ON SOME QUESTIONS OF ERDŐS AND GRAHAM ABOUT EGYPTIAN FRACTIONS

Ernest S. Croot III

Abstract. In this paper we prove that for x sufficiently large, every integer m with

$$1 \leq m \leq \left[\sum_{1 \leq n \leq x} \frac{1}{n} - \frac{9}{2} \frac{(\log \log x)^2 (1 + o(1))}{\log x} \right]$$

can be written as $m = \sum_{1 \leq n \leq x} \epsilon_n / n$, where $\epsilon_i = 0$ or 1.

§1. *Introduction.* Define $N(x)$ to be the set of all positive integers m which can be expressed as

$$m = \frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_k},$$

where k is variable and the n_i's are integers with $1 \leq n_1 < n_2 < \cdots < n_k \leq x$. In [1] Erdős and Graham asked the following questions.

1. What is the smallest number not in $N(x)$?
2. How many numbers are in $N(x)$?

Recently, in [6], Yokota showed that $\{n \ : \ 1 \leq n \leq \log x - 5 \log \log x\} \subseteq N(x)$, thus giving the correct asymptotic to the first two of these questions. In this paper we prove the following results.

Main Theorem. *Define $n(x)$ to be the largest integer such that, whenever $1 \leq n \leq n(x)$, $n \in \mathbb{Z}$, there exist integers $1 \leq n_1 < n_2 < \cdots < n_k \leq x$, for some k, such that*

$$n = \frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_k}.$$

Then

$$\left[\sum_{1 \leq n \leq x} \frac{1}{n} - \frac{9}{2} \frac{(\log \log x)^2 (1 + o(1))}{\log x} \right] \leq n(x) \leq \left[\sum_{1 \leq n \leq x} \frac{1}{n} - \frac{1}{2} \frac{(\log \log x)^2 (1 + o(1))}{\log x} \right].$$

Corollary. *Let $\gamma = \lim_{x \to \infty} \left\{ \sum_{1 \leq n \leq x} 1/n - \log x \right\}$ be Euler's constant. For a given positive integer n there are integers*

$$1 \leq n_1 < n_2 < \cdots < n_k \leq e^{n - \gamma} \left\{ 1 + \left(\frac{9}{2} + o(1) \right) \frac{\log^2 n}{n} \right\},$$

for some k, such that

$$n = \frac{1}{n_1} + \frac{1}{n_2} + \ldots + \frac{1}{n_k}.$$

To answer these questions of Erdős and Graham above, let $\sum_{1\le n\le x} 1/n = m+\delta$, where $m = m(x)$ is the integer part of $\sum_{1\le n\le x} 1/n$, and $\delta = \delta(x)$ is the fractional part.

We have trivially that $N(x) \subseteq \{1,2,...,m\}$ and our main theorem tells us that when x is sufficiently large, $\{1,2,...,m-1\} \subseteq N(x)$. Moreover, if $\delta > ((\frac{9}{2}+o(1))(\log\log x)^2/\log x$ then $m \in N(x)$ so that $N(x) = \{1,2,...,m\}$, and if $\delta < (\frac{1}{2}+o(1))(\log\log x)^2/\log x$ then $m \notin N(x)$ so that $N(x) = \{1,2,...,m-1\}$. Let $D(x) = (\frac{1}{2}+o(1))(\log\log x)^2/\log x$. We believe that the upper bound in the Main Theorem is the truth, which if true would say that for x sufficiently large,

$$N(x) = \begin{cases} \{1,2,...,m\}, & \text{if } \delta > D(x) \\ \{1,2,...,m-1\}, & \text{if } \delta < D(x). \end{cases}$$

To prove the Main Theorem we will need to introduce some notation. For any given prime power p^a and any integer $x \ge 1$, define $S(p^a,x)$ to be the set of integers $n \le x$ such that if $q^b | n$ then $q^b < p^a$. Define

$$\begin{aligned} f(p^a,x) \;:=\; \max_{1\le l\le p-1} \min\Bigg\{ \frac{1}{x_1}+\cdots+\frac{1}{x_k} \;:\; & \frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_k} \equiv l \pmod p, \\ & 1 \le x_1 < \cdots < x_k \le x \text{ each } x_i \in S(p^a,x) \Bigg\} \end{aligned}$$

(let $f=\infty$ if such a 'maximum' does not exist). For $c>0$ let

$$F(x,c) \;=\; \sum_{p^a \le x/\log^c x} \frac{f(p^a, x/p^a)}{p^a} + \sum \left\{ \frac{1}{mp^a} \;:\; x/\log^c x < p^a \le x,\ mp^a \le x \right\}.$$

The idea of the proof of the Main Theorem is as follows: we start with the full sum $\sum_{1\le n\le x} 1/n$ and then try to remove as few terms as we can so that the sum of the remaining terms is an integer. We want the contribution of those terms we remove to be as small as possible and Proposition 1 below tells us that this contribution need be no bigger than $F(x,c)$, for any $c>0$.

