RECIPROCAL SUMS AND COUNTING FUNCTIONS

ALEX RICE

ABSTRACT. Motivated by the gentle exploration of the distribution of prime numbers typical of an undergraduate number theory course, as well as by a recent breakthrough result in arithmetic combinatorics, we explore connections between the counting function $\mathcal{C}_A(x) = |A \cap [1, x]|$ and the reciprocal sum function $S_A(x) = \sum_{n \in A \cap [1,x]} 1/n$ for a set $A \subseteq \mathbb{N}$.

1. Introduction

One of the most ubiquitous proofs of the infinitude of the prime numbers (perhaps second only to Euclid's proof) is due to Euler, in which the divergence of the harmonic series $\sum_{n=1}^{\infty} 1/n$ is shown to imply the divergence of the series $\sum_p 1/p$, where the latter sum is taken over primes. In fact, a more quantitative, but nearly as elementary, version of this proof, using the bound $\sum_{n \le x} 1/n > \log x$, yields the lower bound

$$\sum_{p \le x} \frac{1}{p} > \log\log(x) - 1 \tag{1}$$

for all $x > 1$, which turns out to be a rather accurate estimate for the partial reciprocal sums over the primes. Here and throughout this paper, we use log to denote the natural logarithm, the summation limit $n \le x$ refers to the natural numbers in that range, while $p \le x$ refers to the primes in that range. We take the convention that the set of natural numbers, $\mathbb{N}$, starts at 1.

While (1) is quite satisfying, more detailed investigation into the distribution of prime numbers typically involves estimates on the prime counting function $\pi(x)$, the number of primes that are at most x. The celebrated prime number theorem says that $\pi(x)$ is well-approximated by $x/\log x$, and more accessible theorems of Chebyshev provide upper and lower bounds of this order of magnitude, but all require considerably more effort than (1). Since the derivation of (1) (included later) is so low-tech, it is tempting to inquire as to what information about $\pi(x)$, aside from its tendency toward infinity, can be gleaned from (1) and (1) alone.

Skipping ahead some centuries, Bloom and Sisask [3] recently made headlines by settling the first nontrivial case of a longstanding conjecture of Erdős. They showed that if $A \subseteq \mathbb{N}$ and $\sum_{n \in A} 1/n$ diverges, then A contains a nontrivial three-term arithmetic progression (3AP), in other words a set of the form $\{n, n+d, n+2d\}$ for some $n, d \in \mathbb{N}$. Presented with this formulation out of context, it is natural to wonder how the divergence of the reciprocal sum over the elements of A is directly utilized in the proof. In fact, the answer is not at all, and the result stated in this form is a bit misleading. What is actually achieved is strictly stronger, a highly-anticipated upper bound on the counting function of a set lacking 3APs.

Framed by these two motivating contexts, one classical and one cutting edge, we investigate the relationship between counting functions and reciprocal sums for sets of positive integers. We consider the extent to which an estimate on one of these functions informs the other, and we construct extreme examples to test the boundaries of our observations. Reciprocal sums as descriptors for sets of positive integers were explored previously by Bayless and Klyve in a 2013 Monthly article [1], but that article focuses on estimates for specific, convergent reciprocal sums, and has little overlap with our discussions here.

2. Preliminaries

We begin by introducing our key players.

Definition 1. For a set $A \subseteq \mathbb{N}$ and $x \geq 1$, we define the counting function $\mathcal{C}_A(x) = |A \cap [1, x]|$ and the reciprocal sum function $S_A(x) = \sum_{n \in A \cap [1,x]} 1/n$. Here and throughout we use $|X|$ to denote the number of elements of a finite set X. Further, when making shorthand asymptotic statements involving these or related functions, such as $S_A(x) \to \infty$ or $C_A(x)/x \to 0$, we refer to the behavior as x tends to infinity.

All proofs throughout the paper require only basic calculus and make frequent use of the following two facts:

- **Harmonic Series Estimate:** For all $x \geq 1$ we have

$$\log x < \sum_{n \leq x} \frac{1}{n} \leq \log x + 1. \tag{2}$$

  Better estimates for large x are possible, but these are the bounds yielded by the most straightforward method of drawing $1 \times 1/n$ rectangles and comparing the resulting area to the area under the graph of $1/t$ for $1 \leq t \leq x$.

