MATHEMATICS OF COMPUTATION, VOLUME 29, NUMBER 129
JANUARY 1975, PAGES 29–42

The Number of Distinct Subsums of $\sum_1^N 1/i$

By M. N. Bleicher and P. Erdös

Abstract. In this paper we improve the lower bounds for the number, $S(N)$, of distinct values obtained as subsums of the first N terms of the harmonic series. We obtain a bound of the form

$$S(N) \geqslant e\left(\frac{N \log 2}{\log N} \prod_{3}^{k+1} \log_j N\right)$$

whenever $\log_{k+1} N \geqslant k + 1$, for $k \geqslant 3$. Slight modifications are needed for $k = 1, 2$. We begin by discussing the number $Q_k(N)$ of integers $n \leqslant N, n = p_1 p_2 \cdots p_k$, where $p_i > e^{\alpha p_{i-1}}$, $i = 2, \cdots, k$. We prove that

$$\frac{N}{\log N} \prod_{i=1}^{k+1} \log_i N \leqslant Q_k(N) \leqslant \left(1 + \frac{k}{\log_{k+1} N}\right) \frac{N}{\log N} \prod_{i=3}^{k+1} \log_i N.$$

This bound is valid for $\log_{k+1} N \geqslant k + 1$ and for $1 \leqslant \alpha \leqslant 2(1 - e_2(4)/e_3(4))$. The symbols $\log_i x$ and $e_i(x)$ are defined by

$$e_0(x) = x, \qquad e_{i+1}(x) = e^{e_i(x)},$$

$$\log_0 x = x, \qquad \log_{i+1} x = \log(\log_i x),$$

where $\log x$ denotes the logarithm to the base e.

In this paper we improve the lower bounds given in [2] and [3] for the number, $S(N)$, of distinct values obtained as subsums of the first N terms of the harmonic series. The estimates in [1], [2] and [3] were derived because the upper bound was needed for lower estimates of the denominators of Egyptian fractions. In this paper we concentrate on the lower bounds. We obtain a bound of the form

$$S(N) \geqslant e\left(\frac{N \log 2}{\log N} \prod_{3}^{k+1} \log_j N\right)$$

whenever $\log_{k+1} N \geqslant k + 1$, for $k \geqslant 3$. Slight modifications are needed for $k = 1, 2$; see Corollaries 1, 2, 3 and 4 for more details. In order to do this we begin by discussing the number $Q_k(N)$ of integers $n \leqslant N, n = p_1 p_2 \cdots p_k$ where $p_i > e^{\alpha p_{i-1}}$, $i = 2, \cdots, k$. We first prove that

Received July 26, 1974.

AMS (MOS) subject classifications (1970). Primary 10A99, 10A25; Secondary 10B25.

Key words and phrases. Harmonic series, squarefree integers, prime numbers, distinct subsums.

$$\frac{N}{\log N}\prod_{i=3}^{k+1}\log_i N \leqslant Q_k(N) \leqslant \left(1+\frac{k}{\log_{k+1} N}\right)\frac{N}{\log N}\prod_{i=3}^{k+1}\log_i N.$$

This bound is valid for $\log_{k+1} N \geqslant k+1$ and for $1 \leqslant \alpha \leqslant 2(1-e_2(4)/e_3(4))$. The bounds on N and α are for convenience in evaluating the range of validity and the constants in the inequality, not for essential reasons. The symbols $\log_i x$ and $e_i(x)$ are defined by

$$e_0(x)=x, \qquad e_{i+1}(x)=e^{e_i(x)},$$

$$\log_0 x = x, \qquad \log_{i+1} x = \log(\log_i x),$$

where $\log x$ denotes the logarithm to the base e.

In fact we prove the following slightly stronger version.

THEOREM. *If* $1 \leqslant \alpha \leqslant 2(1-e_2(4)/e_3(4)) = 1.999\cdots$, *then*:

For $k=1$,

$$\frac{N}{\log N}\left(1+\frac{1}{2\log N}\right) \leqslant Q_1(N)=\pi(N) \leqslant \frac{N}{\log N}\left(1+\frac{3}{2\log N}\right),$$

where the lower bound holds for $N \geqslant 59$ *and the upper bound for* $N \geqslant 2$; $Q_1(N)=0$ *for* $N<2$.

For $k=2$,

$$\frac{N}{\log N}\left(\log_3 N+\frac{1}{11}\right) \leqslant Q_2(N) \leqslant \frac{N}{\log N}(\log_3 N+2)$$

where the lower bound holds for $\log_3 N \geqslant 2$ *and the upper bound for* $N \geqslant e_3(-2) = 3.1\cdots$ (*i.e.*, $\log_3 N \geqslant -2$); $Q_2(N)=0$ *for* $N<22$.

