2022 S.T. Yau High School Science Award (Asia)

Research Report

The Team
Registration Number: Math-160

Name of team member: CHEUNG Bock Man, CHAN Hiu Long
School: Baptist Lui Ming Choi Secondary School
City, Country: Hong Kong, China

Name of supervising teacher: LEE Wing Yan
Position: Mathematics Teacher
School/Institution: Baptist Lui Ming Choi Secondary School
City, Country: Hong Kong, China

Title of Research Report: On The Coprime Product Series and Its Divergence and Bounds

Date: 2022/08/31

On The Coprime Product Series and Its Divergence and Bounds

Cheung Bock Man, Chan Hiu Long

Abstract

It has been shown by L. Euler [1] that the sum of the reciprocals of all prime numbers diverges, and its growth is proportional to $\log\log n$. Others [2, 3, 4] have shown that, in particular:

$$\sum_{p\leq n} \frac{1}{p} = \log\log n + M + O\left(\frac{1}{\log^2 n}\right),$$

where $M = 0.261497...$ is a constant.
We estimate and prove an asymptotic formula on the reciprocals of the coprime series,

$$\sum_{\substack{1\leq s_1,s_2\leq n \\ (s_1,s_2)=1}} \frac{1}{s_1 s_2} = \frac{6}{\pi^2}\log^2 n + O(\log n)$$

where the sum runs over natural numbers s_1 and s_2 that satisfy $(s_1, s_2) = 1$ diverges, and in general we show that,

$$\sum_{\substack{1\leq s_1,s_2,\cdots,s_k\leq n \\ (s_1,s_2,\cdots,s_k)=1}} \frac{1}{s_1\cdots s_k} = \frac{1}{\zeta(k)}\log^k n + O(\log^{k-1} n)$$

where the sum runs over $s_1, s_2, \cdots, s_k$ with $(s_1, s_2, \cdots, s_k) = 1$. This implies that the natural density of coprime k-tuples over $\mathbb{N}^k$ is $1/\zeta(k)$.

Keywords: analytic number theory, reciprocals, series, coprime, divergence, bound, density

Acknowledgement:

We would like to express our gratitude to our Mathematics Teacher Ms. Wing Yan Lee, with whom we received tremendous support at various stages of the research process.

We are obliged to Dr. Gary P.T. Choi, who provided many excellent comments and patiently guided us to finish the paper.

A special thanks goes to our principal, Mr. Ka Kit Wan, for inspiring us to start this project.

As a final note, we would like to extend our sincere thanks to all the teachers and friends without whom our project would not be possible. We dedicate this project to all the bright young people who are pursuing their interests.

Commitments on Academic Honesty and Integrity

We hereby declare that we

1. are fully committed to the principle of honesty, integrity and fair play throughout the competition.
2. actually perform the research work ourselves and thus truly understand the content of the work.
3. observe the common standard of academic integrity adopted by most journals and degree theses.
4. have declared all the assistance and contribution we have received from any personnel, agency, institution, etc. for the research work.
5. undertake to avoid getting in touch with assessment panel members in a way that may lead to direct or indirect conflict of interest.
6. undertake to avoid any interaction with assessment panel members that would undermine the neutrality of the panel member and fairness of the assessment process.
7. observe the safety regulation of the laboratory(ies) where we conduct the experiment(s), if applicable.
8. observe all rules and regulations of the competition.
9. agree that the decision of YHSA(Asia) is final in all matters related to the competition.

We understand and agree that failure to honour the above commitments may lead to disqualification from the competition and/or removal of reward, if applicable; that any unethical deeds, if found, will be disclosed to the school principal of team member(s) and relevant parties if deemed necessary; and that the decision of YHSA(Asia) is final and no appeal will be accepted.

(Signatures of full team below)

張博文

Team member: CHEUNG Bock Man

Team member: CHAN Hiu Long

Yan Lee

Supervising teacher: LEE Wing Yan

Noted and endorsed by

Wan

(Signature)

Baptist Lui Ming Choi Secondary School
Hong Kong
(Shatin, N.T.)

Name of the School Principal:

WAN KA KIT

On The Coprime Product Series and Its Divergence and Bounds

31/08/2022

Contents

1 Introduction

Prime numbers have been in the spotlight in Number Theory historically. In fact, the properties of the reciprocals of prime numbers are well known and thoroughly studied, such as their asymptotic growth and density. Another mathematically similar concept is coprimality (Two or more integers whose greatest common divisor is 1). As of date, however, not much attention is received on the reciprocals of products of coprime pairs, which seems like a "general" version of primes.

In this paper, we ask ourselves: How does the series of reciprocals products of coprime numbers asymptotically behave like? Precisely, can we derive an expression for $\sum_{\substack{1\le s_1,s_2\le n\\(s_1,s_2)=1}} \frac{1}{s_1 s_2}$, based on some known facts such as the probability of 2 randomly chosen integers being coprime is $1/\zeta(2) = 6/\pi^2$? To answer this, we will start from some basic facts in prime numbers, and then discuss an asymptotic formula for the series of reciprocal products of coprimes, which holds several implications on the density of coprimes.

At the time of Euclid, the fact that there are an infinite number of prime numbers has been known. Since then, mathematicians have developed new approaches to tackle the problem.

In the 1730s, Leonhard Euler discovered a way [1] to prove that the sum of the reciprocals of all prime numbers diverges, with the key observation that the harmonic series, later recognized as $\lim_{s\to 1}\zeta(s)$, can be factored into a product of primes.

Take $\mathbb{P}$ to be the set of prime numbers.

To see why $\sum_{p\in\mathbb{P}} 1/p$ diverges, we begin with the Euler Product of the Riemann Zeta function,

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p\in\mathbb{P}}^{\infty} (1-p^{-s})^{-1}$$

Taking log on both sides, we have

$$\begin{aligned}\log \sum_{n=1}^{\infty} \frac{1}{n^s} &= -\sum_{p} \log(1-p^{-s}) \\ &= \sum_{p}\sum_{m=1}^{\infty} m^{-1}p^{-ms} \\ &= \sum_{p} p^{-s} + \sum_{p}\sum_{m=2}^{\infty} m^{-1}p^{-ms}\end{aligned}$$

Note that

$$\sum_{p}\sum_{m=2}^{\infty} m^{-1}p^{-ms} < \sum_{p}\sum_{m=2}^{\infty} p^{-m} = \sum_{p} \frac{1}{p(p-1)} < 1$$

It suffices to show that

$$\lim_{s\to 1}\sum_p p^{-s} \to \infty$$

since $\zeta(s) \to \infty$ as $s \to 1$.

