On sets whose subsets have integer mean

Enrique Treviño

LAKE FOREST COLLEGE

Integers 2025
May 17, 2025
Happy birthday Carl and Mel.

Motivating Problem

Consider the following problem that appeared as problem 2 in the 31st Mexican Mathematical Olympiad held in November 2017:

A set with n distinct positive integers is said to be *balanced* if the mean of any k numbers in the set is an integer, for any $1 \leq k \leq n$. Find the largest possible sum of the elements of a balanced set with all numbers in the set less than or equal to 2017.

Sketch of solution

- Consider a balanced set with n elements. Say $S = \{a_1, a_2, \ldots, a_n\}$.
- Let $k \leq n - 1$. Note that by fixing any $k - 1$ terms, the k-th term has to be of the same congruence modulo k for any other number. Therefore, they are all congruent modulo k.
- Since $a_i \equiv a_j \bmod k$ for all pairs i, j and all $k \leq n - 1$, then all the numbers are congruent modulo $M = \mathrm{lcm}\{1, 2, \ldots, n - 1\}$.
- Note that if $n \geq 8$, then a balanced set consists of elements congruent to $\mathrm{lcm}\{1, 2, \ldots, 7\} = 420$. Since we can't have 8 positive integers ≤ 2017 congruent to each other modulo 420, then we need to consider balanced sets with at most 7 elements.
- $S = \{2017, 2017 - 60, \ldots, 2017 - 6 \cdot 60\}$ is the balanced set with 7 elements of maximal sum (12859). If you have 6 elements or less the sum is at most $6 \cdot 2017 < 12859$.

Slight variant

Consider the same problem but with numbers ≤ 3000 instead of ≤ 2017. What happens?

- Since $420 \cdot 7 \leq 3000$, we can fit an 8-element balanced set, namely $\{3000, 3000 - 420, \ldots, 3000 - 7 \cdot 3000\}$. The sum of the elements of this set is 12240.
- The 7-element balanced set $\{3000, 3000 - 60, \ldots, 3000 - 6 \cdot 60\}$ has sum 19740.
- The 7-element balanced set has a higher sum than the 8-element balanced set!

Generalization

- For a positive integer N, let $M(N)$ be the size of the largest balanced set all of whose elements are $\leq N$.
- Let $S(N)$ be the size of the set with maximal sum among balanced sets all of whose elements are $\leq N$.

For what N is $M(N) = S(N)$?

For example $M(2017) = S(2017)$, yet $M(3000) \neq S(3000)$.

Numerics

Using a computer, we can verify that if $N \leq 1000000$, then $M(N) = S(N)$ for

$$1 \leq N \leq 18$$
$$31 \leq N \leq 48$$
$$85 \leq N \leq 300$$
$$571 \leq N \leq 2940$$
$$18481 \leq N \leq 22680$$
$$54181 \leq N \leq 304920$$

Pattern

Consider 18, 48, 300, 2940, 22680, 304920. Let

$$L(n) = \mathrm{lcm}\{1, 2, \ldots, n\}.$$

Then

$$\begin{aligned} 18 &= 3L(3) \\ 48 &= 4L(4) \\ 300 &= 5L(5) \\ 2940 &= 7L(7) \\ 22680 &= 9L(9) \\ 304920 &= 11L(11) \end{aligned}$$

Theorems about $mL(m)$

Theorem

Let p be prime. Then $M(pL(p)) = S(pL(p))$. Furthermore, $M(pL(p)+1) \neq S(pL(p)+1)$.

Theorem

If m is not a prime power, then $M(mL(m)) \neq S(mL(m))$.

Ingredients of the proofs

- To prove $M(pL(p)) = S(pL(p))$ and $M(pL(p)+1) \neq S(pL(p)+1)$ the key is that $L(p) = pL(p-1)$.
- To prove that $M(mL(m)) \neq S(mL(m))$ for m not a prime power. The key is that a balanced set with p elements where p is a prime close to m will have a higher sum than a balanced set with more elements as long as p is close enough to m.
- For non-prime powers close enough is at least larger than $m/2$. This happens due to Bertrand's postulate.

