DENOMINATORS OF EGYPTIAN FRACTIONS II

BY

MICHAEL N. BLEICHER AND PAUL ERDŐS

I. Introduction

A positive fraction a/N is said to be written in Egyptian form if we write

$$a/N = 1/n_1 + 1/n_2 + \cdots + 1/n_k, \quad 0 < n_1 < n_2 < \cdots < n_k,$$

where the n_i are integers. Among the many expansions for each fraction a/N there is some expansion for which n_k is minimal. Let $D(a, N)$ denote the minimal value of n_k.

Define $D(N)$ by $D(N) = \max \{D(a, N): 0 < a < N\}$. We are interested in the behavior of $D(N)$. In our paper [1] we showed that for $N = P$, a prime, $D(P) \geq P \log P$ and that for some constant K and any $N > 1$, $D(N) \leq KN (\log N)^4$. It was surprising that such close upper and lower bounds could be achieved by the simple techniques of [1]. In this paper we refine the techniques of [1] and show that on the one hand for P large enough that $\log_{2r} P \geq 1$,

$$D(P) \geq \frac{P \log P \log_2 P}{\log_{r+1} P \prod_{j=4}^{r+1} \log_j P}$$

and on the other hand that for $\varepsilon > 0$ and N sufficiently large (Theorem 1 and its corollary yield more precise statements), $D(N) \leq (1 + \varepsilon)N (\log N)^2$. We conjecture that the exponent 2 can be replaced by $(1 + \delta)$ for $\delta > 0$.

As part of the proof of the above results we need to analyze the number of distinct subsums of the series $\sum_{i=1}^{N} 1/i$, say $S(N)$. We show that whenever $\log_{2r} N \geq 1$,

$$\frac{\alpha N}{\log N} \prod_{j=3}^{r} \log_j N \leq \log S(N) \leq \frac{N \log_r N}{\log N} \prod_{j=3}^{r} \log_j N$$

for some $\alpha \geq 1/e$.

II. The upper bound for $D(N)$

Let p_k denote the kth prime, and let $\Pi_k = \prod_{i=1}^{k} p_i$. We recall from [1]:

LEMMA 1. *If $0 < r < \sigma(\Pi_k)$ then there are divisors d_i of Π_k such that $r = \sum d_i$.*

Received July 5, 1974; received in revised form January 27, 1976.

LEMMA 2. *For N sufficiently large, if k is chosen so that $\Pi_{k-1} \le N \le \Pi_k$, then*

$$p_k \le \log N\left(1 + \frac{2}{\log\log N}\right).$$

Proof. If $\vartheta(x) = \sum_{p \le x} \log p$ then $\log \Pi_k = \vartheta(p_k)$. We note that p_k is the least prime such that $\vartheta(p_k) \ge \log N$. By [4, Theorem 4], $\vartheta(x) \ge x(1 - (1/2 \log x))$ for large enough x. Thus if

$$x_0 = \log N\left(1 + \frac{1}{\log\log N}\right)$$

then $\vartheta(x_0) \ge \log N$. Let p_0 be the least prime greater than x_0. For x_0 sufficiently large we have [3, p. 323] $p_0 \le x_0 + x_0^{2/3}$. Since $p_k \le p_0$,

$$p_k \le \log N\left(1 + \frac{2}{\log\log N}\right)$$

for N sufficiently large.

LEMMA 2*. *If $N \ge 2$ and $\Pi_{k-1} < N \le \Pi_k$ then $p_k \le 2 \log N/\log 2$.*

Proof. For $N = 2$, $p_k = 2$ and the lemma holds. For $3 \le N \le 6$, $p_k = 3$ and the lemma holds. For $\Pi_2 < N \le \Pi_{16}$ the theorem follows since for $k \le 16$, computation shows that $p_k \le 2 \log \Pi_{k-1}/\log 2$. For $N \ge \Pi_{16}$ we have $\log N \ge 41$. By definition of $\vartheta(x)$, $\log \Pi_k = \vartheta(p_k)$ where p_k is the least prime such that $\vartheta(p_k) \ge \log N$. Since for $x \ge 41$ we have [4, Theorem 4, Corollary] $\vartheta(x) \ge x(1 - (1/\log x))$, we see that

$$\vartheta(x_0) \ge \log N \quad \text{for } x_0 = \log N\left(1 + \frac{3}{2\log\log N}\right) \ge 41.$$

By Betrand's postulate we see that $p_k \le 2x_0$. Since

$$2\left(1 + \frac{3}{2\log\log N}\right) \le 2/\log 2 \quad \text{when } \log N \ge 41,$$

the lemma follows.

LEMMA 3. *If $N \ge 12$, then in the closed interval $[\sqrt{N}, N + \sqrt{N}]$ there are at least $[N/2] + 1$ square-free integers with all prime factors less than N.*

Proof. Let $\Pi^* = \prod_{p<N} p$. Let $D = \{m: \sqrt{N} \le m \le N + \sqrt{N}, m \mid \Pi^*\}$. Let $Q(x)$ be the number of square free integers not exceeding x. Thus

$$|D| \ge Q(N + \sqrt{N}) - Q(\sqrt{N}) - L$$

where L is the number of primes between N and $N + \sqrt{N}$ inclusive. Suppose $N \ge 24^2$, so that $\sqrt{N} \ge 24$. In the interval $[N, N + \sqrt{N}]$ only odd numbers can be prime; there are at most $1 + \frac{1}{2}\sqrt{N}$ odd numbers, and at least four of them are divisible by 3. We deduce that $L \leq (\frac{1}{3}\sqrt{N}) - 3$. From the proof of Theorem 333 in [2] we see that

$$Q(x) = \sum_{d^2 \leq x} \mu(d) \left[\frac{x}{d^2}\right].$$

Thus

$$\begin{aligned} Q(N + \sqrt{N}) - Q(\sqrt{N}) &= \sum_{d \leq \sqrt{N+\sqrt{N}}} \mu(d) \left[\frac{N + \sqrt{N}}{d^2}\right] \\ &\quad - \sum_{d \leq N^{1/4}} \mu(d) \left[\frac{\sqrt{N}}{d^2}\right] \\ &\geq (N + \sqrt{N}) \sum_{d \leq \sqrt{N+\sqrt{N}}} \frac{\mu(d)}{d^2} \\ &\quad - \sqrt{N} \sum_{d \leq N^{1/4}}^{d \leq N^{1/4}} \frac{\mu(d)}{d^2} - [\sqrt{N + \sqrt{N}}]. \end{aligned}$$