Now, informally, since the harmonic series grow in the order of $\log n$, for very large n we have

$$\log \sum_{\substack{i\leq n\\ i\in\mathbb{N}}} \frac{1}{i} \sim \log\log n$$

and so we expect $\sum_{p\leq n} 1/p \sim \log\log n$ for large n.

In fact, there is an asymptotic formula proven by Mertens [2] that for the summation of reciprocals of primes up to n, we have

$$\sum_{p\leq n}\frac{1}{p} = \log\log(n) + M + O\left(\frac{1}{\log n}\right) \tag{1}$$

where $M = \gamma + \sum_p(\log(1-1/p)+1/p) \approx 0.261497212847643$ is a constant and $\gamma \approx 0.577$ is the Euler-Mascheroni Constant.
The error term was improved by E. Landau [3] in 1909, in which by exploiting the prime number theorem he found that it can be further bounded to $O(e^{-(\log n)^{1/14}})$. Others [4] estimated it as

$$\left|\sum_{p\leq n}\frac{1}{p} - (\log\log(n)+M)\right| \leq \frac{1}{10\log^2 n} + \frac{4}{15\log^3 n}$$

Compared with primality, a closely related concept is coprimality. We say the **greatest common divisor** of a and b is d, if d is the largest integer d such that $d|a$ and $d|b$. We write this as $gcd(a,b) = d$, or if no ambiguity is present, $(a,b) = d$. We say two integers a and b are **coprime** if $d = 1$. It is natural to ask whether similar results can be established for a summation of the form

$$\sum_{\substack{1\leq s_1,s_2\leq n\\ (s_1,s_2)=1}} \frac{1}{s_1 s_2}, \tag{2}$$

Or further, for k integers $s_1, s_2, \cdots, s_k$ which are coprime (their greatest common divisor is 1), we are tempted to make the generalization:

$$\sum_{\substack{1\leq s_1,s_2,\cdots,s_k\leq n\\ (s_1,s_2,\cdots,s_k)=1}} \frac{1}{s_1 s_2\cdots s_k} \tag{3}$$

Establishing such a formula for equation (2) and (3) will be one of the main goals in this paper.

An interesting interpretation of the asymptotic formula of the prime numbers is the concept of density, or loosely speaking, how “frequent” prime numbers appear in a set.

Definition 1. *The* **density** *of a set S (whose items are k-tuples of the form $(s_1, s_2, ..., s_k)$ where each s_i is a positive integer) over $\mathbb{N}^n$ is written as $\boldsymbol{\rho}_{\mathbb{N}^n} S$ (or simply $\boldsymbol{\rho} S$) and defined as*

$$\lim_{n\to\infty} \frac{\#\{s \in S : s_i \le n \text{ for all } 0 < i \le k\}}{n^k}.$$

Example 1. *The density of even numbers over $\mathbb{N}$ is $1/2$, and the density of $\mathbb{P}$ over $\mathbb{N}$ is 0.*

By the explicit formula proved in Merten's Second Theorem, a straightforward implication is that the prime numbers has density 0 in the natural numbers, since

$$\lim_{n\to\infty} \frac{\log\log n}{n} = 0.$$

2 Asymptotic Formula of the Sum of Reciprocal Products of Two Coprime Numbers

In this section, our goal is to make an educated guess on the rate of divergence (growth of partial sums) of the sum of reciprocals of coprime series, an analog for the reciprocals of primes, and explain the motivation behind.

2.1 Divergence of the Series

The divergence of

$$\sum_{\substack{1\le s_1, s_2 \le n \\ (s_1,s_2)=1}} \frac{1}{s_1 s_2}$$

can be seen as a corollary to the divergence of the harmonic series.

Proposition 1.

$$\sum_{\substack{1\le s_1, s_2 \le n \\ (s_1,s_2)=1}} \frac{1}{s_1 s_2} \to \infty \text{ as } n \to \infty.$$

Proof. Note that

$$\sum_{\substack{1\le s_1, s_2 \le n \\ (s_1,s_2)}} \frac{1}{s_1 s_2} \ge \sum_{1\le n} \frac{1}{n}$$

by letting $s_2 = 1$.
Since the sum on the right diverges as $n \to \infty$, the sum on the left diverges. $\square$

2.2 An Estimate of the Series

Now that we have shown that the above series diverges, we can set off to make our estimation based on a few observations.

The probability that a number is divisible by any prime p is $1/p$. Hence, the probability for two numbers to be both divisible by p is $1/p^2$, and $1 - 1/p^2$ for at least one to not be divisible by p. Because two numbers are coprime if they share no common divisors, and on top of that we know the probability that any finite collection of divisibility events associated with distinct primes is mutually independent, we have

$$\Pr(\text{two integers are coprime}) = \prod_{p \in \mathbb{P}} \left(1 - \frac{1}{p^2}\right) = \left(\prod_{p \in \mathbb{P}} \left(1 - p^{-2}\right)^{-1}\right)^{-1} = \zeta(2)^{-1}$$

$$\Pr(\text{two integers are coprime}) = \frac{1}{\zeta(2)} \tag{4}$$

Another observation is that the harmonic series $\sum_{i=1}^{n} 1/i \approx \log n + \gamma$, where $\gamma = 0.577...$ is the Euler-Mascheroni constant, so combining the two, it is reasonable to have the following estimate:

$$\begin{aligned}\sum_{\substack{1 \le s_1, s_2 \le n \\ (s_1, s_2) = 1}} \frac{1}{s_1 s_2} &\approx \frac{1}{\zeta(2)} \sum_{s_1=1}^{n} \frac{1}{s_1} \sum_{s_2=1}^{n} \frac{1}{s_2} \\ &\approx \frac{1}{\zeta(2)} (\log^2 n + 2\gamma \log n + \gamma^2) \\ &\approx \frac{1}{\zeta(2)} \log^2 n + \text{constant} \cdot \log n\end{aligned}$$

Therefore, we conjecture that

Theorem 1.
$$\sum_{\substack{1 \le s_1, s_2 \le n \\ (s_1, s_2) = 1}} \frac{1}{s_1 s_2} = \frac{1}{\zeta(2)} \log^2 n + O(\log n).$$

After checking our estimate using direct computation, it further suggests that the error term is $O(\log n)$. Refer to Section 5 for detailed figures.