For $k \geqslant 3$,

$$\frac{N}{\log N}\prod_{3}^{k+1}\log_j N \leqslant Q_k(N) \leqslant \frac{N(\log_{k+1} N+k)}{\log N}\prod_{3}^{k}\log_j N,$$

where the lower bound holds for $\log_{k+1} N \geqslant k+1$ *and the upper bound holds for* $N \geqslant e_{k+1}(-2)$; $Q_k(N)=0$ *for* $N \leqslant e_{k+1}(-.13\cdots) = e_{k-2}(11)$.

Proof. Case 1. $k=1$. In this case $Q_1(N)=\pi(N)$, so that the result is well known, see [4, p. 69].

Case 2. $k=2$. Let $\mathcal{Q}_2(N)$ be those integers counted by $Q_2(N)$; namely

$$\mathcal{Q}_2(N)=\{pq\colon p, q \text{ prime}, e^{\alpha p}<q, pq \leqslant N\}.$$

The Upper Bound for $Q_2(N)$. Let L be the number which satisfies $e^{\alpha L}\cdot L = N$. It follows that

$$Q_2(N)=\sum_{2\leqslant p<L}(\pi(N/p)-\pi(e^{\alpha p})), \tag{1}$$

where p runs through the primes in the indicated interval. We see from the conditions on α that

(2) $$L \leqslant \log N.$$

We thus deduce that

(3) $$Q_2(N) \leqslant \sum_{2 \leqslant p \leqslant \log N} \frac{N}{p \log N/P}\left(1+\frac{3}{2 \log N/P}\right).$$

Since $\log N/P$ is almost constant on the interval under consideration, we obtain

(4) $$Q_2(N) \leqslant \frac{N}{\log(N/\log N)}\left(1+\frac{3}{2\log(N/\log N)}\right)\sum_2^{\log N}\frac{1}{p}.$$

The value of $\Sigma\, 1/p$ is well known, for example see [4, p. 70]. Thus we obtain

(5) $$Q_2(N) \leqslant \frac{N}{\log N}\left(1+\frac{2\log_2 N}{\log N}\right)\left(\log_3 N + B + \frac{1}{\log_2^2 N}\right),$$

which is valid for $N \geqslant 3$ and where $B = .26149 \cdots$. If $N \geqslant e^4$, i.e., $\log_3 N \geqslant \log_2 4 > .326 \cdots$, then this can be simplified to

(6) $$Q_2(N) \leqslant N(\log_3 N + 2)/\log N$$

If $22 \leqslant N \leqslant e^4 < 55$, then $Q_2(N) \leqslant Q_2(54) = 5$ together with $\log_3 N \geqslant 0$ gives the upper bound of the theorem for $k = 2$.

The Lower Bound for $Q_2(N)$. From the definition of $Q_2(N)$ we obtain

(7) $$Q_2(N) = \sum_{1 \leqslant p \leqslant N}\ \sum_{1 \leqslant q \leqslant M} 1,$$

where p and q run over primes in the indicated intervals and $M = \min\{N/p, \log p/\alpha\}$. Let L be such that

(8) $$\alpha N = L \log L,$$

so that $N/\log N < L < eN/\log N$, then

(9) $$Q_2(N) = \sum_{1 \leqslant p \leqslant L}\ \sum_{1 \leqslant q \leqslant (\log p)/\alpha} 1 + \sum_{L < p \leqslant N}\ \sum_{1 \leqslant q \leqslant N/P} 1.$$

Let Σ_1 denote the first double sum and Σ_2 the second. Since $\Sigma_1 \geqslant 0$ we can obtain a lower bound for $Q_2(N)$ by obtaining a lower bound for Σ_2.

The Bounds for Σ_2. From the definition of Σ_2 in (9) we obtain

(10) $$\Sigma_2 = \sum_{L < p \leqslant L'} \pi(N/P) + \sum_{L' < p \leqslant N/2} \pi(N/P),$$

where $L < L' = N/p_l$, p_l is the lth prime with $l \geqslant 7$ to be determined later. We note that

$$\sum_{L'<p\leqslant N/2} \pi(N/p) = \sum_{2\leqslant p\leqslant p_l} \pi(N/p) - l\pi(N/p_l). \tag{11}$$

We shall frequently need to estimate sums of the above type where the index of the summation range over an interval of primes. There is a standard technique for converting the sum to a Stieltjes integral, with respect to $d\vartheta(x)$, integrating by parts twice with $\vartheta(x)$ approximated by x in between to obtain the following well-known lemma.