Since $\sum_{d=1}^{\infty} \mu(d)/d^2 = 1/\zeta(2) = 6/\pi^2$ and $|\mu(d)| \leq 1$ we get

$$\begin{aligned} Q(N + \sqrt{N}) - Q(\sqrt{N}) &\geq \frac{6N}{\pi^2} - [\sqrt{N + \sqrt{N}}] - N \sum_{d > \sqrt{N+\sqrt{N}}} \frac{1}{d^2} \\ &\quad - \sqrt{N} \sum_{N^{1/4} < d \leq \sqrt{N+\sqrt{N}}} \frac{1}{d^2} \\ &> \frac{6N}{\pi^2} - M - \frac{N}{M} - \sqrt{N}\left(\frac{1}{[N^{1/4}]} - \frac{1}{M}\right) \end{aligned}$$

where $M = [\sqrt{N + \sqrt{N}}]$. Since $\sqrt{N + \sqrt{N}} - \sqrt{N - \sqrt{N}} \geq 1$, we see that $M \geq \sqrt{N - \sqrt{N}}$ and hence that the above expression is decreasing in M. Thus we obtain

$$\begin{aligned} Q(N + \sqrt{N}) - Q(\sqrt{N}) &\geq \frac{6N}{\pi^2} - \sqrt{N + \sqrt{N}} - \frac{N}{\sqrt{N + \sqrt{N}}} \\ &\quad - \sqrt{N}\left(\frac{1}{[N^{1/4}]} - \frac{1}{\sqrt{N + \sqrt{N}}}\right) \\ &= \frac{6N}{\pi^2} - \frac{2N}{\sqrt{N + \sqrt{N}}} - \frac{\sqrt{N}}{[N^{1/4}]}. \end{aligned}$$

Thus

$$|D| \geq \frac{6N}{\pi^2} - \frac{2N}{\sqrt{N + \sqrt{N}}} - \frac{\sqrt{N}}{[N^{1/4}]} - \frac{\sqrt{N}}{2} + 3.$$

To show that $|D| \geq N/2$ it suffices to show that

$$0.1079\cdots = \frac{6}{\pi^2} - \frac{1}{2} \geq \frac{2}{\sqrt{N + \sqrt{N}}} + \frac{1}{2\sqrt{N}} + \frac{1}{\sqrt{N}[N^{1/4}]} - \frac{3}{N}$$

which is true for $N = 24^2$, whence for $N \geq 24^2$. On the other hand one can verify directly and/or by special arguments that the lemma is true for $576 \geq N \geq 12$.

LEMMA 4. *If* $\Pi_k(1 - (2/\sqrt{p_k})) \leq r < 2\Pi_k$ *then there are distinct* d_i *such that*

$$d_i \mid \Pi_k,\ d_i > \Pi_{k-1}(p_k + \sqrt{p_k})^{-1} \quad \text{and} \quad r = \sum d_i.$$

Proof. We note, in order to begin a proof by induction, that the lemma is true for $k = 1, 2, 3$, since for these cases $\Pi_{k-1}(p_k + \sqrt{p_k})^{-1} < 1$. We suppose $k \geq 4$ and that the lemma is true for all $k' < k$. Consider the set

$$D = \{d\colon \sqrt{p_k} \leq d < p_k + \sqrt{p_k},\ d \mid \Pi_{k-1}\}.$$

Case 1. $k \geq 6$, i.e., $p_k \geq 13$. Let r be given in the desired range. According to Lemma 3, $|D| \geq (p_k + 1)/2$. Also note that no two elements of D are congruent mod p_k and that none is congruent to zero mod p_k. Let

$$D^* = \{0\} \cup \{\Pi_{k-1}/d;\ d \in D\}.$$

If $d \in D^*$, $d \neq 0$ then $\Pi_{k-1}(\sqrt{p_k})^{-1} \geq d \geq \Pi_{k-1}(p_k + \sqrt{p_k})^{-1}$. We note that $|D^*| \geq (p_k + 3)/2$ and no two elements of D^* are congruent mod p_k. If $r \equiv 2d \bmod p_k$ for some $d \in D^*$, let $D^{**} = D^*\backslash\{d\}$, otherwise let $D^{**} = D^*$. Hence $|D^{**}| \geq (p_k + 1)/2$ and we may apply the Cauchy-Davenport Theorem to find d' and d'', distinct elements of D^{**} such that $r - d' - d'' \equiv 0 \bmod p_k$. Let $r^* = r - d' - d''$. Then

$$r^* \geq r - \frac{2\Pi_{k-1}}{\sqrt{p_k}} \geq \Pi_k\left(1 - \frac{2}{\sqrt{p_k}} - \frac{2}{p_k\sqrt{p_k}}\right).$$