In the next section, we attempt to prove this conjecture rigorously.

3 Proof of the Asymptotic Formula

3.1 Lemmas

Before we show the following results, we shall define the Möbius μ and Euler φ functions.

Definition 2. *Let $n = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$ be the prime factorization of n.*
(a) The **Möbius function**, *$\mu(n)$, is defined as*

$$\mu(1) = 1$$
$$\mu(n) = \begin{cases} (-1)^k & \text{if } a_i = 1 \text{ for all } 1 \le i \le k \\ 0 & \text{otherwise.} \end{cases}$$

(b) The **Euler totient function**, *$\varphi(n)$, is defined as the number of positive integers not exceeding n which are coprime to n.*

To begin with, we first establish a few theorems and lemmas, taken from M. Ram Murty's *Problems in Analytic Number Theory* [5].

Lemma 1. *For all $n \ge 1$,*

$$\sum_{d|n} \mu(d) = \begin{cases} 1, & n = 1 \\ 0, & n > 1. \end{cases}$$

In other words, the sum is a characteristic function of the number 1.

Proof. When $n = 1$, the statement is clearly true. However, when $n > 1$, let $n = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$ as above. Notice that $\mu(d) \neq 0$ if and only if $d = 1$ or a product of distinct primes. Therefore

$$\begin{aligned} \sum_{d|n} \mu(d) &= \mu(1) + \mu(p_1) + \cdots + \mu(p_k) + \mu(p_1 p_2) + \cdots + \mu(p_{k-1} p_k) \\ &\qquad + \cdots + \mu(p_1 p_2 \cdots p_k) \\ &= 1 + \binom{k}{1}(-1) + \binom{k}{2}(-1)^2 + \cdots + \binom{k}{k}(-1)^k \\ &= (1-1)^k \\ &= 0. \end{aligned}$$

□

Lemma 2. *For all $n \ge 1$,*

$$\sum_{d|n} \varphi(d) = n.$$

Proof. Let S be the set $\{1, 2, \cdots, n\}$, and for each d that divides n, let

$$A(d) = \{k \mid (k, n) = d, 1 \le k \le n\}.$$

All such $A(d)$ are disjoint, and their union is precisely S. Now $\#A(d) = \varphi(n/d)$, which is the number of k's such that $(k/d, n/d) = 1$. Therefore

$$\sum_{d|n} \varphi(n/d) = \sum_{d|n} \varphi(d) = n.$$

□

Lemma 3.

$$\frac{\varphi(n)}{n} = \sum_{d|n} \frac{\mu(d)}{d}.$$

Proof. Suppose

$$f(n) = \sum_{d|n} \mu(d)(1/d).$$

Then

$$\begin{aligned}
\sum_{d|n} f(d) &= \sum_{d|n} \sum_{e|d} (d/e)\mu(e) \\
&= \sum_{est=n} s\mu(e) \\
&= \sum_{s|n} s \sum_{e|\frac{n}{s}} \mu(e) \\
&= n \qquad \text{(by Lemma 1)}.
\end{aligned}$$

By Lemma 2, letting f be φ gives us the required formula. □

Lemma 4.

$$\sum_{\substack{n \le x \\ (n,k)=1}} \frac{1}{n} \sim \frac{\varphi(k)}{k} \log x, \text{ for } x \to \infty.$$

Proof. By Lemma 1,

$$\begin{aligned}
\sum_{\substack{n \le x \\ (n,k)=1}} \frac{1}{n} &= \sum_{n \le x} \frac{1}{n} \sum_{d|(n,k)} \mu(d) \\
&= \sum_{d|k} \mu(d) \sum_{\substack{n \le x \\ d|n}} \frac{1}{n} \\
&= \sum_{d|k} \frac{\mu(d)}{d} \sum_{t \le x/d} \frac{1}{t} \\
&= \sum_{d|k} \frac{\mu(d)}{d} \left(\log \frac{x}{d} + O(1)\right),
\end{aligned}$$

by the asymptotic formula for the harmonic series. Therefore,

$$\begin{aligned}
\sum_{d|k} \frac{\mu(d)}{d} \left(\log \frac{x}{d} + O(1)\right) &= \sum_{d|k} \frac{\mu(d)}{d} \log x - \sum_{d|k} \frac{\mu(d)}{d} (\log d + O(1)) \\
&= \sum_{d|k} \frac{\mu(d)}{d} \log x + O(1) \\
&= \frac{\varphi(k)}{k} \log x + O(1) \qquad \text{(by Lemma 3)}
\end{aligned}$$

□

Lemma 5.

$$\sum_{k=1}^{n} \varphi(k) = \frac{3n^2}{\pi^2} + O(n \log n).$$

Proof. See [6]. $\square$

Lemma 6. *(Abel's summation formula) Suppose $\{a_n\}_{n=1}^{\infty}$ is a sequence of real numbers and $f(t)$ is a continuously differentiable function on $[1, x]$. If*

$$A(t) = \sum_{n \leq t} a_n,$$

then

$$\sum_{n \leq x} a_n f(n) = A(x)f(x) - \int_1^x A(t)f'(t)dt.$$

Proof. See [5], pp. 17-18. $\square$

We are equipped with enough tools to prove Theorem 1.