- **Partial Summation:** If $\{b_n\}_{n \in \mathbb{N}}$ is a sequence of real numbers and $\psi : [1, \infty) \to \mathbb{R}$ is continuously differentiable, then

$$\sum_{n \leq x} b_n \psi(n) = B(x)\psi(x) - \int_1^x B(t)\psi'(t)dt,$$

  where $B(x) = \sum_{n \leq x} b_n$. This formula is a convenient discrete-continuous hybrid analog of integration by parts, and can be obtained by writing b_n as $B(n) - B(n-1)$, writing $\psi(n+1) - \psi(n)$ as $\int_n^{n+1} \psi'(t)dt$, and noting that $B(n) = B(t)$ for $n \leq t < n+1$. We will exclusively apply this formula in the case that b_n is the indicator function of a fixed set $A \subseteq \mathbb{N}$ (meaning 1 if $n \in A$ and 0 if $n \notin A$) and $\psi(x) = 1/x$. In this case, we have

$$S_A(x) = \frac{\mathcal{C}_A(x)}{x} + \int_1^x \frac{\mathcal{C}_A(t)}{t^2} dt. \tag{3}$$

3. A Quick, Accurate "Lower Bound" for $\pi(x)$

As promised, in order to keep our discussion complete and self-contained, we include a derivation of (1), the product of an only slightly more careful version of Euler's proof of the infinitude of the primes. This serves as a remarkably obstacle-free path to a quite accurate lower bound on the reciprocal sum over the primes, common and ideal for an undergraduate number theory course. For example, the proof below is very similar to one found in Chapter 1, Section 4 of [12].

Proposition 1. $\sum_{p \leq x} 1/p > \log\log(x) - 1$ *for all* $x > 1$.

Proof. Fix $x > 1$. For each prime $p \leq x$, let $s_p(x) = 1 + 1/p + 1/p^2 + \cdots + 1/p^j$, where p^j is the largest power of p that is at most x. By the distributive property, and the fact that each natural number $n \leq x$ has a factorization into prime powers (we do not even need that such a factorization is unique), the term $1/n$ for each $n \leq x$ appears in the expanded product $\prod_{p \leq x} s_p(x)$, so this product exceeds $\sum_{n \leq x} 1/n$. By the geometric series formula, $s_p(x) \leq \sum_{k=0}^{\infty} p^{-k} = 1 + 1/(p-1)$. Further, we have

$$1 + t = 1 + \int_0^t du \leq 1 + \int_0^t e^u du = e^t$$

for all $t \geq 0$. Writing e^t as $\exp(t)$ when convenient, these facts combine with (2) to yield

$$\log x < \sum_{n \leq x} \frac{1}{n} \leq \prod_{p \leq x} s_p(x) \leq \prod_{p \leq x} \left(1 + \frac{1}{p-1}\right) \leq \prod_{p \leq x} \exp(1/(p-1)) = \exp\left(\sum_{p \leq x} \frac{1}{p-1}\right),$$

so in particular

$$\sum_{p \leq x} \frac{1}{p-1} > \log\log x. \tag{4}$$

Finally, we use a telescoping series to observe that

$$\sum_{p\le x} \frac{1}{p-1} - \frac{1}{p} \le \sum_{2\le n\le x} \frac{1}{n-1} - \frac{1}{n} < \sum_{n=2}^{\infty} \frac{1}{n-1} - \frac{1}{n} = 1,$$

which when subtracted from (4) completes the proof. □

Less standard is the following application of (3), which says that any set $A \subseteq \mathbb{N}$ satisfying something like (1), not necessarily the primes, must contain nearly $n/\log n$ of the first n natural numbers infinitely often.