Since $1/\sqrt{p_{k-1}} - 1/\sqrt{p_k} \geq 1/p_k\sqrt{p_k}$, as is seen by using the mean value theorem on $1/\sqrt{x}$, we deduce that $r^* \geq \Pi_k(1 - (2/\sqrt{p_{k-1}}))$. Let $r' = r^*/p_k$, an integer. Then

$$\Pi_{k-1}\left(1 - \frac{2}{\sqrt{p_{k-1}}}\right) \leq r' < 2\Pi_{k-1},$$

so by induction $r' = \sum d_i$ where $d_i \mid \Pi_{k-1}$, $d_i \geq (p_{k-1} + \sqrt{p_{k-1}})^{-1}\Pi_{k-2}$. It follows that $r = \sum p_k d_i + d' + d''$, and since the d_i were distinct by induction, so are the $p_k d_i$; also, unless either d' or d'' is zero, in which case we discard it from the sum, $d', d'' \not\equiv 0 \bmod p_k$ so that all the terms in the sum are distinct. Clearly

$$d', d'' \geq \frac{\Pi_{k-1}}{p_k + \sqrt{p_k}}.$$

On the other hand, by induction

$$d_i \geq \frac{\Pi_{k-2}}{p_{k-1} + \sqrt{p_{k-1}}},$$

thus

$$d_i p_k \geq \frac{\Pi_{k-2} p_k}{p_{k-1} + \sqrt{p_{k-1}}} \geq \frac{\Pi_{k-1}}{p_k + \sqrt{p_k}}.$$

Case 2. $k = 4, 5$. $p_k = 7, 11$. An easy computation shows that for $p_k = 7$, $D^* = \{0, 5, 6, 10\}$. Every nonzero congruence class mod 7 can be obtained as a sum of two or fewer elements of D^* as follows: $1 \equiv 5 + 10$, $2 \equiv 6 + 10$, $3 \equiv 10 + 0$, $4 \equiv 5 + 6$, $5 \equiv 5 + 0$, and $6 \equiv 6 \bmod 7$. Thus for $r \not\equiv 0 \bmod 7$ we may proceed to define r' as in Case 1. If $r \equiv 0 \bmod 7$, let $r^* = r$ and proceed as in Case 1.

For $p_k = 11$, $D^* = \{0, 2\cdot 3\cdot 7, 5\cdot 7, 2\cdot 3\cdot 5, 3\cdot 7, 3\cdot 5\} \equiv \{0, 9, 2, 8, 10, 4\}$ mod 11. Every congruence class mod 11 can be obtained as a sum of at most three distinct elements of D^* as follows: $0 \equiv 0$, $1 \equiv 10 + 2$, $2 \equiv 2$, $3 \equiv 10 + 4$, $4 \equiv 4$, $5 \equiv 10 + 4 + 2$, $6 \equiv 4 + 2$, $7 \equiv 10 + 8$, $8 \equiv 10 + 9$, $9 \equiv 9$, $10 \equiv 10$. Thus we may define r' and proceed as in Case 1. The proof is completed.

We are now ready to prove:

THEOREM 1. *For every* N, $D(N) \leq \lambda^3(N)N(\ln N)^2$ *where* $2/\log 2 \geq \lambda(N) \geq 1$ *and* $\lim_{N\to\infty} \lambda(N) = 1$.

Proof. Given a/N choose Π_k such that $\Pi_{k-1} < N \leq \Pi_k$. If $N \mid \Pi_k$, then $a/N = b/\Pi_k$. By Lemma 1, $b = \sum d_i$, $d_i \mid \Pi_k$. By reducing the fractions in $\sum d_i/\Pi_k$ we obtain a representation of a/N in which no denominator exceeds $\Pi_k < 2N \log N/\log 2$.

If $N \nmid \Pi_k$ write $a/N = (qN + r)/N\Pi_k$ where r is chosen so that

$$\Pi_k\left(1 - \frac{2}{\sqrt{p_k}}\right) \leq r \leq 2\Pi_k.$$

This can be done since we may assume $a \geq 2$ and since $N \leq \Pi_k$. The fraction q/Π_k can be handled by Lemma 1, as in the paragraph above. We now use Lemma 4 to write r/Π_k in Egyptian form using very small denominators. By Lemma 4, $r = \sum d_i$ where $d_i \mid \Pi_k$, the d_i are distinct and $d_i \geq \Pi_{k-1}(p_k + \sqrt{p_k})^{-1}$. Thus $r/\Pi_k = (\sum d_i)/\Pi_k = \sum 1/n_i'$ where $n_i' = \Pi_k/d_i$. Thus the n_i' are distinct and $n_i' \leq p_k(p_k + \sqrt{p_k})$. It follows that $r/N\Pi_k = \sum 1/n_i$ where $n_i = n_i'N$ and the n_i are distinct from each other as well as from the denominators in the expansion of q/Π_k since these denominators all divide Π_k while $N \mid n_i$ and $N \nmid \Pi_k$. Furthermore

$$n_i \leq Np_k(p_k + \sqrt{p_k}) \leq \lambda^3(N)N(\ln N)^2$$

where $\lambda(N)$ can be chosen to satisfy $2/\log 2 \geq \lambda(N)$ by Lemma 2*, $\lim_{N\to\infty} \lambda(N) = 1$ by Lemma 2, and $\lambda(N) \geq (1 + (1/\sqrt{\log N}))$.

III. The number of distinct subsums of $\sum_{i=1}^{N} 1/i$.

DEFINITION. Let $S(N)$ denote the number of distinct values of $\sum_{k=1}^{N} \varepsilon_k/k$ where the ε_k's take on all possible combinations of values with $\varepsilon_k = 0$ or 1.

To obtain a lower bound for $S(N)$ we begin with the following lemma.

LEMMA 5. *For all $N \geq 3$, $S(N) \geq 2^{N/\log N}$.*

Proof. It is clear that each distinct choice of the ε_p's for p prime yields a different value of $\sum_{p \leq N} \varepsilon_p/p$. Thus $S(N) \geq 2^{\pi(N)}$. Since for $N \geq 17$, $\pi(N) \geq N/\log N$ by Corollary 1 of Theorem 2 of [4], the lemma is true for $N \geq 17$. To verify that the result holds for $3 \leq N \leq 16$, note that both $S(N)$ and $2^{N/\log N}$ are monotone and $2^{4/\log 4} \leq 8 \leq S(3)$, $2^{12/\log 12} < 2^5 \leq S(5)$ and $2^{16/\log 16} < 2^6 = 2^{\pi(13)} \leq S(13)$, where $S(3) = 8$ and $S(5) = 2^5$ are a result of direct verification. Thus the lemma is proved.