3.2 Proof

Proof. *(Theorem 1)*

$$\begin{aligned}
\sum_{\substack{1 \leq s_1, s_2 \leq n \\ (s_1, s_2)=1}} \frac{1}{s_1 s_2} &= \sum_{s_1=1}^{n} \sum_{\substack{s_2 \leq n \\ (s_1, s_2)=1}} \frac{1}{s_1 s_2} \\
&= \sum_{s_1=1}^{n} \frac{1}{s_1} \sum_{\substack{s_2 \leq n \\ (s_1, s_2)=1}} \frac{1}{s_2} \\
&\sim \sum_{s_1=1}^{n} \frac{1}{s_1} \frac{\varphi(s_1)}{s_1} \log n \\
&= \log n \sum_{s_1=1}^{n} \frac{\varphi(s_1)}{s_1^2}
\end{aligned}$$

By Lemmas 5 and 6,

$$\begin{aligned}
\sum_{s=1}^{n} \frac{\varphi(s)}{s^2} &= \frac{3n^2}{\pi^2 n^2} - \int_1^n \left(\frac{3t^2}{\pi^2} + O(t \log t) \right) \left(\frac{-2}{t^3} \right) dt + O\left(\frac{\log n}{n} \right) \\
&= \frac{3}{\pi^2} + \int_1^n \left(\frac{6}{\pi^2 t} + O\left(\frac{\log t}{t^2} \right) \right) dt + O\left(\frac{\log n}{n} \right) \\
&= \frac{3 + 6\log(n)}{\pi^2} + O\left(\frac{\log n}{n} \right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{\substack{1\le s_1,s_2\le n\\(s_1,s_2)=1}} \frac{1}{s_1 s_2} &= \log n\left(\frac{3+6\log(n)}{\pi^2}+O\left(\frac{\log n}{n}\right)\right)\\
&= \frac{6}{\pi^2}\log^2 n+\frac{3}{\pi^2}\log n+O\left(\frac{\log^2 n}{n}\right)\\
&= \frac{1}{\zeta(2)}\log^2 n+O(\log n)
\end{aligned}$$

□

See Figure 5.3 for a computed upper bound for the constant of proportionality of the $O(\log n)$ term.

This theorem immediately implies a generalization of the condition on $(s_1, s_2) = m$:

Corollary 1.

$$\sum_{\substack{1\le s_1,s_2\le n\\(s_1,s_2)=m}} \frac{1}{s_1 s_2} = \frac{1}{\zeta(2)}\frac{\log^2 n}{m^2}+O(\log n)$$

Proof. Writing $s_1 = mr_1$ and $s_2 = mr_2$,

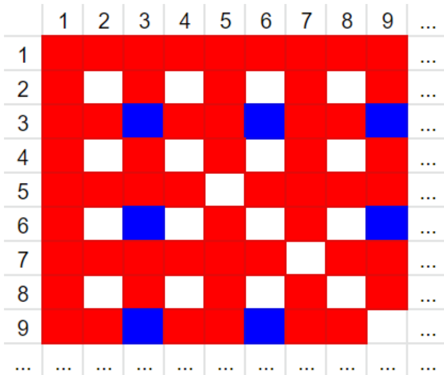
$$\begin{aligned}
\sum_{\substack{1\le s_1,s_2\le n\\(s_1,s_2)=m}} \frac{1}{s_1 s_2} &= \frac{1}{m^2}\sum_{\substack{1\le r_1,r_2\le n/m\\(r_1,r_2)=1}} \frac{1}{r_1 r_2}\\
&= \frac{1}{m^2}\left(\frac{1}{\zeta(2)}\log^2\frac{n}{m}+O\left(\log\frac{n}{m}\right)\right)\\
&= \frac{1}{\zeta(2)}\frac{1}{m^2}\left(\log^2 n-\log m\log n+\log^2 m\right)+O(\log n)\\
&= \frac{1}{\zeta(2)}\frac{\log^2 n}{m^2}-\text{constant}\cdot\log n+\text{constant}+O(\log n)\\
&= \frac{1}{\zeta(2)}\frac{\log^2 n}{m^2}+O(\log n)
\end{aligned}$$

□

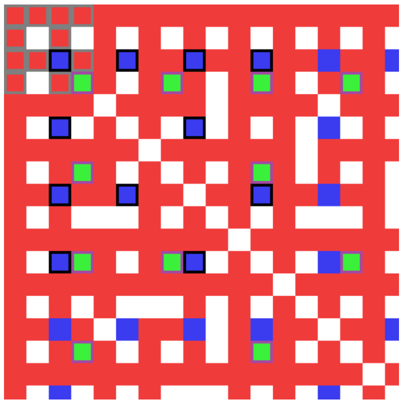
As a remark, this corollary also suggests that the sum diverges for any m.

3.3 Geometric Justification

This corollary can also be viewed geometrically. In the following graph, the number pairs whose gcd is 3 are colored as blue, and pairs whose gcd is 1 are colored as red on an infinite square grid. The blue region is disjoint from the red, and the red area is equal to m^2 times (in this case, 9 times) the blue area, since the regions are similar.



Number pairs with gcds 1, 3 and 4 are highlighted as red, blue and green respectively. A small figure is bordered to highlight the similarity.



4 Asymptotic Formula of the Sum of Reciprocal Products of k Coprime Tuples

4.1 Sum of Reciprocals of k-tuple Coprime Numbers

In view of Proposition 1, it is easy to see that sum of k-tuples of reciprocals of coprime integers $s_1, s_2, \cdots, s_k$ also diverges,

Proposition 2.

$$\sum_{\substack{1 \le s_1, s_2, \cdots, s_k \le n \\ (s_1, s_2, \cdots, s_k)=1}} \frac{1}{s_1 s_2 \cdots s_k} \to \infty \ as \ n \to \infty$$

Proof.

$$\sum_{\substack{1 \le s_1, s_2, \cdots, s_k \le n \\ (s_1, s_2, \cdots, s_k)=1}} \frac{1}{s_1 s_2 \cdots s_k} \ge \sum_{1 \le n} \frac{1}{n}$$

by letting $s_2, \cdots, s_k = 1$. The right hand side again diverges, so we are done. □

However, determining the rate of divergence is not as easy. Based on the results proven above, it is natural to conjecture on the formula of the general case of Theorem 1 for the k-tuple sum.