Proposition 1. *For any given integer x there exists a subset T of the integers $\leq x$ such that $S_0 = S_o(x) = \sum_{t\in T} 1/t$ is an integer satisfying*

$$0 < \sum_{n\leq x} \frac{1}{n} - S_0(x) \leq F(x,c),$$

for all $c \geq 0$.

To see why Proposition 1 is true, we first remove all of those terms in the full sum where n has a prime power factor bigger than $x/\log^c x$, which accounts for the second summand in the definition of $F(x,c)$ above. Call the sum of the remaining terms $S = u/v$, where $gcd(u,v) = 1$. We observe that if $p^a|v$ then $p^a < x/\log^c x$. Let $q^b \leq x/\log^c x$ be the largest prime power dividing v. We try to find numbers $1 \leq x_1 < x_2 < \cdots < x_k \leq x/q^b$, where:

1. $q \nmid x_i$ for $i = 1, 2, ..., k$.
2. All of the prime power factors of the x_i are $< q^b$; and
3. $$q^b S - \tfrac{1}{x_1} - \tfrac{1}{x_2} - \cdots - \tfrac{1}{x_k} \equiv 0 \pmod{q}.$$

(Notice that $q^b S = q^b u/v$ makes sense modulo q because $q^b||v$.) From the definition of the function f we have that if $f(q^b, x/q^b) \neq \infty$ then there are such integers x_i; moreover, there is a choice with

$$\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_k} \leq f(q^b, x/q^b).$$

Let us just assume for the moment that $f(q^b, x/q^b) \neq \infty$ and let $1 \leq x_1 < x_2 < \cdots < x_k \leq x/q^b$ be any choice satisfying

$$\frac{1}{x_1} + \cdots + \frac{1}{x_k} \leq f(q^b, x/q^b).$$

We make the following three deductions.

1. Each of the numbers

$$\frac{1}{q^b x_1}, \frac{1}{q^b x_2}, ..., \frac{1}{q^b x_k}$$

are terms in the sum S.

2. If we remove these terms from S and call the new sum S' so that

$$\frac{u'}{v'} = S' = S - \frac{1}{q^b x_1} - \cdots - \frac{1}{q^b x_k}, \quad \gcd(u', v') = 1,$$

then the largest prime power dividing v' is strictly smaller than q^b.

3. This new sum S' satisfies $S - S' < f(q^b, x/q^b)/q^b$.

Now let r^c be the largest prime power dividing v'. We subtract terms from S', just like when we subtracted from the sum S, to produce yet another sum $S'' = u''/v''$, $gcd(u'', v'') = 1$, where the largest prime power dividing v'' is strictly smaller than $r^c < q^b$. This new sum S'' satisfies

$$S - S'' = (S - S') + (S' - S'') \leq \frac{f(q^b, x/q^b)}{q^b} + \frac{f(r^c, x/r^c)}{r^c}.$$

If we continue subtracting terms in this manner we eventually get down to a sum S_0 where

$$S - S_0 \leq \sum_{2 \leq p^a \leq x/\log^c x} \frac{f(p^a, x/p^a)}{p^a},$$

and S_0 is an integer. Proposition 1 now follows since

$$\sum_{1 \leq n \leq x} 1/n - S_0 \leq F(x, c).$$

We will obtain explicit bounds on $F(x,c)$ by proving the following inequality for $f(p^a, x/p^a)$:

$$f(p^a, x/p^a) \leq \begin{cases} (p-1)p^a/\mathrm{lcm}\{2,3,4,...,p^a\}, & \text{if } p^a < \frac{1}{4}\log x \\ 20/(p^a-1), & \text{if } \frac{1}{4}\log x < p^a \leq \sqrt{x} \\ 4/\log^{\epsilon/3} x, & \text{if } \sqrt{x} < p^a \leq x/(\log^{3+\epsilon} p^a) \end{cases}.$$

To prove the first case, when $p^a < \frac{1}{4}\log x$, we will show that

$$\left\{ \frac{\mathrm{lcm}\{2,3,...,p^a\}}{p^a u} \ : \ 1 \leq u \leq (p-1) \right\} \subseteq S(p^a, x),$$

and moreover that this set has a member in each residue class $\not\equiv 0 \pmod{p}$. Our bound $f(p^a, x/p^a) \leq (p-1)p^a/\mathrm{lcm}\{2,3,...,p^a\}$ then follows. Using the following identity we will have that the contribution of such small prime powers to $F(x,c)$ is < 1.

Lemma 1.

$$\sum \left\{ \frac{p-1}{\mathrm{lcm}\{2,3,...,p^a\}} \ : \ 2 \leq p^a \leq q^b, p, q \text{ prime} \right\} = \frac{\mathrm{lcm}\{2,3,...,q^b\} - 1}{\mathrm{lcm}\{2,3,...,q^b\}}.$$

The bound on $f(p^a, x/p^a)$ where $\frac{1}{4}\log x < p^a \leq \sqrt{x}$ comes directly from the following lemma:

Lemma 2. *If $p \neq 2$ then $f(p^a, p^a) < 20/(p^a - 1)$ for $p^a > 3$.*

From this lemma we will show that the contribution of such prime powers to $F(x,c)$ is $O(1/\log x)$.

