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

$$S(N) \geq e\left(\frac{N \log 2}{\log N} \prod_{i=3}^{k+1} \log_i N\right)$$

whenever $\log_{k+1} N \geq k + 1$, for $k \geq 3$. Slight modifications are needed for $k = 1, 2$.

We begin by discussing the number $Q_k(N)$ of integers $n \leq N$, $n = p_1 p_2 \cdots p_k$, where $p_i > e^{\alpha p_{i-1}}$, $i = 2, \cdots, k$. We prove that

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

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

$$e_0(x) = x, \quad e_{i+1}(x) = e_i(x),$$

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

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

$$S(N) \geq e\left(\frac{N \log 2}{\log N} \prod_{i=3}^{k+1} \log_i N\right)$$

whenever $\log_{k+1} N \geq k + 1$, for $k \geq 3$. 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 $n \leq N$, $n = p_1 p_2 \cdots p_k$ where $p_i > e^{\alpha p_{i-1}}$, $i = 2, \cdots, k$. We first prove that

Received July 26, 1974.


Key words and phrases. Harmonic series, squarefree integers, prime numbers, distinct subsums.
\[ \frac{N}{\log N} \prod_{i=3}^{k+1} \log N \leq Q_k(N) \leq \left(1 + \frac{k}{\log_{k+1} N}\right) \frac{N}{\log N} \prod_{i=3}^{k+1} \log_i N. \]

This bound is valid for \( \log_{k+1} N \geq k + 1 \) and for \( 1 < \alpha < 2(1 - e_2(4)/e_3(4)) \). The bounds on \( N \) and \( \alpha \) 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

\[ e_0(x) = x, \quad e_{i+1}(x) = e_i(x), \]
\[ \log_0 x = x, \quad \log_{i+1} x = \log(\log_i x), \]

where \( \log x \) denotes the logarithm to the base \( e \).

In fact we prove the following slightly stronger version.

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

- For \( k = 1 \),
  \[ \frac{N}{\log N} \left(1 + \frac{1}{2 \log N}\right) \leq Q_1(N) = \pi(N) \leq \frac{N}{\log N} \left(1 + \frac{3}{2 \log N}\right), \]
  where the lower bound holds for \( N \geq 59 \) and the upper bound for \( N \geq 2 \); \( Q_1(N) = 0 \) for \( N < 2 \).

- For \( k = 2 \),
  \[ \frac{N}{\log N} \left(\log_3 N + \frac{1}{11}\right) \leq Q_2(N) \leq \frac{N}{\log N} \left(\log_3 N + 2\right), \]
  where the lower bound holds for \( \log_3 N \geq 2 \) and the upper bound for \( N \geq e_3(-2) = 3.1 \cdots \) (i.e., \( \log_3 N \geq -2 \)); \( Q_2(N) = 0 \) for \( N < 22 \).

- For \( k \geq 3 \),
  \[ \frac{N}{\log N} \prod_{i=3}^{k+1} \log_i N \leq Q_k(N) \leq \frac{N(\log_{k+1} N + k)}{\log N} \prod_{i=3}^{k+1} \log_i N, \]
  where the lower bound holds for \( \log_{k+1} N \geq k + 1 \) and the upper bound holds for \( N \geq e_{k+1}(-2) \); \( Q_k(N) = 0 \) for \( N \leq e_{k+1}(-13 \cdots) = e_{k-2}(11) \).

**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 \( Q_2(N) \) be those integers counted by \( Q_2(N) \); namely

\[ Q_2(N) = \{pq: p, q \ prime, e^{op} < q, pq \leq N\}. \]

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

(1) \[ Q_2(N) = \sum_{2 \leq p < L} \left(\pi(N/p) - \pi(e^{op})\right), \]
where \( p \) runs through the primes in the indicated interval. We see from the conditions on \( \alpha \) that

\[
(2) \quad L \leq \log N.
\]

We thus deduce that

\[
(3) \quad Q_4(N) \leq \sum_{2 < p < \log N} \frac{N}{p \log N/P} \left( 1 + \frac{3}{2 \log N/P} \right).
\]

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

\[
(4) \quad Q_2(N) \leq \frac{N}{\log(N/\log N)} \left( 1 + \frac{3}{2 \log(N/\log N)} \right) \sum_{2} \frac{1}{p}.
\]

