# ON SETS WHOSE SUBSETS HAVE INTEGER MEAN II

**Enrique Treviño**
*Department of Mathematics, Lake Forest College, Lake Forest, IL, 60045, USA*
trevino@lakeforest.edu

*For Carl Pomerance on his 80th birthday*

*Received: , Revised: , Accepted: , Published:*

**Abstract**

We call a finite set of positive integers balanced if all its subsets have integer mean. For a positive integer $N$, let $M(N)$ be the cardinality of the largest balanced set all of whose elements are less than or equal to $N$, and let $S(N)$ be the cardinality of the balanced set with elements less than or equal to $N$ that has maximal sum. It is known that the set $T$ of $N$ for which $M(N) = S(N)$ does not have an asymptotic density. In this paper we prove that $T$ has logarithmic density equal to 0.

## 1. Introduction

As in [5] we say a finite set $A$ of positive integers is balanced if, for any subset $B \subseteq A$, the arithmetic mean of the elements of $B$ is an integer, i.e.,

$$\frac{1}{|B|}\sum_{i\in B} i \in \mathbb{Z}.$$

Inspired by problem 2 of the Mexican Mathematical Olympiad 2017, we consider balanced sets of positive integers bounded above by some integer $N$. For a positive integer $N$, let $M(N)$ be the size of the largest balanced set all of whose elements are at most $N$, and $S(N)$ be the size of the set with maximal sum among balanced sets all of whose elements are at most $N$, with the understanding that if two or more balanced sets achieve the maximal sum, we take the set with most elements[1]. As mentioned in [5], among the first million positive integers $N$, we have $M(N) = S(N)$ for the following values of $N$.

[1]In [5], we didn't realize that $S(N)$ wasn't well-defined because there exist $N$'s for which there are balanced sets of different sizes achieving the maximal sum.

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

It is hard to characterize the values of $N$ on which $M(N) = S(N)$, however, we have some results. For example, if we define $L(n) = \text{lcm}\{1, 2, \ldots, n\}$, Lemma 5 on [5] characterizes when $M(N) = S(N)$ for $qL(q) < N \le pL(p)$ when $q < p$ are large enough consecutive primes for which there is no prime power between $q$ and $p$. The lemma states:

**Lemma 1.** *Suppose $q < p$ are consecutive primes for which there is no prime power in the interval $(q, p)$. Let $k = p - q$.*

1. *For $N$ such that*
$$pL(p) \ge N > \max\left\{\left(p - \frac{1}{2}\right) L(p-1), \frac{L(p)(p^2 - p - q + 1)}{2pk}\right\},$$
*we have $M(N) = S(N)$.*

2. *If*
$$\frac{L(p)(p^2 - p - q + 1)}{2pk} > \left(p - \frac{1}{2}\right) L(p-1),$$
*for $N$ satisfying*
$$qL(q) + 1 \le N < \frac{L(p)(p^2 - p - q + 1)}{2pk},$$
*we have $M(N) \ne S(N)$.*

This lemma was used to show that the upper density for the set $T$ of all values $N$ such that $M(N) = S(N)$ is 1, while the lower density of $T$ is 0.

We will simplify Lemma 1 to the following:

**Lemma 2.** *Suppose $q < p$ are consecutive primes for which there is no prime power in the interval $(q, p)$. Let $k = p - q$. If $qL(q) < N \le pL(p)$, then $M(N) = S(N)$ if and only if*
$$\frac{L(p)(p^2 - p - q + 1)}{2pk} \le N \le pL(p).$$
*Furthermore, when $qL(q) < N < \frac{L(p)(p^2-p-q+1)}{2pk}$, $1 \le M(N) - S(N) \le k$, with all integer values in $[1, k]$ realized for some $N$.*

When these results were presented at Integers 2025, Alexander Borisov asked if the logarithmic density exists, and Tristan Phillips asked if there's a bound for the difference between $M(N)$ and $S(N)$. Tristan Phillips' question can be answered in the negative from Lemma 2. The answer to Borisov's question is our main theorem:

**Theorem 1.** *The logarithmic density for $T = \{N \in \mathbb{N} \mid M(N) = S(N)\}$ is 0. That is,*

$$\lim_{x\to\infty} \frac{1}{\log x} \sum_{\substack{N \le x \\ M(N)=S(N)}} \frac{1}{N} = 0.$$

The proof requires strong theorems about the distribution of prime gaps, in particular we need the sum of consecutive prime gaps to be small. We use the following result of Peck [3] improving on results of Heath-Brown [1, 2] (Stadlmann [4] made further improvements but her results have not been published yet).

**Theorem A.** *Let $p_1, p_2, \ldots$ be the list of primes in order. Let $x$ be a positive real number. Then*

$$\sum_{3 \le p_n \le x} (p_n - p_{n-1})^2 \ll_\epsilon x^{\frac{5}{4}+\epsilon}.$$

In Section 2, we will prove some lemmas about the distribution of gaps of consecutive primes we will need to prove Theorem 1, and we will prove a cleaner version of Lemma 1. In Section 3 we will prove Theorem 1.

### Notation

For the rest of the paper, we will use $L(n) = \operatorname{lcm}\{1, 2, \ldots, n\}$. As a reminder $f(x) \ll g(x)$ means that there exists a positive constant $C$ such that $|f(x)| \le C|g(x)|$. We also say that $f(x) = o(g(x))$ if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0.$$

## 2. Results on gaps of consecutive primes and other useful lemmas

The following lemma, which can be of independent interest, can be thought of as bounding the sum of the largest $z$ consecutive prime gaps for some value of $z$ (in this case for $z \ll \sqrt{x}$).

**Lemma 3.** *Let $p_1, p_2, \ldots$ be the list of primes in order. Let $x$ be a positive real. We say $p_i \in B$ if there is a prime power in the interval from $p_{i-1}$ to $p_i$.*

$$\sum_{\substack{3 \le p_n \le x \\ p_n \in B}} (p_n - p_{n-1}) \ll x^{\frac{8}{9}}.$$

*Proof.* Note that the number of elements in $B$ is bounded above by the number of prime powers less than or equal to $x$. Therefore $|B| \ll \sqrt{x}$. Let $\epsilon > 0$. From Theorem A and Cauchy-Schwarz,

$$\begin{aligned}\sum_{\substack{3 \le p_n \le x \\ p_n \in B}} (p_n - p_{n-1}) &\le \sqrt{|B| \cdot \sum_{3 \le p_n \le x} (p_n - p_{n-1})^2} \\ &\ll_\epsilon \sqrt{x^{\frac{1}{2}} x^{\frac{5}{4}+\epsilon}} = x^{\frac{7}{8}+\frac{\epsilon}{2}} \ll x^{\frac{8}{9}}.\end{aligned}$$

□

The following lemma gives a simple bound for the sum of the logarithm of consecutive prime gaps.

**Lemma 4.** *Let $p_1, p_2, \ldots$ be the list of primes in order. Let $x > 0$ be a real number. Then*

$$\sum_{3 \le p_n \le x} \log (p_n - p_{n-1}) \ll \frac{x \log \log x}{\log x}.$$

*Proof.* Let $N = \pi(x)$. By AM-GM

$$\begin{aligned}\sum_{3 \le p_n \le x} \log (p_n - p_{n-1}) &= \log \left( \prod_{3 \le p_n \le x} (p_n - p_{n-1}) \right) \\ &\le \log \left( \left( \frac{\sum_{3 \le p_n \le x} (p_n - p_{n-1})}{N-1} \right)^{N-1} \right) \\ &= (N-1) \log \left( \frac{p_N - 2}{N-1} \right).\end{aligned}$$

From the Prime Number Theorem, we have $N = \frac{x}{\log x} (1 + o(1))$, hence

$$\sum_{3 \le p_n \le x} \log (p_n - p_{n-1}) \ll \frac{x}{\log x} \log \left( \frac{x}{\frac{x}{\log x}} \right) \ll \frac{x \log \log x}{\log x}.$$

□

The following lemma will allow us to simplify Lemma 1.