LEMMA. *If $f(x) \geqslant 0$ and $f'(x)$ exists and is continuous and $0 < a < b$*

$$\sum_{a<p\leqslant b} f(p) = \left.\frac{f(x)(\vartheta(x)-x)}{\log(x)}\right|_a^b + \int_a^b \frac{f(x)}{\log x}\,dx - \int_a^b (\vartheta(x)-x)\frac{d}{dx}\left(\frac{f(x)}{\log x}\right)dx.$$

We recall from [4] the estimates

$$|\vartheta(x) - x| \leqslant x/(2\log x) \quad \text{for } x \geqslant 563 \tag{12}$$

and

$$\vartheta(x) - x \leqslant x/(2\log x) \quad \text{for } x > 1 \tag{13}$$

and the estimates

$$\frac{x}{\log x}\left(1 + \frac{1}{2\log x}\right) < \pi(x) \quad \text{for } x \geqslant 59, \tag{14}$$

$$\frac{x}{\log x} < \pi(x) \quad \text{for } x \geqslant 17, \tag{15}$$

and

$$\pi(x) < \frac{x}{\log x}\left(1 + \frac{3}{2\log x}\right) \text{ for } x > 1. \tag{16}$$

We use (15) which holds for $N \geqslant 73$ and the lemma to estimate the first sum of (10); thus

$$\sum_{L<p\leqslant L'} \frac{N}{p\log N/p} = N\left\{\left.\frac{\vartheta(x)-x}{x\log x\log N/x}\right|_L^{L'} + \int_L^{L'} \frac{dx}{x\log x\log N/x} - \int_L^{L'} (\vartheta(x)-x)\frac{d}{dx}\left(\frac{1}{x\log x\log N/x}\right)dx\right\}. \tag{17}$$

We next show that

$$\left|\int_L^{L'} (\vartheta(x) - x)\,\frac{d}{dx}\left(\frac{1}{x \log x \log N/x}\right) dx\right| \leqslant \frac{\log_3 N}{2 \log^2 N}. \tag{18}$$

To do this we note that

$$\left|\frac{d}{dx}\left(\frac{1}{x \log x \log N/x}\right)\right| \leqslant \frac{1}{x^2 \log x \log N/x}$$

and that the estimate of (12), $|\vartheta(x) - x| < x/2 \log x$ are both valid for the range $N/\log N \leqslant x \leqslant N/2$ when $N \geqslant e^{8.5}$. Thus

$$\left|\int_L^{L'} (\vartheta(x) - x)\,\frac{d}{dx}\left(\frac{1}{x \log x \log N/x}\right) dx\right| \leqslant \int_L^{L'} \frac{dx}{2x \log^2 x \log N/x}. \tag{19}$$

Since $1/2 \log^2 x$ is almost constant on the interval involved it can be brought out of the integral and replaced by $1/2 \log^2 L$; what remains is the derivative of $-\log_2 N/x$, and we get

$$\int_L^{L'} \frac{dx}{2x \log^2 x \log N/x} \leqslant \frac{1}{2 \log^2 L}\,(-\log_2 N/x)\Big|_L^{L'}, \tag{20}$$

which yields (18).

We next evaluate the first integral in (17) by taking the $1/\log x$ outside the integral as $1/\log L'$ and integrating the rest exactly to obtain

$$\frac{\log_3 N}{\log N}\left(1 + \frac{\log_2 p_l}{\log_3 N} + \frac{\log_2 p_l}{\log N}\right) \leqslant \int_L^{L'} \frac{dx}{x \log x \log N/x}. \tag{21}$$

We next note that

$$\left|\left(\frac{\vartheta(x) - x}{x \log x \log N/x}\Big|_L^{L'}\right)\right| \leqslant \frac{1}{2 \log^2 L \log N/L} + \frac{1}{2 \log^2 L' \log N/L'} \leqslant \frac{1}{2 \log^2 N}. \tag{22}$$

Using (15) and (16), (11) and $N/p_l \geqslant 17$, which holds since $p_l < \log N$ and $\log_3 N \geqslant 2$, we deduce

$$\begin{aligned}\sum_{N/p_l < p \leqslant N/2} \pi\left(\frac{N}{p}\right) &= \sum_{2 \leqslant p \leqslant p_l} \pi\left(\frac{-N}{p}\right) - l\pi\left(\frac{N}{17}\right) \\ &\geqslant \frac{N}{\log N/p_l}\left(\sum_{2 \leqslant p \leqslant p_l} \frac{1}{p} - \frac{l}{p_l}\left(1 + \frac{3}{2 \log N/p_l}\right)\right).\end{aligned} \tag{23}$$

If $l/p_l < B$, then using $N \geqslant e_3(2) > e^{1600}$ and $p_l < \log N$,

$$\sum_{N/p_l \leqslant p \leqslant N/2} \pi\left(\frac{N}{p}\right) \geqslant \frac{N}{\log N}\left(\log_2 p_l + B - \frac{1}{2 \log^2 p_l} - \frac{l}{p_l} + \frac{\log p_l}{\log N}\right). \tag{24}$$

Now with the aid of (10), (11) and (24) as well as (17), (21), (22) and (24) we obtain for $\log_3 N \geqslant 2$ and $l/p_l \leqslant B$,

$$\sum_2 \geqslant \frac{N \log_3 N}{\log N}\left(1 - \frac{\log_2 p_l}{\log_3 N} + \frac{\log p_l}{\log N} - \frac{1}{2 \log N \log_3 N}\right.$$

$$- \frac{1}{2 \log N} + \frac{\log_2 p_l}{\log_3 N} + \frac{B - l/p_l}{\log_3 N} \tag{25}$$

$$\left. - \frac{1}{2 \log^2 p_l \log_3 N} + \frac{\log p_l}{\log_3 N \log N}\right).$$

Taking $p_l = 1597$, $l = 251$ so that all the previous conditions are satisfied and using $B = .261 \cdots$, $l/p_l = .157 \cdots$, $1/2 \log^2 p_l < .0005$ and $\log_3 N \geqslant 2$, we deduce

$$\sum_2 \geqslant \frac{N \log_3 N}{\log N}\left(1 + \frac{1}{11 \log_3 N}\right). \tag{26}$$

Since $Q_2(N) \geqslant \Sigma_1 + \Sigma_2$ and by (13), $\Sigma_1 \geqslant 0$, (26) implies the desired lower bound of the theorem for the case $k = 2$.

