CRM
U. Montreal, May 28, 2018

The Erdős conjecture on primitive sets

Carl Pomerance

Dartmouth College (emeritus)

University of Georgia (emeritus)

Joint work with Jared Duker Lichtman

Suppose that M is a set of multiples: $mn \in M$ whenever $m \in M$. Some examples:

- $\{m \in \mathbb{N} : m \geq m_0\}$,
- $\{m \in \mathbb{N} : \omega(m) \geq k_0\}$ (where $\omega(m) = \sum_{p|m} 1$),
- $\{m \in \mathbb{N} : \Omega(m) \geq k_0\}$ (where $\Omega(m) = \sum_{p^a|m} 1$),
- $\{m \in \mathbb{N} : \sigma(m)/m > 2\}$ (where $\sigma(m) = \sum_{d|m} 1$),
- $\{m \in \mathbb{N} : \sigma(m)/m \geq 2\}$.

A set of multiples M is like an ideal, but it is not necessarily closed under addition.

Like an ideal in $\mathbb{Z}$, a set of multiples in $\mathbb{N}$ is generated by minimal elements. In particular, if $A = A(M)$ is the subset of M consisting of positive integers that are not divisible by any smaller element of M, then evidently

$$M = \{an : a \in A, n \in \mathbb{N}\}.$$

Further, the set A is **primitive**: no element divides another.

Here are the primitive sets corresponding to our examples:

- $A(\{m \in \mathbb{N} : m \geq m_0\}) = ($ a finite set $) \cup \{p \geq m_0 : p$ prime$\}$,

- $A(\{m \in \mathbb{N} : \omega(m) \geq k_0\}) = \{m \in \mathbb{N} : \omega(m) = \Omega(m) = k_0\}$,

- $A(\{m \in \mathbb{N} : \Omega(m) \geq k_0\}) = \{m \in \mathbb{N} : \Omega(m) = k_0\}$,

- $A(\{m \in \mathbb{N} : \sigma(m)/m > 2\}) = \{m \in \mathbb{N} : m$ is primitive abundant$\}$,

- $A(\{m \in \mathbb{N} : \sigma(m)/m \geq 2\}) = \{m \in \mathbb{N} : m$ is primitive nondeficient$\}$.

Lemma. *If M is a set of multiples and $\sum_{a\in A(M)} 1/a < \infty$, then M has a natural density.*

This Lemma applies to just one of our example sets: The primitive nondeficient numbers have a finite reciprocal sum (Erdős, 1934), so the nondeficient numbers have an asymptotic density. We now know (Kobayashi, 2010) that this density is between 0.2476171 and 0.2476475. And we know (Lichtman, 2018) that the sum of reciprocals of primitive nondeficient numbers is between 0.34842 and 0.37937.

The first 3 sets of multiples presented all have asymptotic density 1. The set of abundant numbers has the same density as the set of nondeficient numbers, since the only difference is the set of perfect numbers, which can be shown by other methods to have density 0.

One might guess that a primitive set always has asymptotic density 0. It's true for our 5 examples, and more generally, it's true that the lower asymptotic density of a primitive set must be 0. Somewhat counter-intuitively, the upper asymptotic density need not be 0 !

Here's a construction. The set of integers in the interval $(x, 2x]$ is primitive; let $D(x)$ be the asymptotic density of the set of multiples. We know after Besicovitch (1934) that $\liminf D(x) = 0$. In fact, after work of Erdős, Tenenbaum, and Ford, we now know that

$$D(x) \asymp \frac{1}{(\log x)^c (\log\log x)^{3/2}}, \quad c = 1 - \frac{1 + \log\log 2}{\log 2} = 0.08607\ldots.$$

But already, just from Besicovitch's result, we can construct primitive sets with upper asymptotic density arbitrarily close to 1/2. Namely, choose a very briskly increasing sequence $x_1, x_2, \ldots$ with $D(x_1)$ very small and $D(x_j) \downarrow 0$. Take all numbers in $(x_1, 2x_1]$, all numbers in $(x_2, 2x_2]$ not divisible by any number in $(x_1, 2x_1]$, all numbers in $(x_3, 2x_3]$ not divisible by any number previously chosen, etc.

This result is best possible: The upper asymptotic density of a primitive set is always $< 1/2$.

From now on we only consider primitive sets $A \neq \{1\}$.

Theorem (Erdős, 1935). *If A is a primitive set, then*

$$\sum_{a \in A} \frac{1}{a \log a} < \infty.$$

In fact, the sum $\sum_{a \in A} 1/(a \log a)$ is uniformly bounded as A varies over primitive sets.