Finally, the bound on $f(p^a, x/p^a)$ where $\sqrt{x} < p^a < x/\log^{3+\epsilon} x$ follows from the following Proposition and its corollary.

Proposition 2. *Suppose $\epsilon > 0$ is given. There exists a number N_ϵ such that whenever $n > N_\epsilon$ and $k > \log^{3+2\epsilon} n$, for any set of k distinct primes $2 \leq p_1 < p_2 < \cdots < p_k < \log^{3+3\epsilon} n$ which do not divide n there is a subset*

$$\{q_1, q_2, ..., q_t\} \subseteq \{p_1, p_2, ..., p_k\}$$

such that

$$\frac{1}{q_1} + \frac{1}{q_2} + \cdots + \frac{1}{q_t} \equiv l \pmod{n},$$

for any given l with $0 \leq l < n$.

Corollary to Proposition 2 Let $\delta > 0$ be given. There exists a constant M_δ so that, when $M_\delta < p^a \leq x/\log^{3+\delta} x$,

$$f(p^a, x) < \frac{4}{\log^{\delta/3} p^a}.$$

We will show that the contribution of such prime powers to $F(x, 3+\epsilon)$ is $O(1/\log x)$.

We show that the contribution of the prime powers p^a with $x/\log^{3+\epsilon} x \leq p^a \leq x$ to $F(x, 3+\epsilon)$ is $(\frac{9}{2} + o(1))(\log\log x)^2/\log x$, and so we arrive at

Proposition 3. *For all $\epsilon > 0$ we have*

$$F(x, 3+\epsilon) < 1 + (\frac{1}{2} + o(1))(3+\epsilon)^2(\log\log x)^2/\log x.$$

With Propositions 1 and 3, and the fact that every integer can be written as some Egyptian sum (see [5]), we prove the bound

$$n(x) \geq \left[\sum_{1 \leq n \leq x} \frac{1}{n} - \left(\frac{9}{2} + o(1)\right)\frac{(\log\log x)^2}{\log x}\right],$$

as claimed in the Main Theorem. To get the bound

$$n(x) \leq \left[\sum_{1 \leq n \leq x} \frac{1}{n} - \left(\frac{1}{2} + o(1)\right)\frac{(\log\log x)^2}{\log x}\right]$$

we show that if $1 \leq n_1 < n_2 < \cdots < n_k \leq x$ and

$$\frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_k}$$

is an integer, then none of the n_i's can be divisible by a prime p with $x\log\log x/\log x < p \leq x$. From this and a technical lemma it follows that

$$\sum_{1 \leq n \leq x} \frac{1}{n} - \frac{1}{n_1} - \cdots - \frac{1}{n_k} \geq \sum_{\substack{x\log\log x/\log x < p \leq x \\ mp \leq x}} \frac{1}{mp} > \frac{1}{2}\frac{(\log\log x)^2(1+o(1))}{\log x}.$$

§*2. Proofs and Technical Lemmas.*

Proof of Lemma 1. The lemma holds for $q^b = 2$, since

$$\sum_{2 \leq p^a \leq 2} \frac{p-1}{\mathrm{lcm}\{2, \ldots, p^a\}} = \frac{1}{2}.$$

Assume, for proof by mathematical induction, we have shown that the theorem holds for all prime powers q^b, where $2 \le q^b < r^c$, where r^c is some prime power. We observe that $\mathrm{lcm}\{n \ : \ 2 \le n \le t\} = \mathrm{lcm}\{p^a \ : \ 2 \le p^a \le t,\ p \text{ prime}\}$. Using this and the induction hypothesis we have

$$\begin{aligned}
&\sum_{2 \le p^a \le r^c} \frac{p-1}{\mathrm{lcm}\{2,3,...,p^a\}} \\
&\qquad = \sum_{2 \le p^a < r^c} \frac{p-1}{\mathrm{lcm}\{2,3,...,p^a\}} + \frac{r-1}{\mathrm{lcm}\{2,3,...,r^c\}} \\
&\qquad = \frac{\mathrm{lcm}\{2,3,...,r^c-2,r^c-1\}-1}{\mathrm{lcm}\{2,3,...,r^c-2,r^c-1\}} + \frac{r-1}{\mathrm{lcm}\{2,3,...,r^c\}} \\
&\qquad = \frac{r \cdot \mathrm{lcm}\{2,3,...,r^c-2,r^c-1\}-r}{\mathrm{lcm}\{2,3,...,r^c\}} + \frac{r-1}{\mathrm{lcm}\{2,3,...,r^c\}} \\
&\qquad = \frac{\mathrm{lcm}\{2,3,...,r^c\}-1}{\mathrm{lcm}\{2,3,...,r^c\}},
\end{aligned}$$

and so the theorem follows by mathematical induction.