THEOREM 2. *If $r \geq 1$ and N is large enough that $\log_{2r} N \geq 1$, then*

$$S(N) \geq \exp\left(\alpha \cdot \frac{N}{\log N} \cdot \prod_{j=3}^{r} \log_j N\right)$$

where $\alpha = 1/e$ is a permissible value for α and $\log_1 x = \log x$, $\log_j x = \log(\log_{j-1} x)$.

Proof. The proof is by induction on r.

In order to prove the theorem with the proper constant we make the slightly stronger (as will be shown at the end of the proof) inductive hypothesis

$$(*) \qquad S(N) \geq \exp\left(\prod_{j=3}^{k}\left(1 - \frac{3}{\log_{2j-2} N}\right) \cdot \frac{N}{\log N} \prod_{3}^{k} \log_j N\right)$$

for $\log_{2k} N \geq 1$. The hypothesis $(*)$ is clearly true for $k = 1, 2$ by Lemma 5. We assume the induction hypothesis holds for $k = 1, 2, \ldots, r-1$ and show that it also holds for $k = r \geq 3$.

Let $Q = 2N/\log N$ and $Q' = N/\log_2 N$. Note that $Q' > Q$. We define $\mathscr{P}$ by

$$\mathscr{P} = \{N \geq p \geq Q : p \text{ a prime}\}.$$

Let $T = \{k \leq N: \text{there exists } p \in \mathscr{P}, p \mid k\}$.

$S(N)$ is greater than the number of distinct values of the sume $\sum_{k \in T} \varepsilon_k/k$, which we denote by $T(N)$. We rewrite the sum as

$$\sum_{k \in T} \frac{\varepsilon_k}{k} = \sum_{p \in \mathscr{P}} \frac{1}{p}\left(\sum_{k=1}^{N/p} \frac{\varepsilon_k}{k}\right).$$

Set $\sum_{k=1}^{N/p} \varepsilon_k/k = a_p/b_p$ where $\log b_p = \psi(N/p)$, $\psi(x) = \sum_{p^\alpha \leq x} \log p$. Also

$$a_p \leq 2b_p \log N/p \quad \text{for } p \leq N/3.$$

Thus, if

$$\frac{1}{p}\left(\frac{a_p}{b_p} - \frac{a_p'}{b_p}\right) = \frac{c}{d}, \quad (c, d) = 1,$$

then $p \mid d$ if $p \not\mid (a_p - a_p')$. But for $p \leq N/3$,

$$a_p - a_p' \leq 2b_p \log N/p \leq 2 \log (N/p)e^{\psi(N/p)}.$$

Since $\psi(x) < (1.04)x$ [4, Theorem 12] we see that

$$a_p - a_p' \leq 2 \log (N/Q)e^{(1.04)N/Q} < Q \leq p,$$

since $N \geq e^e$. For $p > N/3$ it is clear that $p \not\mid (a_p - a_p')$.

Thus $p \not\mid (a_p - a_p')$ and $p \mid d$. It follows that distinct choices of a_p/b_p yield distinct sums. Thus $T(N) \geq \prod_{p \in \mathscr{P}} S(N/p)$, so that $S(N) \geq \prod_{p \in \mathscr{P}} S(N/p)$.

We will now evaluate the above product using our inductive hypothesis. First note that

$$\log S(N) \geq \sum_{p \in \mathscr{P}} \log S\left(\frac{N}{p}\right).$$

For simplicity let $S^*(x) = \log S(x)$.

We recall the well-known method using Stieltjes integration with respect to $\vartheta(x)$ and integration by parts by which one evaluates sums where the variable runs over primes [4, p. 74].

LEMMA 6. *If $f'(p)$ exists and is continuous then*

$$\sum_{Q < p \leq Q'} f(p) = \int_Q^{Q'} \frac{f(x)}{\log x}\, dx + \left(\frac{\vartheta(x) - x}{\log x} f(x)\right)\Big|_Q^{Q'} - \int_Q^{Q'} (\vartheta(x) - x) \frac{d}{dx}\left(\frac{f(x)}{\log x}\right) dx.$$

Let $L^*(x) = x/\log x \prod_3^{r-1} \log_j x$, and note that for $Q < p \leq Q'$, $N/p \geq \log_2 N$; hence $\log_{2(r-1)} N/p \geq \log_{2r} N \geq 1$, and the induction assumption tells us that

$$S^*(N/p) \geq \prod_4^r \left(1 - \frac{3}{\log_{2j-2} N}\right) L^*(N/p).$$

We thus obtain

$$\begin{aligned}\left(\prod_{j=4}^r \left(1 - \frac{3}{\log_{2j-2} N}\right)\right)^{-1} S^*(N) &\geq \sum_{Q < p \leq Q'} L^*(N/p) \\ &= \int_Q^{Q'} \frac{L^*(N/x)}{\log x}\, dx + \frac{\vartheta(x) - x}{\log x} L^*(N/x)\Big|_Q^{Q'} \\ &\quad - \int_Q^{Q'} (\vartheta(x) - x)\frac{d}{dx}\left(\frac{L^*(N/x)}{\log x}\right) dx \\ &= S_1 + S_2 + S_3, \quad \text{say.}\end{aligned}$$