Let $f(A) = \sum_{a \in A} 1/(a \log a)$. With $\mathcal{P}$ the set of primes, let $C = f(\mathcal{P}) = 1.63661632336\ldots$, the calculation done by Cohen.

Conjecture (Erdős, 1988). *For A primitive, $f(A) \leq C$.*

Conjecture (Erdős, 1988). *For A primitive, $f(A) \le C$, where $f(A) = \sum_{a \in A} 1/(a \log a)$.*

What do we know about $f(A)$?

Erdős, Zhang (unpublished): $f(A) < 2.886$.

Robin (unpublished): $f(A) < 2.77$.

Erdős, Zhang (1993): $f(A) < 1.84$.

Clark (1995): $f(A) \le e^{\gamma} = 1.78107\ldots$.

The first two results used the original Erdős argument, but the 1993 paper used a new argument. We do not understand the Clark argument, nor have we been successful in duplicating the result.

Let $\mathcal{P}(A)$ denote the set of primes which divide some member of A. Note that the Erdős conjecture is equivalent to the assertion that

$$f(A) \le f(\mathcal{P}(A))$$

for all primitive sets A. Indeed, if $f(A) > f(\mathcal{P}(A))$ for some primitive set A, let A' be A together with every prime not in $\mathcal{P}(A)$, so that A' is primitive, $\mathcal{P}(A') = \mathcal{P}$, and $f(A') > f(\mathcal{P})$.

Let $\mathbb{N}_k = \{n : \Omega(n) = k\}$.

Zhang, (1991): $f(A) \le C$ *if each* $a \in A$ *has* $\Omega(a) \le 4$.

Zhang, (1993): *For each* $k \ge 2$, $f(\mathbb{N}_k) < f(\mathbb{N}_1) = C$.

Banks, Martin, (2013): *If* $\sum_{p \in \mathcal{P}(A)} 1/p < 1.7401\ldots$, *then* $f(A) \le f(\mathcal{P}(A))$.

Banks, Martin, (2013): **Conjecture**: $f(\mathbb{N}_1) > f(\mathbb{N}_2) > f(\mathbb{N}_3) \ldots$.

Bayless, Kinlaw, Klyve, (2018): $f(\mathbb{N}_2) > f(\mathbb{N}_3)$.

Lichtman, P, (2018). *We have the following results.*

- $f(A) < e^{\gamma} + 10^{-5}$.

- *If no member of* A *is divisible by* 8*, then* $f(A) < C + 10^{-5}$.

- *There is an absolute constant* $c > 0$ *such that* $f(\mathbb{N}_k) > c$ *for all* k.

- *Assuming RH and LI, there is a set of primes* $\mathcal{P}_0$ *of relative lower logarithmic density* ≥ 0.995 *such that* $f(A) \leq f(\mathcal{P}(A))$ *when* $\mathcal{P}(A) \subset \mathcal{P}_0$*. Unconditionally,* $\mathcal{P}_0$ *contains all of the odd primes up to* $\exp(10^6)$.

Note: The relative lower logarithmic density of a set of primes $\mathcal{P}_0$ is

$$\liminf_{x\to\infty} \frac{1}{\log\log x} \sum_{\substack{p\in\mathcal{P}_0 \\ p\le x}} \frac{1}{p}\,.$$

Notation: For an integer $a \ge 2$, let

$$p(a) := \min\{p : p \mid a\},$$
$$P(a) := \max\{p : p \mid a\}.$$

A slightly cleaned up version of the 1935 Erdős argument:

For $a \in A$, let $S_a = \{ba : p(b) \geq P(a)\}$. The asymptotic density of S_a is

$$\delta(S_a) = g(a) = \frac{1}{a} \prod_{p<P(a)} \left(1 - \frac{1}{p}\right).$$

Moreover the sets S_a, as a varies over A, are pairwise disjoint (since A is primitive). The union of these sets is contained in the set of multiples of members of A, so

$$\sum_{a \in A} g(a) = \sum_{a \in A} \frac{1}{a} \prod_{p<P(a)} \left(1 - \frac{1}{p}\right) \leq 1.$$

But $\prod_{p<P(a)}(1 - 1/p) \gg 1/\log(P(a)) \geq 1/\log a$, so that $f(A) \ll 1$.

To do a little better, we should be more careful with the step where we say $\prod_{p<P(a)}(1-1/p) \gg 1/\log(P(a)) > 1/\log a$.

Note that as $x \to \infty$, we have $\prod_{p \le x}(1-1/p) \sim 1/(e^\gamma \log x)$. Also, if a is composite, then $a \ge 2P(a)$.