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

\[
(5) \quad Q_2(N) \leq \frac{N}{\log N} \left( 1 + \frac{2 \log_2 N}{\log N} \right) \left( \log_3 N + B + \frac{1}{\log_2 N} \right),
\]

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

\[
(6) \quad Q_2(N) \leq N(\log_3 N + 2)/\log N.
\]

If \( 22 \leq N \leq e^4 < 55 \), then \( Q_2(N) \leq Q_2(54) = 5 \) together with \( \log_3 N \geq 0 \) 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

\[
(7) \quad Q_2(N) = \sum_{1 < p < N} \sum_{1 < q < M} 1,
\]

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

\[
(8) \quad \alpha N = L \log L,
\]

so that \( N/\log N < L < eN/\log N \), then

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

Let \( \Sigma_1 \) denote the first double sum and \( \Sigma_2 \) the second. Since \( \Sigma_1 \geq 0 \) we can obtain a lower bound for \( Q_2(N) \) by obtaining a lower bound for \( \Sigma_2 \).

The Bounds for \( \Sigma_2 \). From the definition of \( \Sigma_2 \) in (9) we obtain

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

where \( L < L' = N/p_i \) is the \( i \)th prime with \( i \geq 7 \) to be determined later. We note that
We shall frequently need to estimate sums of the above type where the index of the summation ranges 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 $f(x) > 0$ and $f'(x)$ exists and is continuous and $0 < a < b$

$$
\sum_{a < p \leq b} f(p) = \frac{f(x)(\vartheta(x) - x)}{\log(x)} \bigg|_a^b + \int_a^b \frac{f(x)}{\log x} \, dx
$$

$$
- \int_a^b (\vartheta(x) - x) \frac{d}{dx} \left( \frac{f(x)}{\log x} \right) \, dx.
$$

We recall from [4] the estimates

(12) $|\vartheta(x) - x| \leq x/(2 \log x)$ for $x \geq 563$

and

(13) $\vartheta(x) - x \leq x/(2 \log x)$ for $x > 1$

and the estimates

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

(15) $\frac{x}{\log x} < \pi(x)$ for $x \geq 17$,

and

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

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

$$
N \sum_{L < p \leq L'} \frac{N}{p \log N/p} = N \left\{ \frac{\vartheta(x) - x}{x \log x \log N/x} \bigg|_{L}^{L'} + \int_L^{L'} \frac{dx}{x \log x \log N/x}
\right.
$$

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

We next show that
THE NUMBER OF DISTINCT SUBSUMS OF $\sum_{1}^{N} 1/i$

(18) \[
\left| \int_{L}^{L'} \left( \Theta(x) - x \right) \frac{d}{dx} \left( \frac{1}{x \log x \log N/x} \right) \, dx \right| \leq \frac{\log_3 N}{2 \log^2 N}.
\]

To do this we note that
\[
\left| \frac{d}{dx} \left( \frac{1}{x \log x \log N/x} \right) \right| \leq \frac{1}{x^2 \log x \log N/x}
\]
and that the estimate of (12), $|\Theta(x) - x| < x/2 \log x$ are both valid for the range $N/\log N \leq x \leq N/2$ when $N \geq e^{8.5}$. Thus

(19) \[
\left| \int_{L}^{L'} (\Theta(x) - x) \frac{d}{dx} \left( \frac{1}{x \log x \log N/x} \right) \, dx \right| \leq \int_{L}^{L'} \frac{dx}{2x \log^2 x \log N/x}.
\]

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

(20) \[
\int_{L}^{L'} \frac{dx}{2x \log^2 x \log N/x} \leq \frac{1}{2 \log^2 L} \left(-\log_2 N/x\right)_{L}^{L'},
\]
which yields (18).

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

(21) \[
\frac{\log_3 N}{\log N} \left( 1 + \frac{\log_2 p_l}{\log_3 N} + \frac{\log_2 p_l}{\log N} \right) \leq \int_{L}^{L'} \frac{dx}{x \log x \log N/x}.
\]

We next note that

(22) \[
\left| \left( \frac{\Theta(x) - x}{x \log x \log N/x} \right)_{L}^{L'} \right| \leq \frac{1}{2 \log^2 L} \log N/L + \frac{1}{2 \log^2 L'} \log N/L' \leq \frac{1}{2 \log^2 N}.
\]