Proposition 2. *If $A \subseteq \mathbb{N}$ and $S_A(x) > \log\log(x) - C$ for a constant C and all $x > 1$, then for every $\epsilon > 0$, $\mathcal{C}_A(n) \ge (1-\epsilon)n/\log n$ for infinitely many $n \in \mathbb{N}$.*

Proof. Fix $A \subseteq \mathbb{N}$. We proceed contrapositively by assuming that there exists $\epsilon > 0$ and $n_0 \in \mathbb{N}$ such that $\mathcal{C}_A(n) \le (1-\epsilon)n/\log n$ for all integers $n \ge n_0$. Further, since $x/\log x$ is increasing and $\mathcal{C}_A(x) = \mathcal{C}_A(\lfloor x \rfloor)$, we in fact have $\mathcal{C}_A(x) \le (1-\epsilon)x/\log x$ for all $x \ge n_0$. Therefore, by (3) and the trivial bound $\mathcal{C}_A(t) \le t$, we have

$$S_A(x) \le 1 + \int_1^{n_0} \frac{1}{t} dt + \int_{n_0}^{x} \frac{(1-\epsilon)}{t\log t} dt \le 1 + \log(n_0) + (1-\epsilon)\log\log x$$

for all $x \ge n_0$. In particular, $S_A(x) < \log\log(x) - C$ provided $\epsilon \log\log x > C + 1 + \log n_0$, which holds for $x > \exp(\exp((C+1+\log n_0)/\epsilon))$. □

While the nonuniformity in the lower bound provided by Proposition 2 certainly makes it much weaker than a global statement of the form $\pi(x) \ge \frac{x}{\log x}(1+o(1))$, it does say something of definitive note: if there is to be a nice global approximation of $\pi(x)$, it cannot be asymptotically smaller than $x/\log x$. This is perhaps the fastest, most elementary path to a statement of this strength about $\pi(x)$ with the correct leading term, and it makes for a fruitful follow-up to Proposition 1 in an undergraduate course, either in class or as an exercise, as an alternative to more intimidating endeavors.

4. Reciprocal Sums and Arithmetic Progressions

In 1936, Erdős and Turan [5] conjectured that if $A \subseteq \mathbb{N}$ contains no 3APs, then $\mathcal{C}_A(x)/x \to 0$ (in other words A has *density* 0), a result proven by Roth [13], now known as Roth's Theorem (not be confused with the Diophantine approximation result of the same name; but it is indeed the same Roth, he was amazing!). The conjecture was generalized to replace 3AP with kAP for any fixed $k \in \mathbb{N}$, a result proven by Szemerédi [14], now known as Szemerédi's Theorem. Since any set satisfying $\sum_{n\in A} 1/n < \infty$ also satisfies $\mathcal{C}_A(x)/x \to 0$ (see Proposition 7), the following conjecture of Erdős calls for a strengthening of Szemerédi's Theorem.

Conjecture 1 (Erdős)**.** *If $A \subseteq \mathbb{N}$ contains no k-term arithmetic progression for some $k \in \mathbb{N}$, then $\sum_{n\in A} 1/n$ converges. Equivalently, if $\sum_{n\in A} 1/n$ diverges, then A contains arbitrarily long arithmetic progressions.*

Remark. Conjecture 1 is sometimes colloquially referred to as the Erdős-Turán Conjecture, despite it originating from Erdős alone, owing to its connection with the previous work of Erdős and Turán, and the fact that Erdős offered a prize for its resolution in memoriam after Turán's death in 1976. For a detailed, well-referenced discussion of this discrepancy in nomenclature, the interested reader may refer to [11].

In the intervening decades, an extensive literature, central to the field of arithmetic combinatorics, has developed on sets without arithmetic progressions. The aforementioned breakthrough of Bloom and Sisask on 3APs, which we state properly below, is the first resolution of a nontrivial case of Conjecture 1.

Theorem 1 (Bloom-Sisask, [3])**.** *There exists $C > 0$ and $\epsilon > 0$ such that if $A \subseteq \mathbb{N}$ contains no three-term arithmetic progression, then*

$$\mathcal{C}_A(x) \le \frac{Cx}{(\log x)^{1+\epsilon}} \quad \textit{for all } x > 1. \tag{5}$$

A pleasing consequence of Theorem 1 is that the primes (which have counting function $\pi(x) \approx x/\log x$), and relatively dense subsets thereof, contain 3APs. In a qualitative sense, this was already known to be true for kAPs for all k (due to Green [7] for $k = 3$ and Green and Tao [9] for $k > 3$), but Theorem 1 assures that the primes contain 3APs for density reasons alone, independent of any other properties. The connection between Theorem 1 and Conjecture 1 comes from (3) and the fact that the improper integral $\int_2^\infty \frac{1}{t(\log t)^{1+\epsilon}}\, dt$ happens to converge, while without the ϵ it diverges, both observable with the substitution $u = \log t$. Therefore, if $A \subseteq \mathbb{N}$ contains no 3AP, then A satisfies (5), hence $\sum_{n\in A} 1/n$ converges by (3). Contrapositively, if $\sum_{n\in A} 1/n$ diverges, then A must contain a 3AP.