The argument of the probability of coprimality can be extended to k integers similar to Equation 4, and is shown [7] to be

$$\Pr(\text{k integers are coprime}) = \frac{1}{\zeta(k)} \tag{5}$$

Therefore, we conjecture that

Theorem 2.

$$\sum_{\substack{1\le s_1,s_2,\cdots,s_k\le n\\(s_1,s_2,\cdots,s_k)=1}} \frac{1}{s_1 s_2\cdots s_k} = \frac{1}{\zeta(k)}\log^k n + O(\log^{k-1} n)$$

Informally, one way to look at Theorem 2 is by observing

$$\begin{aligned}
\sum_{\substack{1\le s_1,s_2,\cdots,s_k\le n\\(s_1,s_2,\cdots,s_k)=1}} \frac{1}{s_1 s_2\cdots s_k} &\approx \frac{1}{\zeta(k)}\prod_i^k \sum_{s_i=1}^n \frac{1}{s_i}\\
&\approx \frac{1}{\zeta(k)}(\log n+\gamma)^k\\
&= \frac{1}{\zeta(k)}\left(\sum_{r=0}^k \binom{k}{r}\gamma^{k-r}\log^r n\right)\\
&\approx \frac{1}{\zeta(k)}\log^k n + O(\log^{k-1} n)
\end{aligned}$$

Numerically, for the $k = 3$ case, the plots match the prediction that the sum has an error of magnitude $\log^2 n$. For details, see Section 6.

4.2 Corollary

Some interesting consequences arise if we infer that Theorem 2 is true. For instance, if we generalize the notion of greatest common divisors, we have

Corollary 2.

$$\sum_{\substack{1\le s_1,s_2,\cdots,s_k\le n\\(s_1,s_2,\cdots,s_k)=m}} \frac{1}{s_1 s_2\cdots s_k} = \frac{1}{\zeta(k)}\frac{\log^k n}{m^k} + O(\log^{k-1} n)$$

It sheds light on why $\frac{1}{\zeta(k)}$ seems to be the only suitable candidate to be the coefficient of the theorem. If we know that for some constant A,

$$\sum_{\substack{1\le s_1,s_2,\cdots,s_k\le n\\(s_1,s_2,\cdots,s_k)=m}} \frac{1}{s_1 s_2\cdots s_k} = A\cdot\frac{\log^k n}{m^k} + O(\log^{k-1} n)$$

The product of k-harmonic sums gives the sum of all possible combinations of $\frac{1}{s_1 s_2 \cdots s_k}$. By grouping them based on their greatest common divisors $(s_1, s_2, \cdots, s_k)$, we have

$$\begin{aligned}
\log^k n \approx \prod_{i=1}^{k} \left(\sum_{s_i} \frac{1}{s_i} \right) &= \sum_{\substack{1 \le s_1, s_2, \cdots, s_k \le n \\ (s_1, s_2, \cdots, s_k) = 1}} \frac{1}{s_1 s_2 \cdots s_k} + \sum_{\substack{1 \le s_1, s_2, \cdots, s_k \le n \\ (s_1, s_2, \cdots, s_k) = 2}} \frac{1}{s_1 s_2 \cdots s_k} + \cdots \\
&\approx A \left(\frac{\log^k n}{1^k} + \frac{\log^k n}{2^k} + \cdots \right) \\
&= A \log^k n \cdot \zeta(k)
\end{aligned}$$

Comparing the coefficients, we have $A \cdot \zeta(k) = 1$, which verifies that $A = \frac{1}{\zeta(k)}$.

In the next section, we will prove Theorem 2 using induction, which heavily relies on this corollary.

4.3 Induction Proof

We begin by stating a lemma which would be particularly useful in the proof:

Lemma 7. *(Dirichlet Series of Euler Totient Function [5]) For* $Re(s) \geq 2$*,*

$$\sum_{k=1}^{\infty} \frac{\varphi(k)}{k^s} = \frac{\zeta(s-1)}{\zeta(s)}$$

Proof. From Lemma 3,

$$\frac{\varphi(n)}{n} = \sum_{d|n} \frac{\mu(d)}{d}$$

We have

$$\sum_{k=1}^{\infty} \frac{\varphi(k)}{k^s} = \sum_{k=1}^{\infty} \frac{\mu(k)}{k^s} \sum_{k=1}^{\infty} \frac{k}{k^s} = \frac{\zeta(s-1)}{\zeta(s)}$$

□

We can use Lemma 7 and the idea described in the last section (by assuming Corollary 2 from the induction hypothesis) to deduce the next case, and inductively prove the general case.

The process is summarized in the diagram below:

$$\begin{tikzcd}
k=2 \text{ (base case)} \arrow[d] & k=3 \arrow[d] & k=4 \arrow[d] & \cdots \\
\text{Corollary for } k=2 \arrow[ur] & \text{Corollary} \arrow[ur] & \text{Corollary} \arrow[ur] & \cdots
\end{tikzcd}$$

Proof. (Theorem 2)

We will prove the statement using mathematical induction for natural numbers $k \geq 2$. The base case is proven in the previous section, so it remains to show the induction step.

Suppose that the statement is true for some positive integers $h \geq 2$. From our induction hypothesis, We can derive a similar expression to our corollary in the $(k = 2)$ case, namely:

$$\sum_{\substack{1 \leq s_1, s_2, \cdots, s_h \leq n \\ (s_1, s_2, \cdots, s_h) = m}} \frac{1}{s_1 s_2 \cdots s_h} = \frac{1}{m^h} \frac{1}{\zeta(h)} \log^h n + O(\log^{h-1} n)$$

We have

$$\begin{aligned} \sum_{\substack{1 \leq s_1, s_2, \cdots, s_{h+1} \leq n \\ (s_1, s_2, \cdots, s_{h+1}) = 1}} \frac{1}{s_1 s_2 \cdots s_{h+1}} &= \sum_{s_{h+1}=1}^{n} \frac{1}{s_{h+1}} \sum_{\substack{1 \leq s_1, s_2, \cdots, s_h \leq n \\ (s_1, s_2, \cdots, s_h, s_{h+1}) = 1}} \frac{1}{s_1 s_2 \cdots s_h} \\ &= \sum_{m=1}^{n} \sum_{\substack{s_{h+1}=1 \\ (m, s_{h+1}) = 1}}^{n} \frac{1}{s_{h+1}} \sum_{\substack{1 \leq s_1, s_2, \cdots, s_h \leq n \\ (s_1, s_2, \cdots, s_h) = m}} \frac{1}{s_1 s_2 \cdots s_h} \end{aligned}$$