Case 3. $k \geqslant 3$. We now proceed by induction on k. Suppose $k > 2$ and that for $2 \leqslant k' < k$ the theorem is true for k replaced by k'; we now show it is true for k.

The Lower Bound for $Q_k(N)$. Let $\mathcal{Q}_k(N)$ denote the set of integers counted by $Q_k(N)$. As before let $L = N/\log N$. We claim that

$$\mathcal{Q}_k(N) \supset \bigcup_{L \leqslant p \leqslant N} \{qp\colon q \in \mathcal{Q}_{k-1}(N/p)\} \tag{27}$$

where the union is disjoint. The disjointness follows from the fact that $p \geqslant L = N/\log N > \log N > q$ and thus distinct choices of p and q yield distinct products. To see the containment we note that since $k \geqslant 3$, q must have at least two prime factors, so that the largest prime factor of q, say p', is at most $N/2p \leqslant \log N/2$; thus

$$\log p \geqslant \log N - \log_2 N \geqslant \alpha\left(\frac{\log N}{2}\right) \geqslant \alpha p', \tag{28}$$

so that qp is one of the integers in $\mathcal{Q}_k(N)$.

The containment (27) leads immediately to the inequality

$$Q_k(N) \geqslant \sum_{L \leqslant p \leqslant L'} Q_{k-1}(N/p), \tag{29}$$

where L' can have any value satisfying $L' \geqslant L$. We define L' by

$$L' = N/e((\log_2 N)^{1/\log_4 N}). \tag{30}$$

With this choice we can show that

$$\log_k N/p \geqslant \log_k N/L' \geqslant (\log_{k+1} N)(1 - (\log_5 N)/\log_4 N). \tag{31}$$

For $k > 3$, (31) yields

$$\log_k N/p \geqslant k; \tag{32}$$

while for $k = 3$ (31) yields

$$\log_3 N/p \geqslant 2, \tag{33}$$

where we have used $\log_{k+1} N \geqslant k + 1$.

From (32) and (33) we see that the hypothesis of the inductively assumed theorem is satisfied for estimating the summands $Q_{k-1}(N/p)$ in (29).

We define $\widetilde{Q}_k(x)$ by

$$\widetilde{Q}_k(x) = \frac{x}{\log x} \prod_3^{k+1} \log_i x; \tag{34}$$

thus in the range of summation in (29) by the inductive hypothesis $\widetilde{Q}_{k-1}(N/p) \leqslant Q_{k-1}(N/p)$.

From the lemma we get

$$\begin{aligned} Q_k(N) \geqslant \frac{\vartheta(x) - x}{\log x} \widetilde{Q}_{k-1}(N/x)\Big|_L^{L'} + \int_L^{L'} \frac{\widetilde{Q}_k(N/x)}{\log x}\, dx \\ - \int_L^{L'} (\vartheta(x) - x) \frac{d}{dx} \frac{\widetilde{Q}_k(N/x)}{\log x}\, dx. \end{aligned} \tag{35}$$

We first obtain lower estimates for the first and last terms in the RHS of (35) and estimate the middle term, which is the main term, last. By (12), the estimate $|\vartheta(x) - x| < x/2 \log x$ is valid in the range under consideration. Since $x/2 \log x$ is increasing in x while $\widetilde{Q}_{k-1}(N/x)$ is decreasing, we see that

$$\left| \frac{\vartheta(x) - x}{\log x} \widetilde{Q}_{k-1}(N/x)\Big|_L^{L'} \right| \leqslant 2 \frac{N}{2 \log^2 N} \cdot \widetilde{Q}_{k-1}(\log N). \tag{36}$$

A straightforward calculation yields

$$\left| \frac{d}{dx}\left(\frac{\widetilde{Q}_{k-1}(N/x)}{\log x} \right) \right| \leqslant \frac{\widetilde{Q}_{k-1}(N/x)}{x \log x}. \tag{37}$$

Thus the absolute value of the last term of the RHS of (35) is bounded above by

$$\int_L^{L'} \frac{\widetilde{Q}_{k-1}(N/x)}{\log^2 x}\, dx \leqslant \frac{1}{\log^2 L} \int_L^{L'} \widetilde{Q}_{k-1}(N/x)\, dx. \tag{38}$$