However, reciprocal sums are not utilized in the proof of Theorem 1, and it is not known if the "true" maximum growth rate of $\mathcal{C}_A(x)$ for a set lacking 3APs (or kAPs for larger k) is anywhere near this "breaking point" of around $x/\log x$ for convergent versus divergent reciprocal sums. In fact, the best constructions of large sets lacking 3APs, dating back to Behrend [2], satisfy $\mathcal{C}_A(x) \geq x\exp(-C\sqrt{\log x})$ for a constant C. This lower bound, while impressively large (asymptotically greater than $x^{1-\epsilon}$ for any fixed $\epsilon > 0$), is still much smaller than the right hand side of (5). Meanwhile, the best-known upper bounds for sets lacking kAPs for $k > 3$ take the form $\mathcal{C}_A(x) \leq Cx/(\log x)^c$ for $k = 4$ (due to Green and Tao [8]), where $C, c > 0$ are constants, and $\mathcal{C}_A(x) \leq Cx/(\log\log x)^{c_k}$, where $c_k > 0$ depends on k, for $k > 4$ (due to Gowers [6]). All of this is to say: the statement of Conjecture 1 is extremely elegant and attractive, but is likely a red herring of sorts. By (3), an upper bound of the form $\mathcal{C}_A \leq \mathcal{C}$ for sets lacking kAPs, where $\mathcal{C}$ is a fixed function such that $\int_1^\infty \mathcal{C}(t)/t^2 dt$ converges, is strictly stronger than Conjecture 1. It is this sort of counting function upper bound to which the field aspires, as opposed to Conjecture 1 itself.

5. General Bounds and Extreme Constructions

In Section 4, we observed an example of how, via the partial summation formula (3), global information about the counting function $\mathcal{C}_A$ immediately yields global information about the reciprocal sum function S_A, for a set $A \subseteq \mathbb{N}$. Conversely, in Section 3, we considered a special case where global information about S_A yielded a weaker, but still notable conclusion about $\mathcal{C}_A$. In this section we explore the latter type of phenomenon in a greater level of generality.

Definition 2. Suppose $f : [1,\infty) \to [0,\infty)$ and $A \subseteq \mathbb{N}$. We say that f is a *lim sup lower bound* for $\mathcal{C}_A$ if, for every $\epsilon > 0$, $\mathcal{C}_A(x_k) \geq (1-\epsilon)f(x_k)$ for some sequence $x_k \to \infty$. This latter condition in particular holds if $\mathcal{C}_A(n) \geq (1-\epsilon)f(n)$ for infinitely many $n \in \mathbb{N}$, and in fact these formulations are equivalent if f is increasing, since $\mathcal{C}_A(x) = \mathcal{C}_A(\lfloor x \rfloor)$. For those familiar with the concept of lim sup from analysis, this definition is equivalent to $\limsup_{x\to\infty} \mathcal{C}_A(x)/f(x) \geq 1$. We define *lim inf upper bound* analogously, with $\mathcal{C}_A(x_k) \geq (1-\epsilon)f(x_k)$ replaced by $\mathcal{C}_A(x_k) \leq (1+\epsilon)f(x_k)$.

For two real-valued functions f, g, we write $f \leq g$ if $f(x) \leq g(x)$ for all x in their common domain. We refer to such bounds as *global*. Roughly speaking, one can think of a lim sup lower (resp. upper) bound as a statement about the *failure* of global upper (resp. lower) bounds. More specifically, if $A \subseteq \mathbb{N}$ and f is a lim sup lower bound for $\mathcal{C}_A$, then no global upper bound of the form $\mathcal{C}_A \leq g$ is possible with g asymptotically smaller than f. We summarize the results that follow below:

(i) As in Proposition 2, global lower bounds (resp. upper bounds) of the form $S_A \geq S$ (resp. $S_A \leq S$), with $S(x) \to \infty$, imply lim sup lower bounds (resp. lim inf upper bounds) for $\mathcal{C}_A$ of the "correct" order of magnitude.