Hence, by the induction hypothesis and Lemmas 4 and 7,

$$\begin{aligned} &\sum_{m=1}^{n} \sum_{\substack{s_{h+1}=1 \\ (m, s_{h+1}) = 1}}^{n} \frac{1}{s_{h+1}} \sum_{\substack{1 \leq s_1, s_2, \cdots, s_h \leq n \\ (s_1, s_2, \cdots, s_h) = m}} \frac{1}{s_1 s_2 \cdots s_h} \\ &\sim \sum_{m=1}^{n} \frac{\varphi(m)}{m} \log n \left(\frac{1}{m^h} \frac{1}{\zeta(h)} \log^h n + O(\log^{h-1} n) \right) \\ &= \frac{\log^{h+1} n}{\zeta(h)} \sum_{m=1}^{n} \frac{\varphi(m)}{m^{h+1}} + O(\log^h n) \\ &\sim \frac{1}{\zeta(h)} \left(\frac{\zeta(h)}{\zeta(h+1)} \right) \log^{h+1} n + O(\log^h n) \\ &= \frac{1}{\zeta(h+1)} \log^{h+1} n + O(\log^h n) \end{aligned}$$

By the principle of mathematical induction, the statement is true for all $k \geq 2$. This completes our proof. $\square$

4.4 Natural Density of Coprime Tuples

By the bounds established in Theorem 2, one can prove the following theorem:

Theorem 3. *The natural density of k-tuples of coprimes in $\mathbb{N}^k$ is* $\dfrac{1}{\zeta(k)}$.

Proof. From the asymptotic form of the harmonic series,

$$\begin{aligned}\sum_{1\le s_1,s_2,\cdots,s_k\le n}\frac{1}{s_1s_2\cdots s_k}&=\prod_i^k\sum_{s_i=1}^n\frac{1}{s_i}\\&\approx(\log n+\gamma)^k\\&=\log^k(n)+O(\log^{k-1}n)\end{aligned}$$

Hence, the density of k-tuples of coprimes in $\mathbb{N}^k$ is

$$\lim_{n\to\infty}\frac{\frac{1}{\zeta(k)}\log^k(n)+O(\log^{k-1}n)}{\log^k(n)+O(\log^{k-1}n)}=\frac{1}{\zeta(k)}.$$

□

Note that $\frac{1}{\zeta(k)}$ increases monotonically as k increases. This is because the density of primes less than n decreases to 0 when n is very large. As such, the probability that 2, 3, $\cdots$, k numbers be coprime would be large as well.

Also, in the case where the *gcd* is m,

Corollary 3. *The density of k-tuples of numbers with gcd m is* $\frac{1}{m^k\zeta(k)}$.

Proof. By Corollary 2 of Theorem 2,

$$\lim_{n\to\infty}\frac{\frac{1}{m^k\zeta(k)}\log^k(n)+O(\log^{k-1}n)}{\log^k(n)+O(\log^{k-1}n)}=\frac{1}{m^k\zeta(k)}.$$

□

Hence, it shows that the density of k-tuples with *gcd* m decreases as m increases.

We remark that here, we take an arithmetical approach instead of arguing in a probabilistic sense of the natural density of these k-tuples.

5 Conclusions and Future Work

In this paper, we have conjectured and rigorously proved the divergence and bounds of the sum of the reciprocals of two coprime numbers (Theorem 1), and an arbitrary number of coprime tuples (Theorem 2). From these bounds, we deduce the natural density of k-tuples of coprimes in $\mathbb{N}^k$ (Theorem 3). We have further generalized our results for k-tuples with *gcd* being any positive integer m, proving the divergence and bounds of their sum (Corollary 1 and 2) and their natural density (Corollary 3).

We end this paper with formulating some open questions.

Question 1. *Can we improve the bound of our estimate?*

The bound of our estimate for general k is only accurate to the $\log^k n$ leading term; the bounding constant for $\log^{k-1} n$ is yet to be known. When k is much greater than 2, we use the binomial formula to conjecture that

$$\sum_{\substack{1\leq s_1,s_2,\cdots,s_k\leq n\\(s_1,s_2,\cdots,s_k)=1}} \frac{1}{s_1 s_2\cdots s_k} = \frac{1}{\zeta(k)}\left(\log^k n + \binom{k}{1}\log^{k-1} n + \binom{k}{2}\log^{k-2} n + \cdots\right)$$

If this is true, we may obtain a bound accurate to $O(\log n)$.

However, when $k = 2$ or some other small integer greater than 2, if we apply the same argument, we run into trouble: what would be the f?

$$\sum_{\substack{1\leq s_1,s_2\leq n\\(s_1,s_2)=1}} \frac{1}{s_1 s_2} = \frac{1}{\zeta(2)}(\log^2 n + 2\log n + O(f)).$$

We may suspect that $f = 1$, but notice that the proof of Theorem 1 relies on Lemma 5, which includes an $O(n\log n)$ term. This poses a limit to how strong the bound could be, if we deduce it purely from the binomial theorem. Thus, we may have to rely on different methods to improve the bound on the $k = 2$ case of the theorem, or we would be stuck on the $O(\log n)$ term.