Proof of Lemma 2. Suppose l is any integer where $1 \le l \le (p-1)$. The number of pairs (x_1, x_2) such that $1 \le x_1 < x_2 \le p^a - 1$, $p \nmid x_1 x_2$ and

$$\frac{1}{x_1} + \frac{1}{x_2} \equiv l \pmod p \tag{1}$$

is $\frac{1}{2}((p-2)p^{2a-2} - 1) > 1$ for $p^a > 3$. Now, one of these pairs must have $x_1 > p^a/10$, since the number of pairs (x_1, x_2) with $x_1 < p^a/10$ satisfying (1) is less than

$$\frac{p^{2a-1}}{10} < \frac{1}{2}\left((p-2)p^{2a-2} - 1\right),$$

whenever $p > 2$. For this pair, we will have

$$\frac{1}{x_1} + \frac{1}{x_2} < \frac{2}{x_1} < \frac{20}{p^a}.$$

Since l was arbitrary, it follows that

$$f(p^a, p^a) < \frac{20}{p^a}.$$

Lemma 3. *Let $g(x)$ be the largest prime power such that* $\mathrm{lcm}\{2,3,4,...,g(x)\} \le$ *x. When $g(x) \ge 2$ we have that $g(x) > \frac{1}{4}\log x$.*

Proof of Lemma 3. Let $h(x)$ be the next prime power after $g(x)$. Since $\pi(x) < 2x/\log x$ when $x \ge 2$ (see [4]) we have when $h(x) \ge 2$ that

$$x < \mathrm{lcm}\{2,3,...,h(x)\} < h(x)^{\pi(h(x))} < h(x)^{2h(x)/\log h(x)} = e^{2h(x)}.$$

By Bertrand's postulate that for $n \geq 2$ there is always a prime p with $n < p < 2n$ we must have $\log x < 2h(x) < 4g(x)$ when $h(x) \geq 2$ and so $g(x) > \frac{1}{4}\log x$ when $g(x) \geq 2$.

Lemma 4.

$$\sum_{\substack{x/\log^c x < p \leq x \\ mp^a \geq x}} \frac{1}{mp} = \frac{c^2}{2}\frac{(\log\log x)^2}{\log x} + O\left(\frac{\log\log x}{\log x}\right).$$

Notice that since

$$\sum_{\substack{a \geq 2 \\ x/\log^c x < p^a \leq x \\ mp^a \leq x}} \frac{1}{mp^a} \ll \sum_{\substack{a \geq 2 \\ x/\log^c x < p^a \leq x \\ mp^a \leq x}} \frac{1}{p^a} \sum_{m \leq \log^c x} \frac{1}{m} \ll \frac{\log\log x}{x/\log^c x} \sum_{\substack{p^a \leq x \\ a \geq 2}} 1 \ll \frac{\log^c x}{\sqrt{x}},$$

Lemma 4 is also true if we replace the sum over primes by a sum over prime powers between $x/\log^c x$ and x.

Proof of Lemma 4. Using the estimate $\sum_{p \leq x} 1/p = \log\log x + B + o(1/\log x)$ (see [4]), we have for any $x/\log^c x < t \leq x$ that

$$\begin{aligned}\sum_{t/e < p < t} \frac{1}{p} &= \log\log t - \log\log\frac{t}{e} + o\left(\frac{1}{\log t}\right) = \log\left(\frac{\log t}{\log t - 1}\right) + o\left(\frac{1}{\log t}\right) \\ &= \frac{1+o(1)}{\log t} = \frac{1+o(1)}{\log x}.\end{aligned}$$

Then,

$$\begin{aligned}\sum_{\substack{x/\log^c x < p \leq x \\ mp \leq x}} \frac{1}{mp} &= \sum_{\substack{x/\log^c x < p \leq x \\ mp \leq x}} \frac{1}{p} \sum_{m \leq x/p} \frac{1}{m} = \sum_{x/\log^c x < p \leq x} \frac{\log(x/p) + O(1)}{p} \\ &= \left(\sum_{1 \leq j \leq [c\log\log x]} \sum_{x/e^j < p \leq x/e^{j-1}} \frac{j + O(1)}{p}\right) \\ &\quad + O\left(\sum_{x/\log^c x < p \leq ex/\log^c x} \frac{c\log\log x}{p}\right) \\ &= \sum_{1 \leq j \leq [c\log\log x]} \frac{j + O(1)}{\log x} + O\left(\frac{\log\log x}{\log x}\right) \\ &= \frac{c^2}{2}\frac{(\log\log x)^2}{\log x} + O\left(\frac{\log\log x}{\log x}\right).\end{aligned}$$

§3. *Proof of Proposition 1.* Let $2 = q_1 < 3 = q_2 < 4 = q_3 < \ldots$ be the sequence of prime powers. Let

$$\frac{u}{v} = \sum_{2 \leq n \leq x} \frac{1}{n} - \sum\left\{\frac{1}{n} : n \leq x,\ n = mp^a,\ p^a > x/\log^c x\right\}.$$