Lemma. *We have* $\prod_{p \le x}(1-1/p) > 1/(e^\gamma \log(2x))$*.*

Conclude: If a is composite, then $\prod_{p<P(a)}(1-1/p) > 1/(e^\gamma \log a)$. And if every member of A is composite, then $f(A) < e^\gamma$.

To go further, we consider various subsets of A defined by the least prime factor of the elements: Let $A(q) = \{a \in A : p(a) = q\}$. Then

$$\sum_{a \in A(q)} g(a) = \sum_{a \in A(q)} \delta(S_a) \ \le \ \delta(\{ba : a \in A(q), p(b) \ge q\}) \ = \ g(q).$$

Note that if $q \notin A(q)$, then every member of $A(q)$ is composite. This implies $g(a) > 1/(e^{\gamma} a \log a)$ for $a \in A(q)$, so that

$$f(A(q)) \ < \ e^{\gamma} g(q) \ = \ \frac{e^{\gamma}}{q} \prod_{p<q} \left(1 - \frac{1}{p}\right).$$

Else, $q \in A(q)$, so that $A(q) = \{q\}$ and $f(A(q)) = 1/(q \log q)$.

Say a prime q is **Mertens** if $e^{\gamma} \prod_{p<q}(1 - 1/p) \le 1/\log q$.
So, if q is Mertens, then $f(A(q)) \le 1/(q \log q)$.
Thus, if every prime in $\mathcal{P}(A)$ is Mertens, then $f(A) \le f(\mathcal{P}(A))$.

A prime q is **Mertens** if $e^\gamma \prod_{p<q}(1-1/p) \le 1/\log q$. And, if every prime in $\mathcal{P}(A)$ is Mertens, then $f(A) \le f(\mathcal{P}(A))$. That is, the Erdős conjecture is true for sets supported on the Mertens primes.

Let's try it out. Is 2 Mertens?

$$\prod_{p<2}\left(1-\frac{1}{p}\right) = 1, \qquad \frac{1}{e^\gamma \log 2} = 0.81001\ldots.$$

So, 2 is not Mertens. ☹

A prime q is **Mertens** if $e^\gamma \prod_{p<q}(1-1/p) \le 1/\log q$. And, if every prime in $\mathcal{P}(A)$ is Mertens, then $f(A) \le f(\mathcal{P}(A))$. That is, the Erdős conjecture is true for sets supported on the Mertens primes.

Let's try it out. Is 2 Mertens?

$$\prod_{p<2}\left(1-\frac{1}{p}\right) = 1, \qquad \frac{1}{e^\gamma \log 2} = 0.81001\ldots .$$

So, 2 is not Mertens. ☹

But every odd prime up to p_{10^8} is indeed Mertens. ☺

Theorem (Lamzouri, 2016). *Assuming RH and LI, the set of real numbers x with $e^\gamma \prod_{p\le x}(1-1/p) < 1/\log x$ has logarithmic density* $0.99999973\ldots$.

Corollary (Lichtman, P, 2018). *Assuming RH and LI, the set of Mertens primes has relative logarithmic density* > 0.995.

We could push the verification of Mertens primes beyond the 10^8th prime, but instead, here's an alternative that allows us to push things *much* higher.

Say a pair of primes q, r is a **Mertens pair** if $q \le r$ and

$$\prod_{q \le p < r} \left(1 - \frac{1}{p}\right) \ \ge \ \frac{\log q}{\log qr}.$$

Recall that with $A(q)$ denoting the set of $a \in A$ with $p(a) = q$, we have

$$\sum_{a \in A(q)} \frac{1}{a} \prod_{p < P(a)} \left(1 - \frac{1}{p}\right) \ \le \ \frac{1}{q} \prod_{p < q} \left(1 - \frac{1}{p}\right).$$

So, if $q, P(a)$ is a Mertens pair for all $a \in A(q)$, then

$$\frac{1}{q} \ \ge \ \sum_{a \in A(q)} \frac{1}{a} \prod_{q \le p < P(a)} \left(1 - \frac{1}{p}\right) \ \ge \ \sum_{a \in A(q)} \frac{1}{a} \cdot \frac{\log q}{\log(qP(a))}.$$