(ii) Global bounds for S_A imply global bounds for $\mathcal{C}_A$, but of much different magnitude than the lim sup lower bounds and lim inf upper bounds referred to in the previous item. Further, constructions show that these "weak" global bounds are essentially sharp and cannot be noticeably improved in general.

(iii) If $\sum_{n\in A} 1/n$ converges, then $\mathcal{C}_A(x)/x \to 0$. However, no stronger global upper bound on $\mathcal{C}_A$ can be deduced from the reciprocal sum, in the sense that any function f satisfying $f(x)/x \to 0$ can be a lim sup lower bound for $\mathcal{C}_A$, with $\sum_{n\in A} 1/n$ arbitrarily small.

We begin by considering the degree to which the argument in the proof of Proposition 2 was specific to the functions $S(x) = \log\log x - C$ and $f(x) = x/\log x$. In fact, the only properties of these functions used in the proof is the tendency of $S(x)$ toward infinity, and the relationship $f(x) = x^2 S'(x)$, which allowed for $S(x)$ to pop up in the integral in the right hand side of (3). We leave it as an exercise to adapt the steps of the proof of Proposition 2 to obtain the following generalization, which accounts for item (i) on our list.

Proposition 3. *Suppose $S : [1,\infty) \to [0,\infty)$ is continuously differentiable with $S(x) \to \infty$, and let $f(x) = x^2 S'(x)$. If $A \subseteq \mathbb{N}$ satisfies $S_A \geq S$ (resp. $S_A \leq S$), then f is a lim sup lower bound (resp. lim inf upper bound) for $\mathcal{C}_A$.*

Remark. In Proposition 3 and several that follow, we assume $S(x)$ is defined and nonnegative for $x \geq 1$. In practice, we may want to apply these propositions for functions like $S(x) = \log\log x - C$ for a constant C, which is only defined and nonnegative for $x \geq \exp(\exp(C))$. In these situations, the propositions still apply perfectly well; one can simply redefine $S(x)$ to be 0 until the desired formula can take over.

As we discussed following Proposition 2, the bounds provided by Proposition 3 are pleasingly sharp, but they do not give global information about the counting function. In the following, we consider the extent to which one *can* deduce global information about $\mathcal{C}_A$ from estimates on S_A, which accounts for the first portion of item (ii) on our list.

Proposition 4. *If $A \subseteq \mathbb{N}$ and $x \geq 1$ then $\exp(S_A(x) - 1) \leq \mathcal{C}_A(x) \leq \left(1 - \frac{1}{\exp(S_A(x)+1)}\right) x$.*

Proof. Fix $A \subseteq \mathbb{N}$ and $x \geq 1$. Let $n_1 < n_2 < \cdots < n_{\mathcal{C}_A(x)}$ be the elements of $A \cap [1,x]$, so in particular $n_k \geq k$ for all $1 \leq k \leq \mathcal{C}_A(x)$. Therefore,

$$S_A(x) = \sum_{k=1}^{\mathcal{C}_A(x)} \frac{1}{n_k} \leq \sum_{k=1}^{\mathcal{C}_A(x)} \frac{1}{k} \leq \log(\mathcal{C}_A(x)) + 1,$$

and hence $\mathcal{C}_A(x) \geq \exp(S_A(x) - 1)$. For the other extreme, we consider the complement $\overline{A} = \mathbb{N} \setminus A$, which satisfies $S_{\overline{A}}(x) = S_{\mathbb{N}}(x) - S_A(x)$. Applying the argument above to $\overline{A}$ yields

$$\mathcal{C}_{\overline{A}}(x) \geq \exp(S_{\mathbb{N}}(x) - S_A(x) - 1) > \exp(\log(x) - S_A(x) - 1) = \frac{x}{\exp(S_A(x)+1)}.$$

The claimed upper bound then follows from the fact that $\mathcal{C}_A(x) \leq x - \mathcal{C}_{\overline{A}}(x)$. $\square$

The global bounds provided by Proposition 4 are seemingly weak. In the case that $S_A(x) \approx \log\log x$, as with the primes, the conclusion yields $c\log x \leq \mathcal{C}_A(x) \leq (1 - C/\log x)x$ for constants $C, c > 0$. Neither extreme is particularly close to $\pi(x) \approx x/\log x$, but the following two constructions show that no noticeable improvements to these global bounds are possible, as promised in the latter portion of item (ii).