Towards stronger statements

Bertrand's postulate is not the best analytic number theory can do in terms of primes close to *m*. Here's a recent theorem of Dudek (2016):

Theorem

For $m \geq e^{e^{33.3}}$, there exists a prime p such that $m^3 \leq p < m^3 + 3m^2$. In particular, there is a prime p such that

$$m^3 < p < (m+1)^3.$$

We can prove a slight variant:

Lemma

For all $m \geq 10^{10^{15}}$ there is a prime p such that

$$m^3 - \frac{1}{3}m^2 < p < m^3.$$

Stronger statements

Theorem

For $m \geq 10^{10^{15}}$ of the form q^k for a prime q and an exponent $k \geq 3$, then $M(mL(m)) \neq S(mL(m))$.

Using results from Carneiro, Milinovich, and Soundararajan (2019) on large prime gaps assuming the Generalized Riemann Hypothesis (GRH), we can prove

Theorem

Assuming GRH, if $m = q^k$ for a prime q and exponent $k \geq 3$, then $M(mL(m)) \neq S(mL(m))$.

Conjecture

Conjecture

$$S(mL(m)) = M(mL(m))$$

if and only if m is prime or $m \in \{4, 9, 121\}$.

The evidence for the conjecture:

- If m is prime, $S(mL(m)) = M(mL(m))$
- If m is not a prime power, $S(mL(m)) \neq M(mL(m))$.
- If m is a large enough prime power with exponent at least 3, $S(mL(m)) \neq M(mL(m))$. (Using GRH, we can remove "large enough")
- The evidence that no other prime squares work is that we've checked up to 1000 and Cramer's heuristics imply it for large enough p^2.

Numerics

Using a computer, we can verify that if $N \leq 1000000$, then $M(N) = S(N)$ for

$$\begin{aligned} 1 &\leq N \leq 18 \\ 31 &\leq N \leq 48 \\ 85 &\leq N \leq 300 \\ 571 &\leq N \leq 2940 \\ 18481 &\leq N \leq 22680 \\ 54181 &\leq N \leq 304920 \end{aligned}$$

Density Question

- Let A be the set of all N for which $S(N) = M(N)$.
- Let $A(x)$ be the set of all $N \leq x$ for which $S(N) = M(N)$.

Does $\lim_{x\to\infty} \frac{A(x)}{x}$ exist?

Upper and lower density definitions

The upper density of a set of natural numbers A is

$$\delta^{+} = \limsup_{x\to\infty} \frac{A(x)}{x}.$$

The lower density is

$$\delta^{-} = \liminf_{x\to\infty} \frac{A(x)}{x}.$$

Our theorems on upper and lower density

Theorem

$$\delta^+ = 1.$$

$$\delta^- = 0.$$

Therefore $\lim_{x\to\infty} \frac{A(x)}{x}$ does not exist.

What was needed for these proofs?

- For δ^+ the idea is as follows. Fix an integer k. If $p > q$ are consecutive primes with no prime powers in between and $p - q \geq k$. Then there is a large interval that contains elements of $A(pL(p))$. In fact this interval is of size at least $\left(1 - \frac{1}{2k}\right) A(pL(p))$ for large enough p. Therefore

$$\delta^+ \geq 1 - \frac{1}{2k}.$$

- By the Prime Number Theorem, the average distance between two primes grows logarithmically, so for any fixed integer k, there are infinitely many primes q satisfying that the next prime p is at least k numbers away. Therefore, we can let $k \to \infty$ to conclude $\delta^+ = 1$.

What was needed for these proofs? II

- For δ^- the idea is as follows. If $p > q$ are consecutive primes with no prime powers in between and $p - q \leq k$. Then there is a large interval that contains elements not in $A(pL(p)/(2k))$. In fact this interval is essentially the size of $A(pL(p)/(2k))$ for large enough p.
- By recent achievements in primes in small gaps by Zhang, Maynard, Tao, and the Polymath group, we know there are infinitely many primes $p > q$ with $p - q \leq 246$. Therefore we can take $k = 246$ and confirm that $\delta^- = 0$.
- There's a small subtlety regarding needing the number of primes $x \geq p > q$ with $p - q \leq 246$ to be bounded below by $\frac{Cx}{\log^{50}(x)} > \sqrt{x}$.

Thank you