Question 2. *Can we find a formula for a more general series involving multiplicative functions,*

$$\sum_{\substack{1\leq s_1,s_2,\cdots,s_k\leq n\\(s_1,s_2,\cdots,s_k)=1}} \frac{f(s_1 s_2\cdots s_k)}{s_1 s_2\cdots s_k}\ ?$$

For example, in the case of two coprime tuples, we have

$$\sum_{\substack{1\leq s_1,s_2\leq n\\(s_1,s_2)=1}} \frac{f(s_1 s_2)}{s_1 s_2} = \sum_{s_1}^{n} \frac{f(s_1)}{s_1} \sum_{\substack{s_2\\(s_1,s_2)=1}}^{n} \frac{f(s_2)}{s_2}$$

so to further evaluate this sum one way is to find a formula for

$$\sum_{\substack{n\leq x\\(n,k)=1}} \frac{f(n)}{n}$$

Our approach used in this paper somehow may be extended the case of general multiplicative functions. Since the result we established matches the informal approximation we stated earlier, namely

$$\sum_{\substack{1\leq s_1,s_2,\cdots,s_k\leq n\\(s_1,s_2,\cdots,s_k)=1}} \frac{1}{s_1s_2\cdots s_k} \approx \frac{1}{\zeta(k)}\prod_i^k \sum_{s_i=1}^n \frac{1}{s_i}$$

this leads us to the question:

Question 3. *Can we always conveniently approximate the sum using the natural density of coprimes? Moreover, is this approximation an equality concerning the asymptotic formula of multiplicative functions?*

In other words, is

$$\sum_{\substack{1\leq s_1,s_2,\cdots,s_k\leq n\\(s_1,s_2,\cdots,s_k)=1}} \frac{f(s_1s_2\cdots s_k)}{s_1s_2\cdots s_k} \sim \frac{1}{\zeta(k)} \sum_{1\leq s_1,s_2,\cdots,s_k\leq n} \frac{f(s_1s_2\cdots s_k)}{s_1s_2\cdots s_k}$$

true in general as $n\to\infty$?

Question 4. *Is there a similar bound for the sum concerning pairwise coprimes, $(s_i,s_j)=1$, where $s_i,s_j\in\{s_1,s_2,\cdots,s_k\}$ and $s_i\neq s_j$?*

We could use another probabilistic argument as our starting point:

Let $l(q;a,b)$ be the probability that if a numbers are chosen randomly from the set $\{0,1,\cdots,q-1\}$, then at least b numbers are equal to zero. This is equivalent to the probability that at least b out of a random integers are 0 modulo q (*i.e.* divisible by q). It turns out that

$$l(q;a,b) = q^{-a}\sum_{j=0}^{a-b}\binom{a}{j}(q-1)^j,$$

By the previous argument, the probability that k numbers are pairwise coprime is

$$\prod_{p\in\mathbb{P}}(1-l(p;k,k-1))$$

A proof of this can be found in [7]. Although there is no known form of this product in terms of ζ, it is highly probable that this product can be written as other functions in terms of Dirichlet series, such as an L-function:

$$L(s,\chi)=\sum_{n=1}^{\infty}\frac{\chi(n)}{n^s}=\prod_{p\in\mathbb{P}}(1-\chi(p)p^{-s})^{-1}$$

where χ is a Dirichlet character, and s is a complex number with real part greater than 1.

Lastly, previous research [8] has demonstrated a similar result of the natural density of the coprime numbers using abstract fields and functions, and we hope that we can extend our research using more sophisticated tools such as Dirichlet characters and Diophantine approximations in analytic number theory. There is still much to explore on this topic, and we hope that our paper has taken the first step towards answering them.

6 Figures of Numerical Computations

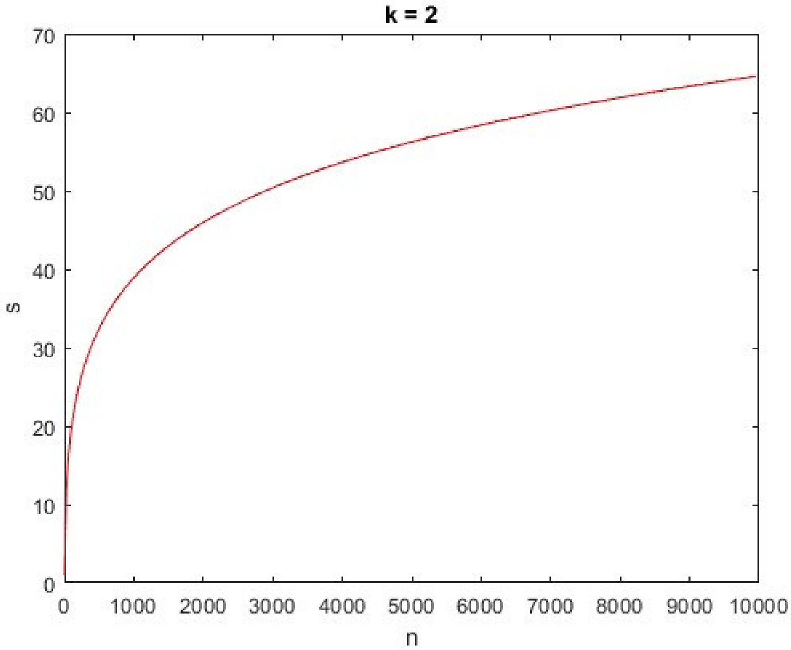


Figure 5.1: $S_2(n) = \sum_{\substack{1 \le s_1, s_2 \le n \\ (s_1, s_2) = 1}} \frac{1}{s_1 s_2}$.

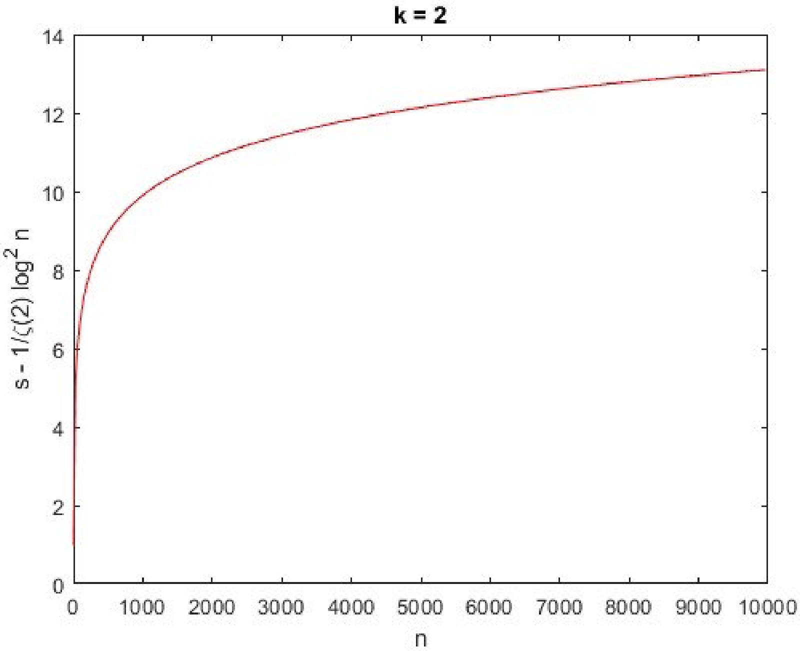


Figure 5.2: The error of $S_2(n)$ from the estimation $\frac{1}{\zeta(2)} \log^2 n$.