Proposition 5. *Suppose $S : [1,\infty) \to [0,\infty)$ is an increasing function satisfying $S \leq S_{\mathbb{N}}$ and*

$$S(n+1) \leq S(n) + \frac{1}{n+1} \tag{6}$$

for all $n \in \mathbb{N}$. If $f : [1,\infty) \to [0,\infty)$ with $f(x) \to \infty$, then there exists $A \subseteq \mathbb{N}$ satisfying $S_A \geq S$ and $\mathcal{C}_A(n) \leq f(n)\exp(S(n))$ for infinitely many $n \in \mathbb{N}$.

Proof. Fix an increasing function $S : [1,\infty) \to [0,\infty)$ satisfying $S \leq S_{\mathbb{N}}$ and (6). Fix $f : [1,\infty) \to [0,\infty)$ tending to infinity. We leave it as an exercise that if $\exp(S(x))/x \not\to 0$, then $A = \mathbb{N}$ satisfies the conclusion of the proposition. For the remainder of the proof, we assume $\exp(S(x))/x \to 0$.

Let $n_0 = 1$ and choose an increasing sequence $\{n_k\}_{k\in\mathbb{N}}$ of natural numbers satisfying $f(n_k) \geq n_{k-1}$ and $(n_{k-1}+1)\exp(S(n_k)) < n_k$ for all $k \in \mathbb{N}$. Let $A_k = \mathbb{N} \cap [n_{k-1}, n_{k-1}\exp(S(n_k))]$, and let $A = \bigcup_{k=1}^{\infty} A_k$. We see that $\mathcal{C}_A(n_k) \leq n_{k-1}\exp(S(n_k)) + 1 \leq f(n_k)\exp(S(n_k))$. To see that $S_A \geq S$, fix $x \geq 1$. Note that we either have $n_{k-1}\exp(S(n_k)) < x < n_k$ or $n_{k-1} \leq x \leq n_{k-1}\exp(S(n_k))$ for some $k \in \mathbb{N}$. In the former case,

$$S_A(x) \geq \sum_{n \in \cup_{j=1}^{k} A_j} \frac{1}{n} \geq 1 + \sum_{n_{k-1} < n \leq n_{k-1}\exp(S(n_k))} \frac{1}{n} \geq \log(n_{k-1}\exp(S(n_k))) - \log(n_{k-1}) = S(n_k) \geq S(x).$$

In the latter case,

$$S_A(x) \geq \sum_{n \in \cup_{j=1}^{k-1} A_j} \frac{1}{n} + \sum_{n_{k-1} \leq n \leq x} \frac{1}{n} \geq S(n_{k-1}) + \sum_{n_{k-1} \leq n \leq x} \frac{1}{n} \geq S(\lfloor x \rfloor) + \frac{1}{n_{k-1}} \geq S(x). \qquad \square$$

Proposition 6. *Suppose $S : [1, \infty) \to [0, \infty)$ is increasing. If $f : [1, \infty) \to [0, \infty)$ with $f(x) \to \infty$, then there exists $A \subseteq \mathbb{N}$ satisfying $S_A \leq S$ and $\mathcal{C}_A(x) \geq n\left(1 - \frac{f(n)}{\exp(S(n))}\right)$ for infinitely many $n \in \mathbb{N}$.*