Recapitulating, if $q, P(a)$ is a Mertens pair for all $a \in A(q)$, then

$$\sum_{a \in A(q)} \frac{1}{a \log(qP(a))} \ \le \ \frac{1}{q \log q}.$$

In particular, if $q \notin A(q)$, then $\log(qP(a)) \le \log a$, so that $f(A(q)) \le 1/(q \log q)$. This holds with equality if $q \in A(q)$. We conclude that if every $a \in A$ has $p(a), P(a)$ a Mertens pair, then the Erdős conjecture is true for A, i.e.,

$$f(A) \le f(\mathcal{P}(A)).$$

Note that if q is a Mertens prime, then q, r is a Mertens pair for all primes $r \ge q$. Indeed,

$$\prod_{q \le p < r}\left(1 - \frac{1}{p}\right) \;\ge\; e^{\gamma} \log q \prod_{p<r}\left(1 - \frac{1}{p}\right) \;>\; \frac{e^{\gamma}\log q}{e^{q}\log(2r)} \;\ge\; \frac{\log q}{\log(qr)}.$$

So, q, r is a Mertens pair for all odd primes q among the first 10^8 primes. But the play we have in going from $\log r$ to $\log(qr)$ gets better when q is large, which now can be assumed. As well, the error in Mertens theorem is smaller for large cut-offs. Using these ideas, we can show every q, r is a Mertens pair for the odd primes $\le \exp(10^6)$.

One would have a very hard time giving an explicit counterexample to the Erdős conjecture using odd primes!

The Erdős–Zhang approach

They have a two-pronged argument. First, they get an upper bound for $f(A)$ when $p(A)$, the least prime dividing any member of A, is large. Then, they use a reverse recurrence to step-by-step lower $p(A)$.

The upper bound when $p(A)$ is large is done by a clever induction and estimates for the nth prime p_n, when n is large. Using only this first part of their argument, together with the Mertens primes, we can show:

Lichtman, P, (2018). *If A is a primitive set of odd numbers, then $f(A) < f(\mathcal{P}(A)) + 10^{-5}$.*

Suppose now that A has no member divisible by 8. Using the theorem for the odd case, if $2 \in A$, then

$$f(A) = f(A \setminus \{2\}) + \frac{1}{2\log 2}$$
$$< f(\mathcal{P}(A) \setminus \{2\}) + 10^{-5} + \frac{1}{2\log 2} \ \le \ C + 10^{-5}.$$

So suppose $2 \notin A$. Let A_0 denote the odd members of A, let A_1 the members of A that are 2 mod 4, and A_2 the members of A that are 4 mod 8. Let $B_1 = A_1/2$, $B_2 = A_2/4$, so that B_1, B_2 are primitive sets of odd numbers. We have

$$f(A_1) = \sum_{b \in B_1} \frac{1}{2b\log(2b)} \ \le \ \frac{1}{2} f(B_1)$$

and similarly, $f(A_2) \le \frac{1}{4} f(B_2)$ (assuming $4 \notin A$). Thus,

$$f(A) = f(A_0) + f(A_1) + f(A_2) < \left(C + 10^{-5} - \frac{1}{2\log 2}\right)\left(1 + \frac{1}{2} + \frac{1}{4}\right) < C.$$

If $4 \in A$, then

$$f(A) < \left(C + 10^{-5} - \frac{1}{2\log 2}\right)\left(1 + \frac{1}{2}\right) + \frac{1}{4\log 4} < C.$$

Thus, the Erdős conjecture is true within 10^{-5} for primitive sets not containing a multiple of 8. (Oddly, the same is true for primitive sets that contain no number that is 4 mod 8.)

Toward Clark's claim that $f(A) \le e^\gamma$

Since $e^\gamma = 1.781072\ldots$ and $C = 1.636616\ldots$, and considering how close to C we have been able to get, it would seem easy to not only prove that e^γ always works, but even surpass it. But so far we have been unsuccessful. However, we have gotten close.

There is a chain of inferences we can use. First, if $2 \in A$, we have the upper bound $C + 10^{-5}$, so we may assume that the even members of A are composite. We've seen that when $A(q)$ is composite, we have

$$f(A(q)) < \frac{e^\gamma}{q} \prod_{p<q} \left(1 - \frac{1}{q}\right),$$

so in the case $q = 2$ we have $f(A(2)) < e^\gamma/2$.

We try to do a little better. Suppose $3 \in A$. Then, no member of $A(2)$ is divisible by 3, and the earlier argument can then be improved to $f(A(2)) < e^{\gamma}/3$. This is enough to get $f(A) < e^{\gamma}$, so we may assume $3 \notin A$.

Knowing that $3 \notin A$ then gives us that $f(A(3)) < e^{\gamma}/6$, and this is a noticeable improvement over $1/(3 \log 3)$, which would have been $f(A(3))$ if $3 \in A$. (For the record, $e^{\gamma}/6 = 0.29684\ldots$ and $1/(3 \log 3) = 0.30341\ldots$.)