Similarly for the main term

$$\int_L^{L'} \frac{\widetilde{Q}_{k-1}(N/x)}{\log x}\, dx \geqslant \frac{1}{\log L'} \int_L^{L'} \widetilde{Q}_{k-1}(N/x)\, dx. \tag{39}$$

Putting together (35), (36), (38), and (39), we obtain

(40)
$$Q_k(N) \geqslant \left(\frac{1}{\log L'} - \frac{1}{\log^2 L}\right)\int_L^{L'} \widetilde{Q}_{k-1}(N/x)\, dx - \frac{N}{\log^2 N}\, \widetilde{Q}_{k-1}(\log N).$$

We can evaluate the integral in (40) by parts with $u = \Pi_3^k \log_j(N/x)$ and $v = -\log_2(N/x)$ to obtain

(41)
$$\int_L^{L'} \widetilde{Q}_{k-1}(N/x)\, dx = -N \log_2 N/x \prod_3^k \log_j N/x \Big|_L^{L'} + \int_L^{L'} \widetilde{Q}_{k-1}(N/x)\left(\sum_{i=3}^{k}\left(\prod_{i=3}^{i} \log_j N/x\right)^{-1}\right) dx.$$

Since

$$\sum_{i=3}^{k}\left(\prod_{i=3}^{i} \log_j N/x\right)^{-1} \geqslant \frac{1}{\log_3 N/x},$$

(41) leads to

(42)
$$\int_L^{L'} \widetilde{Q}_{k-1}(N/x)\, dx \geqslant -N \prod_2^k \log_j N/x \Big|_L^{L'} + \int_L^{L'} \widetilde{Q}_{k-1}(N/x)/\log_3 N/x\, dx.$$

The last integral can be approximated by substituting for $\widetilde{Q}_{k-1}(N/x)$ and simplifying to get

(43)
$$\begin{aligned}\int_L^{L'} \widetilde{Q}_{k-1}(N/x)/\log_3 N/x\, dx &= \int_L^{L'} \frac{N}{x \log N/x} \prod_4^k \log_j N/x\, dx \\ &\geqslant N \prod_4^k \log_j N/L' \cdot \int_L^{L'} \frac{1}{x \log N/x}\, dx \\ &= N \prod_4^k \log_j N/L'(-\log_2 N/x|_L^{L'}) \\ &= N \cdot \prod_4^k \log_j N/L' \left(\log_3 N - \frac{\log_3 N}{\log_4 N}\right).\end{aligned}$$

Substituting this for the last term in (42) while evaluating the first and combining terms, we get

$$\int_L^{L'} \widetilde{Q}_{k-1}(N/x)\,dx = N\left\{\prod_3^{k+1} \log_j N + \left(\prod_4^{k} \log_j N/L'\right)\log_3 N \log_4 N\left(\frac{\log_5 N - 1}{\log_4^2 N}\right)\right\}. \tag{44}$$

Since $1/\log L - 1/\log^2 L' > 1/\log N$, we get from (40), and (44) that

$$\begin{aligned} Q_k(N) \geqslant &\frac{N}{\log N} \prod_3^{k+1} \log_j N \\ &+ \frac{N}{\log N} \log_3 N \log_4 N \prod_4^{k} \log_j N/L'\left(\frac{\log_5 N - 1}{\log_4^2 N}\right) \\ &- \frac{N}{\log N} \cdot \frac{1}{\log_2 N} \prod_4^{k+1} \log_j N. \end{aligned} \tag{45}$$

Since

$$\log_4 N/L' = \log_5 N + \log\left(1 - \frac{\log_5 N}{\log_4 N}\right) \geqslant \log_5 N\left(1 - \frac{2}{\log_4 N}\right),$$

we see that the sum of the last two terms is positive. The desired lower bound follows.

The Upper Bound for $Q_k(N)$. We may suppose $N \geqslant e_{k-2}(11)$, for otherwise $Q_k(N) = 0$.

We begin by establishing the following inequality:

$$\begin{aligned} Q_k(N) \leqslant &\sum_{M \leqslant p \leqslant L} Q_{k-1}(\log p \log_2^2 p) + \sum_{L < p \leqslant L'} Q_{k-1}(N/p) \\ &+ \sum_{L' < p \leqslant N/N_0} Q_{k-1}(N/p) = \Sigma_1 + \Sigma_2 + \Sigma_3, \end{aligned} \tag{46}$$

where $M = e_{k-2}(11)$, a lower bound for the largest prime factor of elements of $\mathcal{Q}_{k-1}$, $L = N/(\log N \cdot \log_2^2 N)$ and $L' = \min\{N/\log_3 N, N/N_0\}$, where N_0 is the smallest element in $\mathcal{Q}_{k-1}$. To see that (46) holds, consider $n \in \mathcal{Q}_k(N)$, factor $n = pq$ where p is the largest prime factor, then n is counted by the appropriate sum depending on the range into which p falls. We see that in the first sum since $q = p_1 p_2 \cdots p_{k-1}$ with $p_{k-1} \leqslant \log p/\alpha$ and $p_i \leqslant \log p_{i+1}/\alpha$, $1 \leqslant i < k-1$, $q \leqslant \log p \log_2 p \cdots \log_{k-1} p \leqslant \log p \log_2^2 p$. The last two sums follow from the fact that $pq = n \leqslant N$ and thus $q \leqslant N/p$.