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

(23) \[
\sum_{N/p_l \leq p \leq N/2} \pi \left( \frac{N}{p} \right) = \sum_{2 \leq p \leq p_l} \pi \left( \frac{N}{p} \right) - \pi \left( \frac{N}{17} \right) \geq \frac{N}{\log N/p_l} \left( \sum_{2 \leq p \leq p_l} \frac{1}{p} - \frac{1}{p_l} \left( 1 + \frac{3}{2 \log N/p_l} \right) \right).
\]

If $l/p_l < B$, then using $N \geq e_3(2) > e^{1.600}$ and $p_l < \log N$,

(24) \[
\sum_{N/p_l \leq p \leq N/2} \pi \left( \frac{N}{p} \right) \geq \frac{N}{\log N} \left( \log_2 p_l + B - \frac{1}{2 \log^2 p_l} - \frac{l}{p_l} + \frac{\log p_l}{\log N} \right).
\]

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

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use
\[ \sum_{2} \geq \frac{N \log_3 N}{\log N} \left( 1 - \frac{\log p_1}{\log_3 N} + \frac{\log p_1}{\log_3 N} - \frac{1}{2 \log N \log_3 N} \right) \]

(25)

\[- \frac{1}{2 \log N} + \frac{\log p_1}{\log_3 N} + \frac{B - \log p_1}{\log_3 N} \]

\[- \frac{1}{2 \log^2 p_1 \log_3 N} + \frac{\log p_1}{\log_3 N \log N} \].

Taking \( p_1 = 1597, \ l = 251 \) so that all the previous conditions are satisfied and using \( B = .261 \cdots, \ l/p_1 = .157 \cdots, \ 1/2 \log^2 p_1 < .0005 \) and \( \log_3 N \geq 2 \), we deduce

(26)

\[ \sum_{2} \geq \frac{N \log_3 N}{\log N} \left( 1 + \frac{1}{11 \log_3 N} \right). \]

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

Case 3. \( k \geq 3 \). We now proceed by induction on \( k \). Suppose \( k > 2 \) and that for \( 2 \leq k' < k \) 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 \( Q_k(N) \) denote the set of integers counted by \( Q_k(N) \). As before let \( L = N/\log N \). We claim that

(27)

\[ Q_k(N) \supset \bigcup_{L \leq p < N} \{ qp: q \in Q_{k-1}(N/p) \} \]

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

(28)

\[ \log p \geq \log N - \log_2 N \geq \alpha \left( \frac{\log N}{2} \right) \geq \alpha p', \]

so that \( qp \) is one of the integers in \( Q_k(N) \).

The containment (27) leads immediately to the inequality

(29)

\[ Q_k(N) \supset \sum_{L \leq p < L'} Q_{k-1}(N/p), \]

where \( L' \) can have any value satisfying \( L' \geq L \). We define \( L' \) by

(30)

\[ L' = N/e((\log_2 N)^{1/\log_4 N}). \]

With this choice we can show that

(31)

\[ \log_k N/p \geq \log_k N/L' \geq (\log_{k+1} N)(1 - (\log_5 N)/\log_4 N). \]
For \( k > 3 \), (31) yields

\[(32) \quad \log_k N/p \geq k; \]

while for \( k = 3 \) (31) yields

\[(33) \quad \log_3 N/p \geq 2, \]

where we have used \( \log_{k+1} N \geq k + 1 \).

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

\[(34) \quad \widetilde{Q}_k(x) = \frac{x}{\log x} \prod_{3}^{k+1} \log x; \]

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

From the lemma we get

\[(35) \quad \int_L^{L'} \widetilde{Q}_k(N/x) dx \leq 2 \frac{N}{\log^2 N} \cdot \widetilde{Q}_{k-1}(\log N). \]

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 \( |\theta(x) - x| < x/2 \log x \) 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

\[(36) \quad \left| \frac{\theta(x) - x}{\log x} \widetilde{Q}_{k-1}(N/x) \right|_L^{L'} \leq 2 \frac{N}{\log^2 N} \cdot \widetilde{Q}_{k-1}(\log N). \]

A straightforward calculation yields

\[(37) \quad \left| \frac{d}{dx} \left( \frac{\widetilde{Q}_{k-1}(N/x)}{\log x} \right) \right| \leq \frac{\widetilde{Q}_{k-1}(N/x)}{x \log x}. \]

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

\[(38) \quad \int_L^{L'} \frac{\widetilde{Q}_{k-1}(N/x)}{\log^2 x} dx \leq \frac{1}{\log^2 L} \int_L^{L'} \widetilde{Q}_{k-1}(N/x) dx. \]