We shall estimate the absolute values of S_2 and S_3 and then the value of S_1, the main term. We use the estimate [4, p. 70] $|\vartheta(x) - x| < x/(2 \log x)$ to obtain

$$|S_2| \le \frac{N}{\log^2 N} \prod_4^r \log_j N$$

as follows:

$$|S_2| \le \frac{Q'}{2 \log^2 Q'} L^*(N/Q') + \frac{Q}{2 \log^2 Q} L^*(N/Q)$$

$$= \frac{NL^*(\log_2 N)}{2 \log_2 N (\log N - \log_3 N)^2} + \frac{NL^*\left(\frac{\log N}{2}\right)}{\log N (\log N + \log 2 - \log_2 N)^2}$$

$$\le \frac{N}{2 \log^2 N \cdot \log_2 N \left(1 - \frac{\log_3 N}{\log N}\right)^2} \cdot \frac{\log_2 N}{\log_3 N} \prod_5^{r+1} \log_j N$$

$$+ \frac{N}{\log^3 N \left(1 + \frac{\log 2 - \log_2 N}{\log N}\right)^2} \cdot \frac{\log N}{2 (\log_2 N - \log 2)} \prod_4^r \log_j N$$

$$\le \frac{N}{2 \log^2 N} \prod_4^r \log_j N$$

$$\cdot \left(\frac{\log_{r+1} N}{\log_3 N \log_4 N \left(1 - \frac{\log_3 N}{\log N}\right)^2} + \frac{1}{\left(1 - \frac{\log_2 N}{\log N}\right)^2 (\log_2 N - \log 2)} \right)$$

$$\le \frac{N}{\log^2 N} \prod_4^r \log_j N.$$

A straightforward calculation yields

$$\left| \frac{d}{dx} \frac{L^*(N/x)}{\log x} \right| \le \frac{N}{x^2 \log x \log N/x} \prod_3^{r-1} \log_j N/x$$

for x in the prescribed range. Thus

$$|S_3| \le \int_Q^{Q'} \frac{N}{2x \log^2 x \log N/x} \prod_3^{r-1} \log_j N/x \, dx.$$

Using the facts that $N/x \le \log N$ and $2\log^2 x \ge (3/2)\log^2 N$ for all x in the range of integration, we see that

$$
\begin{aligned}
|S_3| &\le \frac{2N\prod_4^r \log_j N}{3\log^2 N}\int_Q^{Q'} \frac{dx}{x\log N/x} \\
&= \frac{2N\prod_4^r \log_j N}{3\log^2 N}\left(-\log_2 N/x|_Q^{Q'}\right) \\
&\le \frac{2N\prod_4^r \log_j N}{3\log^2 N}\left(-\log_2 N/x|_{N/\log N}^{Q'}\right) \\
&\le \frac{N\prod_3^r \log_j N}{\log^2 N}.
\end{aligned}
$$

We next obtain a lower bound for S_1:

$$
\begin{aligned}
S_1 &= \int_Q^{Q'} \frac{N}{x\log x\log N/x}\prod_3^{r-1}\log_j N/x\,dx \\
&\ge \frac{N}{\log N}\int_Q^{Q'} \frac{\prod_3^{r-1}\log_j N/x}{x\log N/x}\,dx.
\end{aligned}
$$

With $u = \prod_3^{r-1}\log_j N/x$ and $v = -\log_2 N/x$ we integrate by parts to obtain

$$
\begin{aligned}
\int_Q^{Q'} &\frac{1}{x\log N/x}\cdot\prod_3^{r-1}\log_j N/x\,dx \\
&= -\prod_2^{r-1}\log_j N/x|_Q^{Q'} - \int_Q^{Q'}\frac{1}{x\log N/x}\left(\sum_{i=3}^{r-1}\prod_{j=i+1}^{r-1}\log_j N/x\right)dx \\
&\ge \prod_2^{r-1}\log_j N/Q - \prod_4^{r+1}\log_j N - 2\prod_5^r\log_j N/Q\int_Q^{Q'}\frac{dx}{x\log N/x} \\
&\ge \prod_3^r \log_j N\left(1-\frac{5}{2\log_4 N}\right),
\end{aligned}
$$

where we have used that

$$
\prod_2^{r-1}\log_j\frac{x}{2} \ge \left(1-\frac{2}{\log x}\right)\prod_2^{r-1}\log_j x \quad \text{for } \log_2 N \le x \le \log N.
$$

Substituting this in the lower bound for S_1 we obtain

$$
S_1 \ge \frac{N}{\log N}\cdot\prod_3^r\log_j N\left(1-\frac{5}{2\log_4 N}\right).
$$

Combining the estimates for S_1, $|S_2|$ and $|S_3|$ we obtain

$$\left(\prod_{j=4}^{r}\left(1-\frac{3}{\log_{2j-2} N}\right)\right)^{-1} S^*(N)$$

$$\geq \frac{N}{\log N}\prod_{3}^{r}\log_j N\left\{1-\frac{5}{2\log_4 N}-\frac{2}{\log N\log_3 N}-\frac{1}{\log N}\right\}$$

$$\geq \left(1-\frac{3}{\log_4 N}\right)\frac{N}{\log N}\prod_{3}^{r}\log_j N$$

which satisfies (*). Thus (*) holds for all $r \geq 1$.

Since we know $\log_{2r} N \geq 1$ we deduce that $\log_{2j-2} N \geq e^{2r-2j+2}$. Thus

$$\prod_{j=3}^{r}\left(1-\frac{3}{\log_{2j-2} N}\right) \geq \prod_{j=3}^{r}\left(1-\frac{3}{e^{2r-2j+2}}\right)$$

$$= \prod_{j=1}^{r-2}\left(1-\frac{3}{e^{2j}}\right)$$

$$\geq \prod_{1}^{\infty}\left(1-\frac{3}{e^{2j}}\right)$$

$$\geq 1/e,$$

where the last inequality follows from the facts that for $0 \leq x \leq 3/e^2 = 0.406\ldots$, $\log(1-x) \geq -3x/2$ and $-(3/2)\sum_{i=1}^{\infty} 3/e^{2j} = -0.526\cdots > -1$.

The theorem is proved.

LEMMA 7. *For* $N \geq 1$, $S(N) \leq 2^N$.

Proof. The result follows immediately since there are 2^N distinct choices for ε_i, $1 \leq i \leq N, \varepsilon_i = 0$ or 1.

LEMMA 8. *For* $\log_2 N \geq 1$, $S(N) \leq \exp(N/\log_2 N)$.
For $\log_4 N \geq 1$, $S(N) \leq \exp(N\log_2 N/\log N)$.