Proof. Fix an increasing function $S : [1, \infty) \to [0, \infty)$ and let $\tilde{S} = S_{\mathbb{N}} - S$. Note that since S is nonnegative, $\tilde{S}$ satisfies $\tilde{S} \leq S_{\mathbb{N}}$, and since S is increasing, $\tilde{S}$ satisfies (6). Therefore, we can apply Proposition 5 to produce a set $B \subseteq \mathbb{N}$ satisfying $S_B \geq \tilde{S}$ and $\mathcal{C}_B(n) \leq f(n) \exp(\tilde{S}(n))$ for infinitely many $n \in \mathbb{N}$. Let $A = \overline{B} = \mathbb{N} \setminus B$, so $S_B \geq \tilde{S}$ implies $S_A = S_{\mathbb{N}} - S_B \leq S_{\mathbb{N}} - \tilde{S} = S$. Further, $\mathcal{C}_B(n) \leq f(n) \exp(\tilde{S}(n))$ implies

$$\mathcal{C}_A(n) = n - \mathcal{C}_B(n) \geq n - f(n) \exp(S_{\mathbb{N}}(n) - S(n)) > n - f(n) \exp(\log n - S(n)) = n\left(1 - \frac{f(n)}{\exp(S(n))}\right)$$

for infinitely many $n \in \mathbb{N}$. $\square$

In Proposition 6, we see that if $S(x) \to \infty$, we can take f to be any function that grows more slowly than $\exp(S)$, say $f = \sqrt{\exp(S)}$, to ensure that the $1 - \frac{f(x)}{\exp(S(x))} \to 1$. In this case, Proposition 6 says, in particular, that $g(x) = x$ is a lim sup lower bound for $\mathcal{C}_A$ (or in other words A has *upper density* 1).

Corollary 1. *If $S : [1, \infty) \to [0, \infty)$ with $S(x) \to \infty$, then there exists $A \subseteq \mathbb{N}$ such that $S_A \leq S$ and, for every $\epsilon > 0$, $\mathcal{C}_A(n) \geq (1 - \epsilon)n$ for infinitely many $n \in \mathbb{N}$.*

The results discussed in this section thus far are either restricted to, or most meaningfully applied to, reciprocal sum functions tending to infinity. One can also inquire as to what information about a counting function can be ascertained from the fact that a reciprocal sum *converges*. This is addressed in the following proposition, which in particular clarifies why Conjecture 1 is a true strengthening of Szemerédi's Theorem.

Proposition 7. *If $A \subseteq \mathbb{N}$ and $\sum_{n \in A} 1/n < \infty$, then $\mathcal{C}_A(x)/x \to 0$.*

Proof. We approach the claim contrapositively, fixing $A \subseteq \mathbb{N}$ such that $\mathcal{C}_A(x)/x \not\to 0$. In other words, there exists $\delta > 0$ such that $\mathcal{C}_A(n) \geq \delta n$ for infinitely many $n \in \mathbb{N}$. Let $n_0 = 0$, and choose a sequence $\{n_k\}_{k \in \mathbb{N}}$ of natural numbers such that $n_k \geq 2n_{k-1}/\delta$ and $\mathcal{C}_A(n_k) \geq \delta n_k$ for all $k \in \mathbb{N}$.

Then,

$$\sum_{n \in A} \frac{1}{n} = \sum_{k=1}^{\infty} \sum_{n \in A \cap (n_{k-1}, n_k]} \frac{1}{n} \geq \sum_{k=1}^{\infty} \frac{\mathcal{C}_A(n_k) - n_{k-1}}{n_k} \geq \sum_{k=1}^{\infty} \frac{\delta n_k - \delta n_k/2}{n_k} = \sum_{k=1}^{\infty} \frac{\delta}{2} \to \infty. \qquad \square$$

To complete the results promised in (iii) and conclude our discussion, we show that Proposition 7 is essentially sharp, in that no stronger global information on a counting function is available strictly from the fact that the reciprocal sum is convergent, or even very small.