Choose r so that q_r is the largest prime power dividing u/v (notice that $q_r \le x/\log^c x$) and let $u_r/v_r = u/v$. Define $T_r := \{n \le x : p^a | n \Rightarrow p^a \le x/\log^6 x\}$ so that $u_r/v_r = \sum_{t\in T_r} 1/t$. We shall recursively define u_j/v_j and T_j, for $j = r-1, r-2, ..., 0$ where $u_j/v_j = \sum_{t\in T_j} 1/t$, $T_j \subseteq T_{j+1}$, and so that

(i) $\{n \le x \ : \ p^a | n \Rightarrow p^a \le q_j\} \subseteq T_j$, and

(ii) $p^a | v_j \Rightarrow p^a \le q_j$,

where we take $q_0 = 1$. Then we take $T = T_0$ in the Proposition since (ii) implies $v_0 = 1$ so that u_0/v_0 is an integer.

If q_j does not divide v_j, let $T_{j-1} = T_j$ and $u_{j-1}/v_{j-1} = u_j/v_j$; otherwise, assume q_j divides v_j and suppose q_j is some power of the prime p. Let $l \equiv q_j u_j/v_j \pmod p$ and select, if we can, integers $1 \le x_1 < x_2 < \cdots < x_k \le x/q_j$, each belonging to $S(q_j, x/q_j)$, so that

$$\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_k} \equiv l \pmod p.$$

Then let $S_j := \{q_j x_1, q_j x_2, ..., q_j x_k\}$. Note that, by (i), $S_j \subseteq T_j$. Define $T_{j-1} := T_j \setminus S_j$. Thus

$$\frac{u_{j-1}}{v_{j-1}} = \sum_{t\in T_{j-1}} \frac{1}{t} = \frac{u_j}{v_j} - \sum_{s\in S_j} \frac{1}{s}.$$

We see immediately that (i) above is satisfied. Now

$$q_j\left(\frac{u_j}{v_j} - \sum_{s\in S_j}\frac{1}{s}\right) = \frac{q_j u_j}{v_j} - \left(\frac{1}{x_1} + \cdots + \frac{1}{x_k}\right) \equiv 0 \pmod p.$$

Since $q_j \nmid v_{j-1}$ we have that (ii) is satisfied. Finally note that, by definition

$$\sum_{s\in S_j} \frac{1}{s} \le \frac{f(q_j, x/q_j)}{q_j},$$

and so the Proposition follows.

§4. *Proof of Proposition 2 and its Corollary.*

Proof of Proposition 2. Suppose that b is coprime to n, and let $r_n(a/b)$ denote the least residue of $ab^{-1} \pmod n$ in absolute value. The number of subsets of $\{p_1, ..., p_k\}$ whose sum of reciprocals is $\equiv l \pmod n$ is then given by

$$S_l \ := \ \frac{1}{n}\sum_{h=0}^{n-1} e\left(\frac{-hl}{n}\right)\prod_{j=1}^{k}\left(1 + e\left(\frac{r_n(h/p_j)}{n}\right)\right),$$

where $e(x)$ is defined to be $e^{2\pi i x}$. Define

$$P(h) \ := \ \prod_{j=1}^{k}\left(1 + e\left(\frac{r_n(h/p_j)}{n}\right)\right).$$

We will show that

$$|P(h)| < \frac{2^k}{n}, \tag{2}$$

when $h \neq 0$ and when n is sufficiently large. It will then follow that

$$|S_l| = \left| \frac{1}{n} \sum_{h=0}^{n-1} P(h) \right| > \frac{1}{n} \left\{ 2^k - \sum_{h=1}^{n-1} \frac{2^k}{n} \right\} = \frac{2^k}{n^2} > 0,$$

and thus there is at least one subset of $\{p_1, ..., p_k\}$ with the desired property.

To prove (2) we note that

$$\begin{aligned} |P(h)| &= \left| \prod_{j=1}^{k} e\left(\frac{r_n(h/p_j)}{2n}\right) \left\{ e\left(\frac{-r_n(h/p_j)}{2n}\right) + e\left(\frac{r_n(h/p_j)}{2n}\right) \right\} \right| \\ &= 2^k \left| \prod_{j=1}^{k} \cos\left(\pi \frac{r_n(h/p_j)}{n}\right) \right|. \end{aligned} \tag{3}$$

We may write

$$r_n(h/p_j) = \frac{s_j n + h}{p_j},$$

where $0 \leq h \leq (n-1)$ and s_j is an integer satisfying $-[\frac{1}{2}p_j] < s_j \leq [\frac{1}{2}p_j]$. Define $L(x) := \log^{2+2\epsilon} x + 1$. We will now show that when n is sufficiently large at least $\frac{1}{2}k$ of the s_j's have the property that $|s_j| > L(n)$. Since, if we suppose there are infinitely many n where at least $\frac{1}{2}k$ of the s_j's satisfy $|s_j| \leq L(n)$ then, by the pigeonhole principle, there is a number m with $|m| \leq L(n)$ such that $s_j = m$ for at least

$$\frac{k/2}{2L(n)+1} > \frac{\log^{3+2\epsilon} n}{4\log^{2+2\epsilon} n + 6} \gg \log n$$

of the primes p_j dividing $mn + h$ when n is sufficiently large. However, this is impossible for large n since $|mn+h| < |n(L(n)+1)| < n^2$ has $o(\log n)$ distinct prime factors. Thus when n is sufficiently large at least $\frac{1}{2}k$ of the s_j's satisfy $|s_j| > L(n)$.