Proof of Lemma 8. Let $Q = N/\log N$. Let

$$\mathscr{P} = \{p: Q < p \leq N\},$$

$$Z_1 = \{k \leq N: \text{there exists } p \in \mathscr{P}, p \mid k\}$$

and

$$Z_2 = \{k \leq N: k \notin Z_1\}.$$

Thus we may write

$$\sum_{k=1}^{N}\frac{\varepsilon_k}{k} = \sum_{k\in Z_1}\frac{\varepsilon_k}{k} + \sum_{k\in Z_2}\frac{\varepsilon_k}{k}.$$

Let $S_i(N)$ denote the number of distinct values of the sum with $k \in Z_i$ as the ε_k's take on all possible values with $\varepsilon_k = 0$ or 1. As before $S^*(N) = \log S(N)$ and $S_i^*(N) = \log S_i(N)$, $i = 1, 2$.

The case $\log_2 N \geq 1$.

Subcase A. $N \geq 10^8$. We estimate $S_1^*(N)$ first. From the definition of Z_1 we see that

$$|Z_1| = \sum_{p \in \mathscr{P}} \left[\frac{N}{p}\right] \leq N \sum_{p \in \mathscr{P}} \frac{1}{p}.$$

Using the estimates of [4, Theorem 5 and corollary], we obtain

$$|Z_1| \leq N\left(\log_2 N - \log_2 Q + \frac{1}{\log^2 N} + \frac{1}{2\log^2 Q}\right).$$

Since $S_1(N) \leq 2^{|Z_1|}$, it follows that

$$S_1^*(N) \leq N(\log 2)\left(\log_2 N - \log_2 Q + \frac{1}{\log^2 N} + \frac{1}{2\log^2 Q}\right). \tag{1}$$

We now estimate $S_2^*(N)$. Suppose $\sum_{k \in Z_2} \varepsilon_k/k = a/b$, then independent of the choice of the ε_k's we may choose $b = \text{l.c.m. } Z_2$. From the definitions of $\psi(x)$ and $\vartheta(x)$ [2, pp. 340–341] we deduce that $\log b = \psi(N) - (\vartheta(N) - \vartheta(Q))$. Since $\psi(x) = \sum_{k=1}^{\infty} \vartheta(x^{1/k})$, one can show $\psi(x) - \vartheta(x) < 1.5x^{1/2}$ (see [4, Theorem 13]). Hence we see that $\log b \leq \vartheta(Q) + 1.5\sqrt{N}$. On the other hand

$$\frac{a}{b} \leq \sum_{i=1}^{N} \frac{1}{i} \leq \log N + \gamma + \frac{1}{N}$$

where $\gamma = 0.57\cdots$ is Euler's constant. Thus we see that the number of distinct possibilities for a is at most $b(\log N + \gamma + 1/N)$. It follows that

$$S_2(N) \leq (\log N + \gamma + 1/N)\exp(\vartheta(Q) + 1.5\sqrt{N}).$$

Whence

$$S_2^*(N) \leq \log(\log N + \gamma + 1/N) + \vartheta(Q) + 1.5\sqrt{N}. \tag{2}$$

Since $S^*(N) \leq S_1^*(N) + S_2^*(N)$ we can now estimate $S^*(N)$.

By the above estimates (1) and (2) for $S_1^*(N)$ and $S_2^*(N)$ we get

$$\begin{aligned} S^*(N) \leq \frac{N}{\log_2 N}\Bigg\{&\log 2\left((\log_2 N)^2 - \log_2 Q \log_2 N + \frac{\log_2 N}{(\log N)^2} + \frac{\log_2 N}{2(\log Q)^2}\right) \\ &+ \frac{\log(\log N + \gamma + 1/N)\cdot \log_2 N}{N} \\ &+ \frac{1.02\log_2 N}{\log N} + \frac{1.5\log_2 N}{\sqrt{N}}\Bigg\} \end{aligned}$$

where we have used [4, Theorem 9] for the penultimate term. A straightforward calculation shows that for $\log_2 N \geq 1$ the term in the braces is decreasing when $N \geq 10^8$, and is less than 1.

Subcase B. $10^8 \geq N \geq e^e$. If $\log_2 N \leq 1/\log 2 = 1.4\cdots$, i.e., $N \leq 68.8\cdots$, then $2^N \leq \exp(N/\log_2 N)$ and the desired inequality holds.

For $N = 69, 70, 71, 72$, or 73 we note by direct calculation from the definition that $|Z_1| \leq 23 \leq N\cdot(23/69) = N/3$. Thus

$$S_1^*(N) \leq \frac{N\log 2}{3} \leq \frac{N\log 2}{\log_2 N}\cdot\left(\frac{1}{2}\right),$$

$$\begin{aligned} S_2^*(N) &\leq \log(\log(N+\gamma+1/N) + \vartheta(Q) + 1.5\sqrt{N}) \\ &\leq \frac{N}{\log_2 N}\left\{\frac{\log(\log(N+\gamma+1/N))\log_2 N}{N} + \frac{\log_2 N}{\log N}\frac{1.5\log_2 N}{\sqrt{N}}\right\}. \end{aligned}$$

Since $S^*(N) \leq S_1^*(N) + S_2^*(N)$ we obtain

$$S^*(N) \leq \frac{N}{\log_2 N}\left\{\frac{\log 2}{2} + \frac{\log(\log(N+1))\log_2 N}{N} + \frac{\log_2 N}{\log N} + \frac{1.5\log_2 N}{\sqrt{N}}\right\}.$$

Since the term in braces is less than 1 for $69 \leq N < 74$, the inequality hold for $N < 74$.

For $74 \leq N \leq 10^8$ we use the estimates of [4, Theorems 18, 20, and 13] to obtain the desired result in a manner analogous to the case when $N \geq 10^8$. The difference in the cases $74 \leq N \leq 10^8$ and $N \geq 10^8$ are all consequences of the different estimates for $\sum 1/p$ and $\vartheta(x)$. The calculations are left to the reader.