For the remainder of the proof we suppose that $L' = N/\log_3 N$, for otherwise the last sum in (46) is zero and the range on the middle sum is shortened. In either case the inductive assumption applies to each $Q_{k-1}(N/p)$ of the middle sum.

To estimate Σ_1 we note that there are at most $\pi(L)$ summands in which each is at most $Q_{k-1}(\log L \log_2^2 L)$ using the estimate $\pi(x) \leqslant 2x/\log x$ and the inductive estimate for Q_{k-1} we obtain

$$\sum\nolimits_1 \leqslant \frac{2L}{\log L} \cdot \frac{\log L}{\log_2 L} (\log_k L + k - 1) \prod_3^{k-1} \log_j L$$

$$\leqslant \frac{2N}{\log N} \cdot \frac{1}{\log_2 N/\log N} (\log_k N + k - 1) \prod_3^{k-1} \log_j N \tag{47}$$

$$\leqslant \frac{3}{\log_2 N} \cdot \frac{N}{\log N} (\log_k N + k - 1) \prod_3^{k-1} \log_j N.$$

We next consider Σ_3. There are at most $\pi(N/22)$ summands each of size at most $Q_{k-1}(N/L') = Q_{k-1}(\log_3 N)$. Hence we conclude

$$\sum\nolimits_3 \leqslant \frac{2N}{22 \log N/22} \cdot \frac{\log_3 N}{\log_4 N} (\log_{k+3} N + k - 1) \prod_6^{k+2} \log_j N$$

$$\leqslant \frac{1}{10 \log_4^2 N} \frac{N}{\log N} (\log_{k+1} N + k - 1) \prod_3^{k} \log_j N. \tag{48}$$

We now turn our attention to Σ_2 which yields the main term. By use of the inductive hypothesis, the choice $L = N/\log N$, the estimate $\log_j(\log x \log_2^2 x) \leqslant (\log_{j+1} x)(1 + 2/\log_2 x)$, for $j \geqslant 3$, and the lemma we deduce

$$\sum\nolimits_2 = \sum_{L<p\leqslant L'} \frac{N}{p \log N/p} (\log_k N/p + k - 1) \prod_3^{k-1} \log_j N/p$$

$$\leqslant N(\log_{k+1} N + k - 1)\left(1 + \frac{2}{\log_2 N}\right)^k \prod_4^k \log_j N \sum_{L<p\leqslant L'} \frac{1}{p \log N/p}$$

$$\leqslant N(\log_{k+1} N + k - 1)\left(1 + \frac{2}{\log_2 N}\right)^k \prod_4^k \log_j N \tag{49}$$

$$\cdot \left\{ \int_L^{L'} \frac{dx}{x \log x \log N/x} + \int_L^{L'} (\vartheta(x) - x) \frac{d}{dx}\left(\frac{1}{x \log x \log N/x}\right) dx + \frac{\vartheta(x) - x}{x \log x \log N/x}\bigg|_L^{L'} \right\}.$$

The last terms in the braces have been evaluated earlier in formulae (18) and (22), where in those formulae slightly different values of L and L' were used. The $1/\log x$ can be taken outside the integral as $1/\log L$ and the rest integrated exactly to yield

$$\sum\nolimits_2 \leqslant N(\log_{k+1} N + k - 1)\left(1 + \frac{2}{\log_2 N}\right)^k \prod_4^k \log_j N$$
$$\cdot \left\{ \frac{1}{\log L} \log_2 N/x \Big|_L^{L'} + \frac{\log_3 N}{2 \log^2 N} + \frac{1}{2 \log^2 N} \right\}$$

(50)

$$\leqslant \frac{N}{\log N} (\log_{k+1} N + 2) \prod_3^k \log_3 N$$
$$\cdot \left\{ \left(1 + \frac{2}{\log_2 N}\right)^k \left(\left(1 + \frac{2 \log_2 N}{\log N}\right)\left(1 - \frac{\log_5 N}{\log_3 N}\right) + \frac{1}{2 \log N} + \frac{1}{2 \log N \log_3 N} \right) \right\}.$$

Recalling that $L' = N/\log_3 N$ or, equivalently $\log_3 N \geqslant N_0 \geqslant 22$, we deduce that $\log_5 N \geqslant 1$. Hence we see that the quantity in the braces is less than 1.

It follows from (50), (48) and (47) that

(51)
$$Q_k(N) \leqslant \frac{N}{\log N} (\log_{k+1} N + k - 1) \prod_3^k \log_0 N \left\{ 1 + \frac{1}{10 \log_4^2 N} + \frac{3}{\log_2 N} \right\}$$
$$\leqslant \frac{N}{\log N} (\log_{k+1} N + k) \prod_3^k \log_j N,$$

which is the desired upper bound.