It follows that, when n is sufficiently large, at least $\frac{1}{2}k$ of the p_j's satisfy

$$|r_n(h/p_j)| = \left| \frac{s_j n + h}{p_j} \right| > \left| \frac{(s_j - 1)n}{p_j} \right| > \frac{n}{\log^{1+\epsilon} n}.$$

We have for such primes p_j that, when n is sufficiently large,

$$\begin{aligned} \left| \cos\left(\frac{\pi r_n(h/p_j)}{n}\right) \right| &= \left| 1 - \frac{1}{2}\left(\frac{\pi r_n(h/p_j)}{n}\right)^2 + O\left(\left(\frac{r_n(h/p_j)}{n}\right)^4\right) \right| \\ &< 1 - \frac{\pi^2}{2\log^{2+2\epsilon} n} + O\left(\frac{1}{\log^{4+4\epsilon} n}\right), \end{aligned}$$

and so, from (3), since $k > \log^{3+2\epsilon} n$ we have that

$$|P(h)| < 2^k \left(1 - \frac{\pi^2}{2\log^{2+2\epsilon} n} + O\left(\frac{1}{\log^{4+4\epsilon} n}\right)\right)^{k/2} \ll 2^k e^{-\frac{\pi^2 \log n}{4}} = o\left(\frac{2^k}{n}\right),$$

which was just what we needed to show in order to prove our Proposition.

Proof of Corollary. Let $N_{\delta/3}$ be as in Proposition 2, and let N'_δ be the smallest number so that when $x \geq N'_\delta$ there are at least $\log^{3+\frac{2}{3}\delta} x + 1$ primes q with $\frac{1}{2}\log^{3+\delta} x \leq q \leq \log^{3+\delta} x$ where $q \neq p$. Suppose l is any number where $0 \leq l \leq (p-1)$ and p^a is any prime power where

$$x/\log^{3+\delta} x \geq p^a > M_\delta \ := \ \max\{N_{\delta/3}, N'_\delta\}.$$

By Proposition 2 there are primes q_1, q_2, ..., q_t with

$$\frac{1}{2}\log^{3+\delta} x \leq q_1 < q_2 < \cdots < q_t \leq \log^{3+\delta} x$$

and $t \leq log^{3+\frac{2}{3}\delta} x + 1$ such that

$$\frac{1}{q_1} + \frac{1}{q_2} + \cdots + \frac{1}{q_t} \equiv l \pmod{p}.$$

Also

$$\frac{1}{q_1} + \cdots + \frac{1}{q_t} < \frac{t}{\frac{1}{2}\log^{3+\delta} x} \leq \frac{2(\log^{3+\frac{2}{3}\delta} x + 1)}{\log^{3+\delta} x} < \frac{4}{\log^{\delta/3} x}.$$

Since this bound holds for all l with $0 \leq l \leq (p-1)$ the Corollary follows.

§5. *Proof of Proposition 3.* Let $g(x)$ be as in Lemma 3. We may then write

$$F(x, 3+\epsilon) = A(x) + B(x) + C(x) + D(x) + E(x),$$

where

$$\begin{aligned}
A(x) &= \sum_{2 \leq p^a \leq g(x)} \frac{f(p^a, x/p^a)}{p^a}, \\
B(x) &= \sum_{g(x) < p^a \leq \sqrt{x}} \frac{f(p^a, x/p^a)}{p^a}, \\
C(x) &= \sum_{\sqrt{x} < p^a \leq x/\log^6 x} \frac{f(p^a, x/p^a)}{p^a}, \\
D(x) &= \sum_{x/\log^6 x < p^a \leq x/\log^{3+\epsilon} x} \frac{f(p^a, x/p^a)}{p^a}, \text{ and} \\
E(x) &= \sum \left\{ \frac{1}{mp^a} : x/\log^{3+\epsilon} x < p^a \leq x, \ mp^a \leq x \right\}
\end{aligned}$$

From Lemma 4 we have that $E(x) = (\frac{1}{2}+o(1))(3+\epsilon)^2(\log\log x)^2/\log x$. We will show that $A(x) < 1$ and that $B(x)$, $C(x)$ and $D(x)$ are each $O(1/\log x)$ and so the Proposition will follow.