We then show that we may assume $5 \notin A$, $7 \notin A$, etc., up to the prime 234,473, after which the argument breaks down. Using that the remaining primes to the 10^8th prime are Mertens, and the Erdős–Zhang argument for the higher cases, gives us the earlier announced result.

Lichtman, P, (2018). *For any primitive set A, $f(A) < e^{\gamma} + 10^{-5}$.*

Recall the Banks–Martin conjecture that $f(\mathbb{N}_1) > f(\mathbb{N}_2) > f(\mathbb{N}_3) > \ldots$, where $\mathbb{N}_k$ is the set of n with $\Omega(n) = k$. We prove that $f(\mathbb{N}_k) \gg 1$, that is, the numbers $f(\mathbb{N}_k)$ are bounded above 0.

The idea is to use the Sathe–Selberg theorem (or an earlier result of Erdős) which implies that when $|k - \log\log x| \le \sqrt{\log\log x}$, we have uniformly as $x \to \infty$

$$\sum_{\substack{n \le x \\ \Omega(n)=k}} 1 \sim \frac{x}{\log x} \frac{(\log\log x)^k}{k!}.$$

Looking at this from another angle, the asymptotic holds uniformly for all values of x between e^{e^k} and $e^{e^{k+\sqrt{k}}}$ as $k \to \infty$.

Then for $1 \le j \le \sqrt{k}$, we use the asymptotic formula, partial summation, and Stirling to show that

$$\sum_{\substack{n \in \mathbb{N}_k \\ e^{e^{k+j-1}} < n \le e^{e^{k+j}}}} \frac{1}{n \log n} \gg \frac{1}{\sqrt{k}}.$$

Now sum on j.

Essentially the same argument shows that $f(\mathbb{N}_k^0) \gg 1$, where $\mathbb{N}_k^0$ is the set of squarefree members of $\mathbb{N}_k$.

To put the Erdős conjecture in perspective, one might like to look at other "metrics" than $1/(a \log a)$.

- In Banks–Martin they look at $f_c(A) = \sum_{a \in A} 1/a^c$, where $c > 1$. They show that $f_c(A) \le f_c(\mathcal{P}(A))$ if $c > 1.1403659$.

- Let $g(A) = \sum_{a \in A} g(a)$, where $g(a) = a^{-1} \prod_{p < P(a)} (1 - 1/p)$. We showed above that $g(A(q)) \le g(q)$ for every prime q. Thus, $g(A) \le g(\mathcal{P}(A))$ for every primitive set A.

- Consider $h(A) = \sum_{a \in A} h(a)$, where $h(a) = 1/(a \log P(a))$ is "asymptotically proportional" to $g(a)$. Here, the Erdős conjecture fails. We have computed that $h(\mathbb{N}_2) > h(\mathbb{N}_1)$.

Some unsolved problems:

- Prove the Erdős conjecture for odd numbers, or for squarefree numbers. (We have it for squarefree even numbers.)

- We've shown that if A is a primitive set of even numbers, then $f(A) < e^\gamma/2$. Do better. (We conjecture $f(A) \le 1/(2\log 2) = 0.7213\ldots$, while $e^\gamma/2 = 0.8905\ldots$.)

- Prove the logarithmic density of Mertens primes is the same as the logarithmic density in Lamzouri's theorem (which is the same as in the $\pi(x), \mathrm{li}(x)$ race). Feel free to assume RH and LI!

- Is there a "race" with $\sum_{p \ge x} 1/(p \log p)$ and $1/\log x$? We can show that if $\sum_{p \ge q} 1/(p \log p) \le 1/\log q$ for every prime $q \ge 5$, then the Erdős conjecture holds for numbers coprime to 6. We have shown that the inequality holds for $3 < q \le p_{10^8}$.

Lets close with an easier, and now solved problem: What is the fewest number of primitive sets whose union contains $[2, x]$?

Lets close with an easier, and now solved problem: What is the fewest number of primitive sets whose union contains $[2, x]$?

Well, with primitive sets of the form $[y, 2y)$, we can cover $[2, x]$ with $\lfloor \log x / \log 2 \rfloor$ primitive sets. Is this best possible?

Lets close with an easier, and now solved problem: What is the fewest number of primitive sets whose union contains $[2, x]$?

Well, with primitive sets of the form $[y, 2y)$, we can cover $[2, x]$ with $\lfloor \log x / \log 2 \rfloor$ primitive sets. Is this best possible?

Yes. The powers of 2 must be in different primitive sets. (Ayla Gafni, last week at the CANT problem session in response to the question of Lichtman)

Merci