq \prod_1^t s_i/9p_{t_k}(\log p_{t_k})^2(\log_2 p_{t_k})^3$. Since

$$\left(1 - \frac{2}{\sqrt{s_t}}\right) \prod_1^t s_i \leqq r_k < 2 \prod_1^t s_i,$$

by Lemma 5, we have $r_k = \sum d_i'$, where d_i' are distinct divisors of $\prod_1^t s_i$ and $d_i' \geqq \prod_1^t s_i/3s_t^2 \log s_t$. Also $s_t/2 < \sqrt{p_{t_k}}$. Thus we have $d_i' > \prod_1^t s_i/3s_t^2 \log s_t > \prod_1^t s_i/9p_{t_k}(\log p_{t_k})^2(\log_2 p_{t_k})^3$ for P large.

REMARK. Erdös and Graham [2] also asked size of the smallest integer not in $N(n)$. From the above result, it is at least $\log n(1/2 - \epsilon(n))$, where $\epsilon(n) \longrightarrow 0$ as $n \longrightarrow \infty$.

REFERENCES

1. M. N. Bleicher and P. Erdös, *Denominators of Egyptian Fractions II*, Illinois J. Math. **20** (1976), 598–613.

2. P. Erdös and R. L. Graham, *Old and new problems in combinatorial number theory*, L'Enseignement Mathématique, Université de Geneve, Imprimerie Kundig Geneve, 1980.

3. J. Rosser and L. Schoenfeld, *Approximate formula for some functions of prime numbers*, Illinois J. Math. **6** (1962), 64–94.

4. H. Yokota, *Denominators of Egyptian Fractions*, J. Number Theory **28** (1988), 258–271.

5. ———, *On a problem of Bleicher and Erdös*, J. Number Theory **30** (1988), 198–207.

Hiroshima Institute of Technology
Itsukaichi, Hiroshima, Japan