For each prime power p^a define

$$U_{p^a} := \left\{ \frac{\mathrm{lcm}\{2,3,...,p^a\}}{p^a u} : 1 \le u \le (p-1) \right\}.$$

We have that $U_{p^a} \subset S(p^a, x/p^a)$, for $p^a \le g(x)$. Also, for each l with $1 \le l \le (p-1)$, there is an element $y \in U_{p^a}$ such that $1/y \equiv l \pmod p$. Thus for $p^a \le g(x)$ we have

$$f(p^a, x/p^a) \le \frac{1}{\min(U_{p^a})} = \frac{p^a(p-1)}{\mathrm{lcm}\{2,3,...,p^a\}}.$$

From this and Lemma 1 it follows that

$$\begin{aligned} A(x) &= \sum_{2 \le p^a \le g(x)} \frac{f(p^a, x/p^a)}{p^a} \\ &\le \sum_{2 \le p^a \le g(x)} \frac{p-1}{\mathrm{lcm}\{2,3,...,p^a\}} = \frac{\mathrm{lcm}\{2,3,...,g(x)\}-1}{\mathrm{lcm}\{2,3,...,g(x)\}}. \end{aligned}$$

Thus $A(x) < 1$.

By Lemma 2, if $p \ne 2$ and $p^a \le \sqrt{x}$ then $f(p^a, x/p^a) \le f(p^a, p^a) < 20/p^a$. Thus by this and Lemma 3 we have for $g(x) \ge 2$ that

$$\begin{aligned} B(x) &= \sum_{g(x) < p^a \le \sqrt{x}} \frac{f(p^a, x/p^a)}{p^a} < \sum_{\frac{1}{4}\log x < p^a \le \sqrt{x}} \frac{f(p^a, x/p^a)}{p^a} \\ &\le \sum_{2^a > \frac{1}{4}\log x} \frac{1}{2^a} + \sum_{\substack{\frac{1}{4}\log x < p^a \le \sqrt{x} \\ p \text{ odd}}} \frac{20}{p^{2a}} < \frac{4}{\log x} + 20 \sum_{n > \frac{1}{4}\log x} \frac{1}{n^2} \\ &= O\left(\frac{1}{\log x}\right). \end{aligned}$$

By the Corollary to Proposition 2 when x is sufficiently large and $p^a \le x/\log^6 x$ we have $f(p^a, x) < 4/\log p^a$. Then for x sufficiently large we have

$$C(x) = \sum_{\sqrt{x} < p^a < x/\log^6 x} \frac{f(p^a, x/p^a)}{p^a} < 4 \sum_{\sqrt{x} < p^a < x/\log^6 x} \frac{1}{p^a \log p^a} = O\left(\frac{1}{\log x}\right)$$

by the Prime Number Theorem.

Again by the Corollary to Proposition 2 we have when x is sufficiently large and $p^a \le x/\log^{3+\epsilon}(x)$ that $f(p^a, x) < 4/\log^{\epsilon/3} p^a = O(1/\log^{\epsilon/3} x)$. Thus

$$\begin{aligned} D(x) &= \sum_{x/\log^6(x) < p^a \le x/\log^{3+\epsilon}(x)} \frac{f(p^a, x/p^a)}{p^a} \ll \frac{4}{\log^{\epsilon/3} x} \sum_{x/\log^6 x < p^a \le x/\log^{3+\epsilon} x} \frac{1}{p^a} \\ &= O\left(\frac{\log\log x}{\log^{1+\epsilon/3} x}\right). \end{aligned}$$

§6. *Proof of the Main Theorem.* From Propositions 1 and 3 we have that there is an integer integer $\tau(x) \in N(x)$ with

$$\tau(x) > \sum_{1 \le n \le x} \frac{1}{n} - 1 - \frac{9}{2}\frac{(\log\log x)^2(1+o(1))}{\log x}.$$

Thus

$$\left[\sum_{1 \le n \le x} \frac{1}{n} - \frac{9}{2}\frac{(\log\log x)^2(1+o(1))}{\log x}\right] \le \tau(x) \le \left[\sum_{1 \le n \le x} \frac{1}{n}\right].$$

Therefore there exists an x_0 so that when $x > x_0$ we have

$$\begin{aligned} 0 &\le |\tau(x+1) - \tau(x)| \\ &\le \left[\sum_{1 \le n \le x+1} \frac{1}{n}\right] - \left[\sum_{1 \le n \le x} \frac{1}{n} - \frac{9}{2}\frac{(\log\log x)^2(1+o(1))}{\log x}\right] \le 1. \end{aligned}$$

It follows that when $x > x_0$ is sufficiently large so that $\tau(x_0) \le n \le \tau(x)$, then $n = \tau(t)$ for some $x_0 \le t \le x$. In particular this says that if $\tau(x_0) \le n \le \tau(x)$ then there exist integers $1 \le n_1 < n_2 < \cdots < n_k \le x$, for some k, so that

$$n = \frac{1}{n_1} + \cdots + \frac{1}{n_k}.$$

As a consequence of the main result in [5] (and [6]) we have that for x sufficiently large and $1 \le n \le \tau(x_0)$, there exist integers $1 \le n_1 < n_2 < \cdots < n_k \le x_0$, for some k, so that

$$n = \frac{1}{n_1} + \cdots + \frac{1}{n_k}.$$

We conclude that when $x > x_0$,

$$n(x) \ge \tau(x) \ge \left[\sum_{1 \le n \le x} \frac{1}{n} - \frac{9}{2}\frac{(\log\log x)^2(1+o(1))}{\log x}\right]$$

as claimed.