Similarly for the main term

\[(39) \quad \int_L^{L'} \frac{\widetilde{Q}_{k-1}(N/x)}{\log x} dx \geq \frac{1}{\log L} \int_L^{L'} \widetilde{Q}_{k-1}(N/x) dx. \]

Putting together (35), (36), (38), and (39), we obtain
We can evaluate the integral in (40) by parts with $u = \frac{1}{\log L} - \frac{1}{\log^2 L}$ and $v = -\frac{N}{\log N}$ to obtain

$$Q_k(N) \geq \left( \frac{1}{\log L} - \frac{1}{\log^2 L} \right) \int L^{' \prime} \tilde{Q}_{k-1}(N/x) \, dx - \frac{N}{\log N} \tilde{Q}_{k-1}(\log N).$$

(40)

Since

$$\sum_{i=3}^{k} \left( \frac{1}{\log N/x} \right)^{-1} \geq \frac{1}{\log N/x},$$

(41) leads to

$$\int L^{' \prime} \tilde{Q}_{k-1}(N/x) \, dx \geq -N \prod_{3}^{k} \log_j N/x \left| \frac{L^{' \prime}}{L} \right| + \int L^{' \prime} \tilde{Q}_{k-1}(N/x) \left( \sum_{i=3}^{k} \left( \prod_{i=3}^{j} \log_j N/x \right)^{-1} \right) \, dx.$$

(42)

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

$$\int L^{' \prime} \tilde{Q}_{k-1}(N/x) \log N/x \, dx = \int L^{' \prime} \frac{N}{x \log N/x} \prod_{4}^{k} \log_j N/x \, dx \geq N \prod_{4}^{k} \log_j N/L^{' \prime} \cdot \int L^{' \prime} \frac{1}{x \log N/x} \, dx$$

(43)

$$= N \prod_{4}^{k} \log_j N/L^{' \prime} (-\log_2 N/x L^{' \prime})$$

$$= N \cdot \prod_{4}^{k} \log_j N/L^{' \prime} \left( \log_3 N - \frac{\log_3 N}{\log_4 N} \right).$$

Substituting this for the last term in (42) while evaluating the first and combining terms, we get
\[ \int_{L'}^{L} \tilde{Q}_{k-1}(N/x) \, dx \]

(44)

\[ = N \left( \prod_{j=3}^{k+1} \log_j N + \left( \prod_{j=4}^{k} \log_j N/L' \right) \log_3 N \log_4 N \left( \frac{\log_5 N - 1}{\log_4^2 N} \right) \right). \]

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

\[ Q_k(N) \geq \frac{N}{\log N} \prod_{j=3}^{k+1} \log_j N \]

(45)

\[ + \frac{N}{\log N} \log_3 N \log_4 N \prod_{j=4}^{k} \log_j N/L' \left( \frac{\log_5 N - 1}{\log_4^2 N} \right) \]

\[ - \frac{N}{\log N} \cdot \frac{1}{\log_2 N} \prod_{j=4}^{k+1} \log_j N. \]

Since

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

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** \( N \geq e_{k-2}(11) \), for otherwise \( Q_k(N) = 0 \).

We begin by establishing the following inequality:

\[ Q_k(N) \leq \sum_{M \leq p < L} Q_{k-1}(\log_2 p \log_2^2 p) + \sum_{L < p < L'} Q_{k-1}(N/p) \]

(46)

\[ + \sum_{L' < p \leq N/N_0} Q_{k-1}(N/p) = \sum_1 + \sum_2 + \sum_3, \]

where \( M = e_{k-2}(11) \), a lower bound for the largest prime factor of elements of \( Q_{k-1} \), \( L = N/(\log N \cdot \log_2^2 N) \) and \( L' = \min \{ N/\log_3^2 N, N/N_0 \} \), where \( N_0 \) is the smallest element in \( Q_{k-1} \). To see that (46) holds, consider \( n \in Q_k(N) \), 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 \( p_{k-1} \leq \log p/\alpha \) and \( p_i \leq \log p_i+1/\alpha, \ 1 \leq i < k-1, \)

\( q \leq \log p \log_2 p' \cdots \log_k-1 p' \leq \log p \log_2^2 p \). The last two sums follow from the fact that \( pq = n \leq N \) and thus \( q \leq N/p \).