**Lemma 5.** *Suppose $5 \leq q < p$ are consecutive primes. Let $k = p - q$. Then*

$$\frac{L(p)(p^2 - p - q + 1)}{2pk} > \left(p - \frac{1}{2}\right) L(p-1).$$

*Proof.* First note that $L(p-1) = \frac{L(p)}{p}$. Therefore, the inequality reduces to trying to show

$$\frac{p^2 - p - q + 1}{2k} > p - \frac{1}{2}.$$

We want to show

$$k < \frac{p^2 - p - q + 1}{2p - 1} = \frac{p^2 - 2p + k + 1}{2p - 1}.$$

Therefore, we want to show

$$k < \frac{p-1}{2}.$$

From Bertrand's postulate, for $n > 3$ there is always a prime $r$ such that $n < r < 2n - 2$, so when $p > 5$, we have $n = (p+1)/2 > 3$, which implies there is a prime $r$ such that $\frac{p+1}{2} < r < p - 1$. In particular, $\frac{p+1}{2} < q < p - 1$ and hence $k = p - q < \frac{p-1}{2}$. □

We can now prove Lemma 2

*Proof of Lemma 2.* The proof of the characterization follows almost directly from Lemma 1 together with Lemma 5. The one possible exception is the case $N = \frac{L(p)(p^2-p-q+1)}{2pk}$. However, in the proof of Lemma 5 in [5], the $>$ symbol can be changed to a $\geq$ symbol once the definition of $S(N)$ is done correctly to account for the fact that multiple balanced sets could achieve the maximal sum, and that in the case of a tie in the sum, we take the set with more elements.

For the result about the difference between $M(N)$ and $S(N)$, it follows from Lemmas 2 and 5 in [5], where for $(m-1)L(m-1)+1 \leq N \leq mL(m)$ we have $M(N) = m$, while for $qL(q) + 1 \leq N < \frac{L(p)(p^2-p-q+1)}{2pk}$ we have $S(N) = q$. Therefore, in the range where $S(N) = q$ we have different intervals where $M(N) = q+1, q+2, \ldots, p$, namely

$$(qL(q), (q+1)L(q)], ((q+1)L(q), (q+2)L(q)], \ldots, \\ ((q+k-1)L(q), \frac{L(p)(p^2 - p - q + 1)}{2pk}),$$

where we used that $L(q+i) = L(q)$ for $0 \leq i < k$ and the last interval is not empty because of Lemma 5. □

## 3. Proof of the main theorem

Now we are ready to prove our theorem about logarithmic density.

*Proof of Theorem 1.* Let $x$ be a large real number. Suppose $p$ is the largest prime such that $pL(p) \leq x$. From the Prime Number Theorem, we have $\log L(p) = p + o(p)$. It follows that $p \leq (1+o(1))\log x$.

Suppose $q < r$ are consecutive primes such that there is no prime power between $q$ and $r$. Let $k = r - q$. Then from Lemma 2,

$$\sum_{\substack{qL(q)<N\leq rL(r)\\ M(N)=S(N)}} \frac{1}{N} = \sum_{\frac{L(r)(r^2-r-q+1)}{2rk}\leq N\leq rL(r)} \frac{1}{N}$$
$$\ll \log\left(\frac{rL(r)2rk}{L(r)(r^2-r-q-1)}\right) \ll \log k. \quad (1)$$