Suppose now that $1 \le n_1 < n_2 < \cdots < n_k \le x$ has the property that

$$\frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_k}$$

is an integer. We claim that none of the n_i's has a prime factor greater than $x \log\log x / \log x$: for suppose p is such a prime and let $n_{i_1} = pm_1$, $n_{i_2} = pm_2, \ldots, n_{i_l} = pm_l$ be all the n_i's divisible by p. Since $(x \log\log x/\log x)m_i < pm_i \le x$ we have that such $m_i < \log x/\log\log x$ and therefore $l < \log x/\log\log x$. Also, since

$$\frac{1}{n_1} + \cdots + \frac{1}{n_k}$$

is an integer, we must have $p \nmid v$ and yet

$$p \mid m_1 m_2 \cdots m_l \left(\frac{1}{m_1} + \frac{1}{m_2} + \cdots + \frac{1}{m_l}\right) < l\left(\frac{\log x}{\log\log x}\right)^{l-1} < \left(\frac{\log x}{\log\log x}\right)^{l}$$
$$< \left(\frac{\log x}{\log\log x}\right)^{\frac{\log x}{\log\log x}} < \frac{x\log\log x}{\log x} < p.$$

We conclude that no n_i is divisible by $p > \frac{x\log\log x}{\log x}$. Using this fact and Lemma 4 we get the bound

$$\sum_{1\le n\le x}\frac{1}{n} - \sum_{1\le i\le k}\frac{1}{n_i} > \sum_{\substack{\frac{x\log\log x}{\log x}<p\le x\\ mp\le x}} \frac{1}{mp} = \frac{1}{2}\frac{(\log\log x)^2(1+o(1))}{\log x},$$

and so

$$n(x) \le \left[\sum_{1\le n\le x}\frac{1}{n} - \frac{1}{2}\frac{(\log\log x)^2(1+o(1))}{\log x}\right]$$

as claimed.

§7. *Proof of the Corollary to the Main Theorem.* Fix an $\epsilon > 0$. Let m be a sufficiently large positive integer and select x for which

$$m < \sum_{1\le n\le x}\frac{1}{n} - \left(\frac{9}{2}+\epsilon\right)\frac{(\log\log x)^2}{\log x}$$
$$= \log x + \gamma + O\left(\frac{1}{x}\right) - \left(\frac{9}{2}+\epsilon\right)\frac{(\log\log x)^2}{\log x},$$

From the Main Theorem we know that for m sufficiently large there exist integers $1 \le n_1 < n_2 < \cdots < n_k \le x$, for some k, so that

$$m = \frac{1}{n_1} + \cdots + \frac{1}{n_k}. \tag{4}$$

Since $x > e^{n+O(1)}$, we have that $n < \log x + \gamma - (\frac{9}{2}+\epsilon+o(1))(\log^2 n)/n$. Thus, as long as x satisfies

$$x > e^{n-\gamma+(\frac{9}{2}+\epsilon+o(1))(\log^2 n)/n} = e^{n-\gamma}\left\{1 + \left(\frac{9}{2}+\epsilon+o(1)\right)\frac{\log^2 n}{n}\right\},$$

equation (4) above has a solution with $n_k < x$. Since $\epsilon > 0$ was arbitrary, the Corollary follows.

§8. *Acknowledgements*

I would like to thank Andrew Granville for substantially shortening the proofs of some of the key Propositions and Lemmas in the paper, for encouraging me to prove sharper and sharper versions of the Main Theorem, and for generously dontating his time reading through the paper. I would also like to thank Greg Martin and Christian Elsholtz for some enlightening comments and conversations I had with them by email and in person.

References

1. Pál Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory*, L'Enseignement Mathématique Université de Genève, 1980, pp. 39-40, 103.
2. Richard K. Guy, *Unsolved Problems in Number Theory, Second Edition*, Springer Verlag, 1994, pp. 158-166.
3. Trygve Nagell, *Introduction to Number Theory, Second Edition*, Chelsea Publishing Company, 1964, pp. 54-67.
4. J. B. Rosser and Lowell Shoenfeld, *Approximate Formulas for Some Functions of Prime Numbers*, Illinois Journal of Mathematics **6** (1962), 70.
5. Hisashi Yokota, *On Number of Integers Representable as a Sum of Unit Fractions, II*, Journal of Number Theory **67** (1997), 162-169.
6. ————-, *Corrigendum*, Journal of Number Theory **72** (1998), 150.