Thus the first half of Lemma 8 is established.

The case $\log_4 N \geq 1$. In this case $N \geq 10^8$. From (1) and (2) we get

$$\begin{aligned} S^*(N) \leq \frac{N\log_2 N}{\log N}\Bigg\{&\log 2\left(\log N - \frac{\log_2 Q\log N}{\log_2 N}\right. \\ &\left.+ \frac{1}{\log N\log_2 N} + \frac{\log N}{2\log_2 N\log^2 Q}\right) \\ &+ \left(\frac{\log(\log N+1)\log N}{N\log_2 N} + \frac{1.02}{\log_2 N} + \frac{1.5\log N}{\sqrt{N}\log_2 N}\right)\Bigg\}. \end{aligned}$$

Using the estimates

$$\log N - \frac{\log_2 Q\log N}{\log_2 N} \leq 1 + \frac{\log_2 N}{\log N}$$

in the above inequality yields

$$\begin{aligned} S^*(N) \leq \frac{N\log_2 N}{\log N}\Bigg\{&\log 2\left(1 + \frac{\log_2 N}{\log N} + \frac{1}{\log N\log_2 N} + \frac{\log N}{2\log_2 N\log^2 Q}\right) \\ &+ \left(\frac{\log(\log N+1)}{\log N\log_2 N}\cdot\frac{\log^2 N}{N} + \frac{1.02}{\log_2 N} + \frac{1.5\log N}{\sqrt{N}\log_2 N}\right)\Bigg\}. \end{aligned}$$

An easy calculation shows that in the range under consideration, $\log_4 N \geq 1$, each term in the parentheses is decreasing. Trivial numerical estimates show that for $\log_4 N = 1$ the quantity in braces is less than 1.

Lemma 8 is proved.

LEMMA 9. *Let* $Q = N/\log N$ *and* $Q' = N/\log_2 N$. *Suppose that* $\log_6 N \geq 1$. *Then*

$$\sum_{Q<p\leq Q'} \frac{1}{p \log (N/p)} \leq \frac{\log_3 N}{\log N}\left(1 - \frac{\log_4 N}{2 \log_3 N}\right).$$

Proof. This is proved by using Lemma 6 almost exactly the same way it was used in the paragraphs following its proof, except that in this case $f(x)$ is simpler and slight adjustments must be made since we are deriving an upper bound. The details are left to the reader.

THEOREM 3. *For* $r \geq 1$ *and* $\log_{2r} N \geq 1$,

$$S(N) \leq \exp\left(\frac{N \log_r N}{\log^2 N \log_2 N} \prod_{j=1}^{r} \log_j N\right).$$

Proof. The values $r = 1, 2$ yield the statements of Lemma 8. We suppose the result is true for $r - 1 \geq 2$ and show that it holds for r.

We divide the integers less than N in a way similar to that in the proof of Theorem 2. Let $Q = N/\log N$ and $Q' = N/\log_2 N$. We define Z_1 and Z_2 by

$$Z_1 = \{k \leq N: \text{there exists } p,\, Q < p < N,\, p \mid k\}$$

and

$$Z_2 = \{k \leq N: k \notin Z_1\}.$$

Thus

$$\sum_{k=1}^{N} \frac{\varepsilon_k}{k} = \sum_{k \in Z_1} \frac{\varepsilon_k}{k} + \sum_{k \in Z_2} \frac{\varepsilon_k}{k}.$$

If $S_i(N)$ denotes the number of distinct values of the sums over Z_i as the ε_k's take on all possible values with $\varepsilon_k = 0$ or 1, then $S(N) \leq S_1(N)S_2(N)$. We estimate each of $S_1(N)$ and $S_2(N)$ separately. Let $S_i^*(N) = \log S_i(N)$; then $S^*(N) \leq S_1^*(N) + S_2^*(N)$.

We estimate $S_2^*(N)$ first. For any choice of ε_k's we may write

$$\sum_{k \in Z_2} \frac{\varepsilon_k}{k} = \frac{a}{b} \quad \text{where } a < \left(\sum_{i=1}^{N} \frac{1}{i}\right) b \text{ and } b = \text{l.c.m.}\,(Z_2).$$

As in the proofs of Lemma 8, we obtain from (2),

$$\begin{aligned} S_2^*(N) &\leq \log(\log N + 1) + \vartheta(Q) + 1.5\sqrt{N} \\ &\leq \log_2 N + 1/\log N + N/\log N + N/\log^2 N + 1.5\sqrt{N} \qquad (4) \\ &\leq 2N/\log N \end{aligned}$$

where we have used [4, Theorem 4] and $1/(2 \log Q) < 1/\log N$ for the values of N under consideration.

We now turn to an estimation of $S_1(N)$. We rewrite the sum as follows

$$\sum_{k \in Z_1} \frac{\varepsilon_k}{k} = \sum_{Q<p<N} \frac{1}{p}\left(\sum_{k=1}^{N/p} \frac{\varepsilon_k}{k}\right)$$

where the ε_k's on the internal sums (which properly should be $\varepsilon_{p,k}$) are independently taking on all possible combinations of values of 0 or 1. We see from this representation that

$$S_1^*(N) \le \sum_{Q<p\le N} S^*(N/p).$$

We break the sum in two parts as follows:

$$\Sigma_1 = \sum_{Q<p\le Q'} S^*(N/p), \qquad \Sigma_2 = \sum_{Q'<p\le N} S^*(N/p). \tag{5}$$

Notice that for $Q < p \le Q'$ we have $N/p \ge \log_2 N$ and thus

$$\log_{2(r-1)} N/p \ge \log_{2r} N \ge 1$$

so that the induction hypothesis for $r - 1$ is satisfied for N/p in the first sum. For the second sum we will use the estimates of Lemmas 7 and 8 which yield $S^*(x) \le x \log 2$ and $S^*(x) \le (x \log_2 x)/\log x$. We estimate Σ_2 first.