For the remainder of the proof we suppose that \( L' = N/\log_3^2 N \), 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 \( \sum_1 \) 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 $\pi(x) \leq 2x/\log x$ and the inductive estimate for $Q_{k-1}$ we obtain

$$
\sum_{i=1}^{k-1} \leq \frac{2L}{\log L} \cdot \frac{\log L}{\log_2 L} (\log_k L + 1) \prod_{j=3}^{k-1} \log_j L
$$

(47)

$$
\leq \frac{2N}{\log N} \cdot \frac{1}{\log_2 N} (\log_k N + 1) \prod_{j=3}^{k-1} \log_j N
$$

$$
\leq \frac{3}{\log_2 N} \cdot \frac{N}{\log N} (\log_k N + 1) \prod_{j=3}^{k-1} \log_j N.
$$

We next consider $\Sigma_3$. 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

$$
\sum_{i=1}^{k-1} \leq \frac{2N}{22 \log N/22} \cdot \frac{\log_3 N}{\log_4 N} (\log_{k+3} N + 1) \prod_{j=6}^{k-1} \log_j N
$$

(48)

$$
\leq \frac{1}{10 \log_4 N} \cdot \frac{N}{\log N} (\log_{k+1} N + 1) \prod_{j=3}^{k} \log_j N.
$$

We now turn our attention to $\Sigma_2$ which yields the main term. By use of the inductive hypothesis, the choice $L = N/\log N$, the estimate $\log_j(\log x \log_2^2 x) \leq (\log_{j+1} x)(1 + 2/\log_2 x)$, for $j \geq 3$, and the lemma we deduce

$$
\sum_{i=1}^{k-1} \leq \frac{N}{\log N} \cdot \log_k N/p + \frac{1}{\log L} \cdot \log_k L
$$

$$
\leq N(\log_{k+1} N + 1) \left(1 + \frac{2}{\log_2 N}\right) \prod_{j=4}^{k} \log_j N \sum_{L < p < L'} \frac{1}{\log N/p}
$$

(49)

$$
\leq N(\log_{k+1} N + 1) \left(1 + \frac{2}{\log_2 N}\right) \prod_{j=4}^{k} \log_j N
$$

$$
\cdot \left\{ \log L \int_{L}^{L'} \frac{dx}{x \log x \log \log N/x} + \int_{L}^{L'} \frac{d}{dx} \left(\log(x) - x\right) \frac{1}{x \log x \log \log N/x} \right\}
$$

$$
+ \frac{\log(x) - x}{x \log x \log \log N/x} \bigg|_{L'}.
$$

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
\[
\sum_{2} \leq N (\log_{k+1} N + k - 1) \left(1 + \frac{2}{\log_2 N}\right)^k \prod_{4}^{k} \log_j N
\]
\[
\cdot \left\{ \frac{1}{\log L} \log_{2} N^x \left| \frac{L'}{L} + \frac{\log_3 N}{2 \log^2 N} + \frac{1}{2 \log^2 N} \right\} \right.
\]
\[
\leq \frac{N}{\log N} (\log_{k+1} N + 2) \prod_{3}^{k} \log_3 N
\]
\[
\cdot \left\{ \left(1 + \frac{2}{\log_2 N}\right)^k \left(1 + \frac{2 \log_2 N}{\log N}\right) \left(1 - \frac{\log_5 N}{\log_3 N}\right)
\right.
\]
\[
+ \frac{1}{2 \log N} + \frac{1}{2 \log \log_3 N}\right\}.
\]

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

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

\[
Q_k(N) \leq \frac{N}{\log N} (\log_{k+1} N + k - 1) \prod_{3}^{k} \log_0 N \left\{ 1 + \frac{1}{10 \log_3 N} + \frac{3}{\log_2 N}\right\}
\]
\[
(51)
\]

which is the desired upper bound.

The Number of Distinct Subsums of \( \Sigma_1^N 1/i \); a Lower Bound. Let \( Q(N) = \bigcup_{k=1}^{N} Q_k(N) \) and \( Q(N) = \Sigma_1^N Q_k(N) \), 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 \( \epsilon_k \) assume all the \( 2^N \) possible combinations with \( \epsilon_k = 0, 1 \), then \( S(N) \geq 2^Q(N) \).