The Number of Distinct Subsums of $\Sigma_1^N 1/i$; a Lower Bound. Let $\mathcal{Q}(N) = \bigcup_{k=1}^{\infty} \mathcal{Q}_k(N)$ and $Q(N) = \Sigma_1^{\infty} Q_k(N)$, where we have taken $\alpha = 3/2$ in defining $Q_k(N)$. Since for any N only finitely many $Q_k(N)$ are nonzero, there is no difficulty with the sum.

In order to relate the problem of distinct values of subsums of $\Sigma_1^N 1/i$ to the previous problem we first prove the following theorem.

THEOREM. *If $S(N)$ denotes the number of distinct values of $\Sigma_1^N \epsilon_k/k$ as the ϵ_k assume all the 2^N possible combinations with $\epsilon_k = 0, 1$, then $S(N) \geqslant 2^{Q(N)}$.*

Before proving the theorem we point out some immediate consequences of this theorem in combination with the previous theorem's lower bounds for $Q_k(N)$.

COROLLARY 1. *For $N \geqslant 2$,*

$$S(N) \geqslant 2^{\pi(N)} \geqslant e\left(\frac{N \log 2}{\log N}\left(1 + \frac{1}{2} \log N\right)\right).$$

COROLLARY 2. *For $\log_3 N \geqslant 2$,*

$$S(N) \geqslant e\left(\frac{N \log 2}{\log N}\left(\log_3 N + \frac{12}{11} + \frac{1}{2 \log N}\right)\right).$$

COROLLARY 3. *For* $k \geqslant 3$ *and* $\log_{k+1} N \geqslant k + 1$,

$$S(N) \geqslant e\left(\frac{N \log 2}{\log N} \prod_{3}^{k+1} \log_j N\right).$$

It may be noted that these corollaries improve the results on lower bounds for $S(N)$ obtained in [2] in two ways. The first is that the constant $1/e$ in the bound in [2] is replaced by the larger $\log 2(\log_3 N + 12/11 + 1/2 \log N)$ in Corollary 2 and by $\log 2$ in Corollary 3. The second is the validity of the formula for a given k is extended to much smaller values of N.

Combining Corollaries 2 and 3 above with Theorem 3 of [2] we obtain

COROLLARY 4. *For* $\log_{2r} N \geqslant 1$ *and* $r \geqslant 2$, *choose* t *such that* $e_t(1) \geqslant 2r - t - 1$. *Let* $k = 2r - t - 1$. *Then* $k \geqslant r$ *(equality only for* $r = 2, 3$*) and*

$$e\left(\frac{N \log 2}{\log N} \prod_{3}^{k+1} \log_j N\right) \leqslant S(N) \leqslant e\left(\frac{N \log_r N}{\log N} \prod_{3}^{r} \log_j N\right).$$

Proof of Corollary 4. From the definition of k we see that if $\log_{2r} N \geqslant 1$ then $\log_{k+1} N \geqslant e_t(1) \geqslant k$; hence Corollary 3 gives the lower bound for $r \geqslant 3$. For $r = k = 2$ it is easy to see that $\log_4 N \geqslant 1$ implies $\log_3 N \geqslant 2$, hence Corollary 2 gives the lower bound. The upper bound is from Theorem 3 of [2]. The comment about equality of k and r is a trivial calculation. In fact, for $r = 4$, $k = 5$, while for $r = 5$, $k = 7$. The corollary is proved.

Proof of the Theorem. The idea of the proof is simple. We show that for each sequence $n_1, n_2, n_3, \cdots, n_k$ of distinct elements of $Q(N)$ we get a distinct value for $\Sigma 1/n_i$. Since $n_i \leqslant N$ and there are $2^{Q(N)}$ such sequences, the lower bound follows, if we can show the values are all distinct. Thus the theorem will be established if we prove the following lemma.

LEMMA. *Let* $n_1, n_2, \cdots, n_k$ *and* $m_1, m_2, \cdots, m_l$ *be two sequences of elements of* $Q(N)$; *the elements in each of these sequences being distinct from other elements of that sequence. Then* $\Sigma 1/n_i = \Sigma 1/m_i$ *if and only if* $k = l$ *and, after possibly renumbering,* $n_i = m_i$, $i = 1, 2, \cdots, k$.

Proof of the Lemma. We prove the "only if". The "if" half is trivial.

Let P be the largest prime factor of the product of the n_i and m_i. Let $n_1, n_2, \cdots, n_{k'}$ and $m_1, m_2, \cdots, m_{l'}$ be all those n_i and m_i in increasing order which have P as a factor. The proof is by induction on the size of P.

If $P = 2, n_i, m_i \in \{1, 2\}$ and clearly the distinctness of different sums is true. Similarly for $P = 3$ when $n_i, m_i \in \{1, 2, 3\}$.