Proposition 8. *Suppose $f : [1, \infty) \to [0, 1]$ with $f(x) \to 0$. For every $\epsilon > 0$, there exists $A \subseteq \mathbb{N}$ such that $\mathcal{C}_A(n) \geq f(n)n$ for infinitely many $n \in \mathbb{N}$ and $\sum_{n \in A} 1/n < \epsilon$.*

Proof. Fix $0 < \epsilon < 1$ and $f : [1, \infty) \to [0, 1]$ satisfying $f(x) \to 0$ as $x \to \infty$. Choose an increasing sequence $\{n_k\}_{k=1}^{\infty}$ of natural numbers satisfying $n_k, f(n_k)^{-1} \geq 2^{k+2}/\epsilon$ for all $k \in \mathbb{N}$. Let $A_k = \mathbb{N} \cap ((1 - f(n_k))n_k, n_k]$, and let $A = \bigcup_{k=1}^{\infty} A_k$. For every $k \in \mathbb{N}$ we have $\mathcal{C}_A(n_k) \geq |A_k| \geq f(n_k)n_k$, and further

$$\sum_{n \in A} \frac{1}{n} \leq \sum_{k=1}^{\infty} \sum_{n \in A_k} \frac{1}{n} \leq \sum_{k=1}^{\infty} \frac{|A_k|}{\min(A_k)} \leq \sum_{k=1}^{\infty} \frac{f(n_k)n_k + 1}{(1 - f(n_k))n_k} < \sum_{k=1}^{\infty} \frac{2(f(n_k)n_k + 1)}{n_k} \leq \sum_{k=1}^{\infty} \frac{\epsilon}{2^k} = \epsilon. \qquad \square$$

A particular consequence of Proposition 8 is the fact that two sets of natural numbers can have nearly identical reciprocal sum functions, up to a tiny fixed error, but have counting functions that behave very differently. For example, let P be the set of primes, let A be the set yielded by Proposition 8 with $\epsilon = .00001$ and $f(x) = 1/\log\log\log x$, and let $B = P \cup A$. We can certainly say that $S_B(x)$ is well-approximated by $\log\log x$, just as $S_P(x)$ is, but in fact the relationship is even stronger, as $S_B(x)$ is almost *exactly* the same as $S_P(x)$, with $|S_B(x) - S_P(x)| < .00001$ for all x. However, while P consistently contains around $n/\log n$ of the first n natural numbers, B contains at least $n/\log\log\log n$ of the first n natural numbers infinitely often, a dramatic difference in density.

Acknowledgements: Proposition 7 was inspired by Chapter 10 of [4], and also appears as an exercise in Chapter 3 of [12]. In the special case where $S(x) = \log\log x - C$, exercises in Chapter 3 of [10] are similar to Propositions 4 and 5.

References

[1] J. Bayless, D. Klyve, *Reciprocal sums as a knowledge metric: theory, computation, and perfect numbers*, Amer. Math. Monthly, vol. 120 (2013), no. 9, 822-831.

[2] F. A. Behrend, *On sets of integers containing no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U. S. A. 32 (1946), 331-332.

[3] T. Bloom, O. Sisask, *Breaking the logarithmic barrier in Roth's theorem on arithmetic progressions*, preprint (2020), arXiv:2007.03528

[4] P. L. Clark, *Number theory: a contemporary introduction*, http://alpha.math.uga.edu/~pete/4400FULL2018.pdf

[5] P. Erdős, P. Turán, *On some sequences of integers*, Jour. Lond. Math. Soc., s1-11 (1936), 261-264.

[6] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. 11 (2001), no. 3, 465-588.

[7] B. Green, *Roth's theorem in the primes*, Ann. of Math. 161 (2005), no. 3, 1609-1636.

[8] B. Green, T. Tao, *New bounds for Szemerédi's theorem, III: A polylogarithmic bound for $r_4(N)$*, Mathematika 63 (2017), no. 3, 944-1040.

[9] B. Green, T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. 167 (2008), no. 2, 481-547.

[10] A. J. Hildebrand, *Introduction to analytic number theory, math 531 lecture notes*, https://faculty.math.illinois.edu/~hildebr/ant/

[11] E. Naslund (https://mathoverflow.net/users/12176/eric-naslund), *The Erdős-Turán conjecture or the Erdős conjecture?*, URL (version: 2013-08-05): https://mathoverflow.net/q/132648

[12] P. Pollack, *Not always buried deep: a second course in elementary number theory*, AMS, 2009, http://pollack.uga.edu/NABDofficial.pdf

[13] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. 28 (1953), 104-109.

[14] E. Szmerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. 27 (1975), 199-245.

Department of Mathematics, Millsaps College, Jackson, MS 39210

Email address: riceaj@millsaps.edu