$$\Sigma_2 \le \sum_{Q'<p\le N/E} \frac{N \log_2 N/p}{p \log N/p} + \sum_{N/E<p\le N} \frac{N}{p} \log 2$$

where E is chosen so that $\log_4 E = 1$. The first sum can be estimated by the use of Lemma 6 with

$$f(p) = \frac{\log_2 (N/p)}{p \log (N/p)}.$$

After some calculation one gets

$$\sum_{Q'<p<N/E} f(p) \le \frac{N \log_4^2 N}{\log N}.$$

Using the standard estimates [4, Theorem 5] for $\sum 1/p$ one obtains

$$\sum_{N/E<p\le N} \frac{N}{p} \log 2 \le \frac{N \log E}{\log N}.$$

We thus obtain

$$\Sigma_2 \le \frac{N}{\log N} (\log_4^2 N + \log E). \tag{6}$$

We now estimate Σ_1 from (5), where we substitute for $S^*(N/p)$ the bound given by the induction hypothesis to obtain

$$\begin{aligned} \Sigma_1 &\le \sum_{Q<p\le Q'} \frac{N \log_{r-1} (N/p)}{p \log^2 (N/p) \log_2 (N/p)} \prod_{j=1}^{r-1} \log_j (N/p) \\ &< \frac{N \log_{r-1} N/Q}{\log N/Q \log_2 N/Q} \prod_{j=1}^{r-1} \log_j (N/Q) \sum_{Q<p\le Q'} \frac{1}{p \log N/p}, \end{aligned} \tag{7}$$

where we have used the fact that

$$\frac{\log_{r-1} N/x}{\log N/x \log_2 N/x} \prod_{j=1}^{r-1} \log_j (N/x)$$

is decreasing in the interval $Q \leq x \leq Q'$ since the two terms in the denominator cancel into the numerator and the rest of the numerator is clearly decreasing in x. But $N/Q = \log N$ and $\sum 1/(p \log N/p)$ can be estimated by Lemma 9; thus

$$\Sigma_1 < \frac{N \log_r N}{\log_2 N \log_3 N}\left(\prod_{j=2}^{r} \log_j N\right)\frac{\log_3 N}{\log N}\left(1 - \frac{\log_4 N}{2 \log_3 N}\right).$$

The above can be rewritten as

$$\Sigma_1 < \frac{N \log_r N}{\log^2 N \log_2 N}\left(\prod_{j=1}^{r} \log_j N\right)\left(1 - \frac{\log_4 N}{2 \log_3 N}\right). \tag{8}$$

We combine (4), (6), and (8) to obtain

$$\begin{aligned} S^*(N) \leq{}& \frac{N \log_r N}{\log N}\left(\prod_{j=3}^{r} \log_j N\right) \\ &\times \left\{1 - \frac{\log_4 N}{2 \log_3 N} + \frac{\log_4^2 N + \log E}{\log_r N \prod_{j=3}^{r} \log_j N} + \frac{2}{\log_r N \prod_{j=3}^{r} \log_j N}\right\}. \end{aligned} \tag{9}$$

It is not difficult to verify that the quantity in braces in (9) is less than 1; hence,

$$S^*(N) < \frac{N \log_r N}{\log N} \prod_{j=3}^{r} \log_j N. \tag{10}$$

But (10) is clearly equivalent to the inequality of Theorem 3, which is thus proven.

IV. A lower bound for $D(P)$

The proof is virtually the same as that for Theorem 2 of [1] except that we have a better bound for $S(N)$.

THEOREM 4. *If P is a prime then for P large enough that* $\log_{2r} P \geq 1$

$$D(P) \geq \frac{P \cdot \log P \cdot \log_2 P}{\log_{r+1} P \prod_{j=4}^{r+1} \log_j P}.$$

Proof. For each a/P, $1 \leq a < P$, write

$$\frac{a}{P} = \frac{1}{P}\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_{t_a}}\right) + \frac{1}{y_1} + \frac{1}{y_2} \cdots \frac{1}{y_{s_a}}$$

where $x_i < x_{i+1}$, $(x_i, P) = (y_i, P) = 1$, and x_{i_a} is minimal for all expansions of a/P. Let $N = \max\{x_{i_a} : 1 \leq a < P\}$. Each value of a requires a different value of

$$\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_{t_a}} = \sum_{k=1}^{N} \frac{\varepsilon_k}{k}$$

for some choice of ε_k's. Thus N must be such that $S(N) \geq P$, the value $a = 0$ corresponding to the choice of all $\varepsilon_k = 0$. From Theorem 3 we see that for P large enough that $\log_{2r} P \geq 1$, N must be bigger than

$$\frac{\log P \cdot \log_2 P}{\log_{r+1} P \prod_{j=4}^{r+1} \log_j P},$$

since for that value $S^*(N) < \log P$. The desired inequality follows.

There are both heuristic and experimental reasons to suppose that the order of $D(N)/N$ is largest for $N = P$, a prime. This could be established if one could prove that for $(M, N) = 1$, $D(MN) \leq D(M) \cdot D(N)$, since we already know [1, Theorem 5] that $D(P^k) \leq 2D(P)P^{k-1}$. Exact estimates for $D(P)$ seem difficult since $D(P)/P$ is not monotone.

Bibliography

1. M. N. Bleicher and P. Erdös, *Denominators of Egyptian fractions*, J. Number Theory, vol. 8 (1976), to appear.
2. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, fourth edition, Oxford Univ. Press, Oxford, 1962.
3. K. Prachar, *Primzahlverteilung*, Springer-Verlag, Heidelberg, 1957.
4. J. B. Rosser and L. Schoenfeld, *Approximate formula for some functions of prime numbers*, Illinois J. Math., vol. 6 (1962), pp. 64–94.

University of Wisconsin
Madison, Wisconsin