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 \( N \geq 2 \),

\[
S(N) \geq 2^N \geq e \left( \frac{N \log 2}{\log N} \left(1 + \frac{1}{2 \log N}\right) \right).
\]

**Corollary 2.** For \( \log_3 N \geq 2 \),
\[ S(N) \geq e \left( \frac{N \log 2}{\log N} \left( \log_3 N + \frac{12}{11} + \frac{1}{2 \log N} \right) \right). \]

**Corollary 3.** For \( k \geq 3 \) and \( \log_{k+1} N \geq k + 1 \),
\[ S(N) \geq e \left( \frac{N \log 2}{\log N} \prod_{j=3}^{k+1} \log_j N \right). \]

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 \( \log_{2r} N \geq 1 \) and \( r \geq 2 \), choose \( t \) such that \( e^t(1) \geq 2r - t - 1 \). Let \( k = 2r - t - 1 \). Then \( k \geq r \) (equality only for \( r = 2, 3 \)) and
\[ \frac{N \log 2 \prod_{j=3}^{k+1} \log_j N}{\log N} \leq S(N) \leq e \left( \frac{N \log_{2r} N}{\log N} \prod_{j=3}^{r} \log_j N \right). \]

**Proof of Corollary 4.** From the definition of \( k \) we see that if \( \log_{2r} N \geq 1 \) then \( \log_{k+1} N \geq e^t(1) \geq k \); hence Corollary 3 gives the lower bound for \( r \geq 3 \). For \( r = 2 \) it is easy to see that \( \log_3 N \geq 1 \) implies \( \log_3 N \geq 2 \), 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 \( n_1, n_2, n_3, \ldots, n_k \) of distinct elements of \( Q(N) \) we get a distinct value for \( \sum 1/n_i \). Since \( n_i \leq N \) 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 \( n_1, n_2, \ldots, n_k \) and \( m_1, m_2, \ldots, m_l \) be two sequences of elements of \( Q(N) \); the elements in each of these sequences being distinct from other elements of that sequence. Then \( \sum 1/n_i = \sum 1/m_i \) if and only if \( k = l \) and, after possibly renumbering, \( n_i = m_{i'} \), \( i = 1, 2, \ldots, k \).

**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 \( n_1, n_2, \ldots, n_k \) and \( m_1, m_2, \ldots, m_l \) 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 \( P \geq 5 \) and that for sequences which have only prime factors less than \( P \), distinct sequences yield distinct values.
Define \( a/b \), a reduced fraction, by

\[
\frac{a}{b} = \sum_{i=1}^{k'} \frac{1}{n_i} - \sum_{i=1}^{l'} \frac{1}{m_i}.
\]

We may assume \( a/b \geq 0 \), since otherwise we may interchange the \( m_i \) and \( n_i \) and proceed.

Let \( n_i = Pn'_i \) and \( m_i = Pm'_i \), thus

\[
\frac{a}{b} = \frac{1}{P} \left( \sum_{i=1}^{k'} \frac{1}{n'_i} - \sum_{i=1}^{l'} \frac{1}{m'_i} \right).
\]

We next show that

\[
k' = l' \quad \text{and} \quad n'_i = m'_i, \quad i = 1, 2, \ldots, k'.
\]

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 \( Q(N) \) 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 \( Q(N) \) 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 \( d = \Pi_{e^r < P} = e^{\delta(Q)} \) is a common multiple for the \( n'_i \) and \( m'_i \). Thus

\[
\sum_{i=1}^{k'} \frac{1}{n'_i} - \sum_{i=1}^{l'} \frac{1}{m'_i} = \frac{c}{d}
\]

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 \( Q(N) \), we see that \( Q \log Q \log_2 Q \cdots \log_r Q \geq n, m \) where \( r \) is chosen so that \( e^2 > \log_r Q \geq 2 \). Thus \( c/d \leq e^{Q/2}1/i < 2 \log Q + 1 \). Hence \( c < 3d \log Q \). It follows that

\[
c < 3d \log Q < 3e^{\delta(Q)} \log Q < e^{3\delta(Q)/2} < P.
\]

(Note: For \( Q = 2, 3 \) a different argument is needed to show that \( c < P \) since \( 3 \log Q > e^{\delta(Q)/2} \). 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 \( P \nmid b \).

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

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

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

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 \( \sum_{i=1}^{N} \epsilon_k/k \) by setting \( \epsilon_k = 1 \) for members of the subset and \( \epsilon_k = 0 \) otherwise. Thus \( S(N) \geq 2^{Q(N)} \), as claimed.

The theorem is established.

Mathematics Department
University of Wisconsin
Madison, Wisconsin 53706

Bell Laboratories
Murray Hill, New Jersey 07974


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