On the other hand, if $q < r$ are consecutive primes for which there is a prime power between $q$ and $r$, we have

$$\sum_{\substack{qL(q)<N\leq rL(r)\\ M(N)=S(N)}} \frac{1}{N} \leq \sum_{qL(q)<N\leq rL(r)} \frac{1}{N}$$
$$\ll \log\left(\frac{rL(r)}{qL(q)}\right) = \log(L(r)) - \log(L(q)) + \log\left(\frac{r}{q}\right) \ll r - q. \quad (2)$$

Let $p_1, p_2, \ldots$ be the list of primes in order with $p_n = p$. For notational purposes, let $p_0 = 1$ and $\mathcal{P}$ be the set of prime powers. From (1) and (2), we get

$$\sum_{\substack{N\leq x\\ M(N)=S(N)}} \frac{1}{N} \leq 1 + \sum_{i=1}^{n+1} \sum_{\substack{p_{i-1}L(p_{i-1})<N\leq p_iL(p_i)\\ M(N)=S(N)}} \frac{1}{N}$$
$$= 1 + \sum_{\substack{1\leq i\leq n+1\\ (p_{i-1},p_i)\cap\mathcal{P}\neq\emptyset}} \sum_{\substack{p_{i-1}L(p_{i-1})<N\leq p_iL(p_i)\\ M(N)=S(N)}} \frac{1}{N} + \sum_{\substack{1\leq i\leq n+1\\ (p_{i-1},p_i)\cap\mathcal{P}=\emptyset}} \sum_{\substack{p_{i-1}L(p_{i-1})<N\leq p_iL(p_i)\\ M(N)=S(N)}} \frac{1}{N}$$
$$\ll \sum_{\substack{2\leq i\leq n+1\\ (p_{i-1},p_i)\cap\mathcal{P}\neq\emptyset}} (p_i - p_{i-1}) + \sum_{\substack{2\leq i\leq n+1\\ (p_{i-1},p_i)\cap\mathcal{P}=\emptyset}} \log(p_i - p_{i-1}).$$

Note that $p_n \ll \log x$, so $p_{n+1} \ll \log x$. Now, applying Lemmas 4 and 3, we get

$$\sum_{\substack{N\leq x\\ M(N)=S(N)}} \frac{1}{N} \ll \sum_{\substack{2\leq i\leq n+1\\ (p_{i-1},p_i)\cap\mathcal{P}\neq\emptyset}} (p_i - p_{i-1}) + \sum_{\substack{2\leq i\leq n+1\\ (p_{i-1},p_i)\cap\mathcal{P}=\emptyset}} \log(p_i - p_{i-1})$$
$$\ll (\log x)^{\frac{8}{9}} + \frac{\log x \log\log\log x}{\log\log x} = o(\log x).$$

Therefore

$$\lim_{x\to\infty} \frac{1}{\log x} \sum_{\substack{N\le x\\ M(N)=S(N)}} \frac{1}{N} = 0.$$

□

**Acknowledgements.** I would like to thank Paul Pollack for some very helpful conversations about the problem. I would also like to thank Alexander Borisov and Tristan Phillips for the inspiring questions they asked when I presented the results of [5] at Integers 2025. I am grateful to the anonymous referee that made important corrections to the paper and gave important suggestions that improved the quality of the paper. This paper is dedicated to Carl Pomerance celebrating his 80th birthday. Carl's influence on my career has been enormous, and I am very grateful for his kind, supportive, and exceptional mentorship.

## References

[1] D. R. Heath-Brown, The differences between consecutive primes, *J. London Math. Soc. (2)* **18** (1978), no. 1, 7–13.

[2] ———, The differences between consecutive primes. III, *J. London Math. Soc. (2)* **20** (1979), no. 2, 177–178.

[3] A. S. Peck, Differences between consecutive primes, *Proc. London Math. Soc. (3)* **76** (1998), no. 1, 33–69.

[4] Julia Stadlmann, On the mean square gap between primes, preprint `arXiv: 2212.10867`.

[5] Enrique Treviño, On sets whose subsets have integer mean, *Integers* **21** (2021), #A79.