We now suppose that $P \geqslant 5$ and that for sequences which have only prime factors less than P, distinct sequences yield distinct values.

Define a/b, a reduced fraction, by

$$\frac{a}{b} = \sum_1^{k'} \frac{1}{n_i} - \sum_1^{l'} \frac{1}{m_i}. \tag{52}$$

We may assume $a/b \geqslant 0$, since otherwise we may interchange the m_i and n_i and proceed.

Let $n_i = Pn_i'$ and $m_i = Pm_i'$; thus

$$\frac{a}{b} = \frac{1}{P}\left(\sum_1^{k'} \frac{1}{n_i'} - \sum_1^{l'} \frac{1}{m_i'}\right). \tag{53}$$

We next show that

$$k' = l' \quad \text{and} \quad n_i' = m_i', \qquad i = 1, 2, \cdots, k'. \tag{54}$$

If $a = 0$ then the claim follows by induction since the n_i' and m_i' have largest prime factor less than P.

We thus consider the case $a \neq 0$ and derive a contradiction.

Since the n_i and m_i are in $\mathcal{Q}(N)$ and P was the largest prime factor if we choose Q to be the largest prime such that $e^{3Q/2} < P$, then we know from the definition of $\mathcal{Q}(N)$ that no prime factor of any n_i' or m_i' exceeds Q. Since all the n_i and m_i are squarefree, we see that $d = \Pi_{P<Q} P = e^{\vartheta(Q)}$ is a common multiple for the n_i' and m_i'. Thus

$$\sum_1^{k'} \frac{1}{n_i'} - \sum_1^{l'} \frac{1}{m_i'} = \frac{c}{d} \tag{55}$$

for some positive integer c. Since the largest prime factor of the n_i' and m_i' is at most Q and the n_i' and m_i' are in $\mathcal{Q}(N)$, we see that $Q \log Q \log_2 Q \cdots \log_r Q \geqslant n_i, m_i$ where r is chosen so that $e^2 > \log_r Q \geqslant 2$. Thus $c/d \leqslant \Sigma_1^{Q^2} 1/i < 2 \log Q + 1$. Hence $c < 3d \log Q$. It follows that

$$c < 3d \log Q < 3e^{\vartheta(Q)} \log Q < e^{3\vartheta(Q)/2} < P. \tag{56}$$

(Note: For $Q = 2, 3$ a different argument is needed to show that $c < P$ since $3 \log Q > e^{\vartheta(Q)/2}$. A trivial calculation suffices.)

Since $0 < c < P$ it follows that $P \nmid c$. Since $a/b = 1/P \cdot c/d$ and $(a, b) = 1$, we see that $P \nmid a$ and $P | b$.

But by hypothesis $\Sigma 1/n_i = \Sigma 1/m_i$, thus

$$\frac{a}{b} = \sum_1^{k'} \frac{1}{n_i} - \sum_1^{l'} \frac{1}{m_i} = \sum_{i>l'} \frac{1}{m_i} - \sum_{i>k'} \frac{1}{n_i} = \frac{r}{s}$$

where we may take $s = e^{\vartheta(P-1)}$, since all the n_i, $i > k'$, and all the m_i, $i > l'$, have prime factors less than P. We deduce that $P \nmid s$; but $a/b = r/s$ and $(a, b) = 1$ and $P | b$, thus $P | s$, a contradiction. Thus $a/b = 0$, and as noted before the equalities of (54) follow. But (54) implies $n_i = m_i$ for $i = 1, 2, \cdots, k' = l'$. Thus

$$\sum_{i=k'+1}^{k} \frac{1}{n_i} = \sum_{i=k'+1}^{l} \frac{1}{m_i}$$

and all prime factors are less than P. By induction $k = l$ and $n_i = m_i$ for $i = k' + 1, k' + 2, \cdots, k$.

The lemma is established.

Conclusion of the Proof of the Theorem. From the lemma we see that every distinct subset of $Q(N)$ yields a distinct value for $\Sigma_1^N \epsilon_k/k$ by setting $\epsilon_k = 1$ for members of the subset and $\epsilon_k = 0$ otherwise. Thus $S(N) \geqslant 2^{Q(N)}$, as claimed.

The theorem is established.

Mathematics Department
University of Wisconsin
Madison, Wisconsin 53706

Bell Laboratories
Murray Hill, New Jersey 07974

1. M. N. BLEICHER & P. ERDÖS, "Denominators of Egyptian fractions," *J. Number Theory* (To appear.)

2. M. N. BLEICHER & P. ERDÖS, "Denominators of Egyptian fractions. II" (To appear.)

3. M. N. BLEICHER & P. ERDÖS, "The number of distinct subsums of 1 + 1/2 + 1/3 + 1/4 + · · · + 1/*N*, *Notices Amer. Math. Soc.*, v. 20, 1973. Abstract #706-10-3.

4. J. B. ROSSER & L. SCHOENFELD, "Approximate formulas for some functions of prime numbers," *Illinois J. Math.*, v. 6, 1962, pp. 64–94. MR 25 #1139.