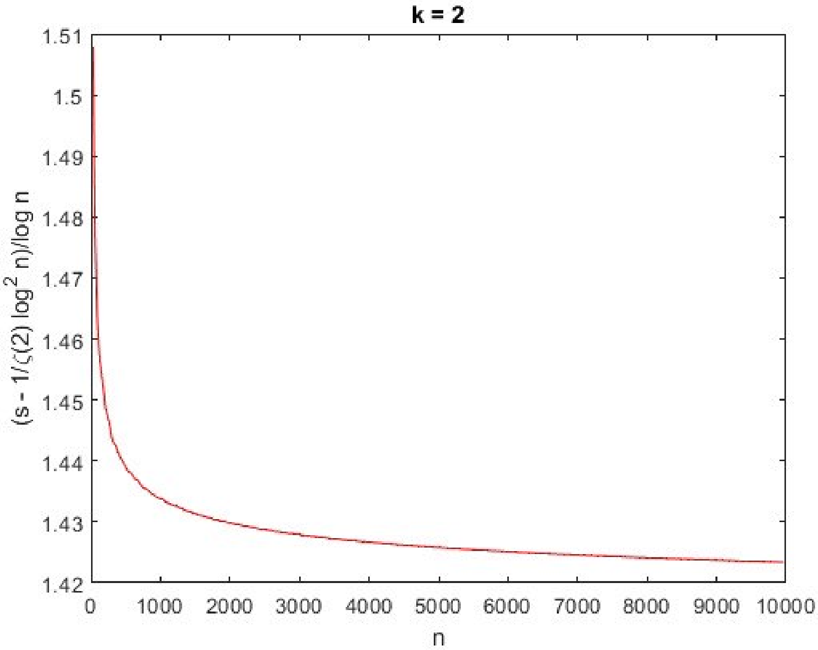


Figure 5.3: The constant C in $S_2(n) = \frac{1}{\zeta(2)} \log^2 n + C \log(n) + O(1)$ has an upper bound of 1.43 as suggested by the computational result.

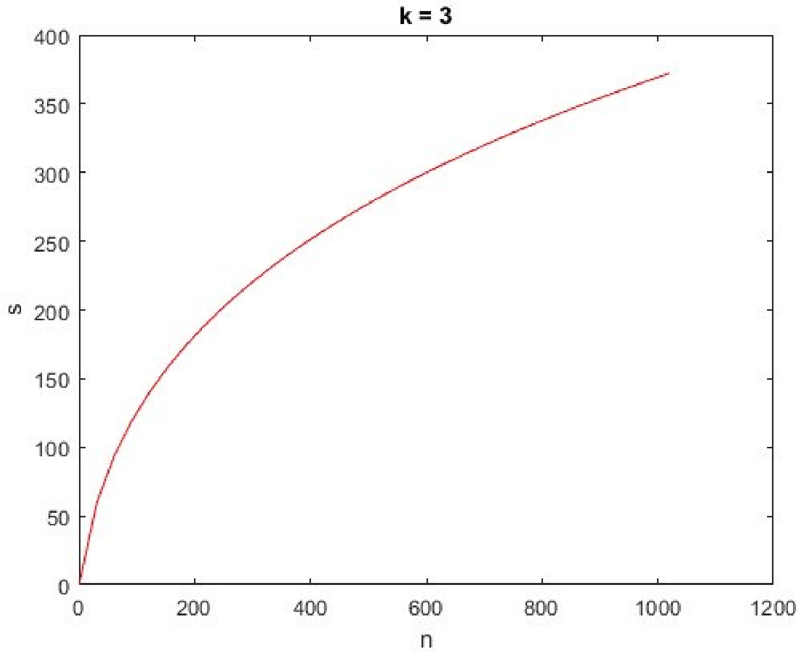


Figure 5.4: $S_3(n) = \sum_{\substack{1 \le s_1, s_2, s_3 \le n \\ (s_1, s_2, s_3) = 1}} \frac{1}{s_1 s_2 s_3}$.

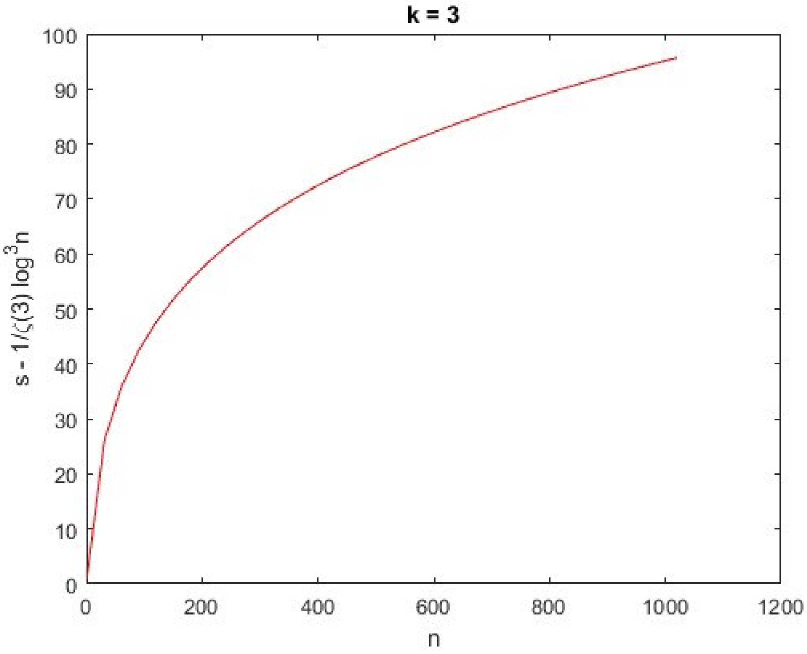


Figure 5.5: The error of $S_3(n)$ from the estimation $\frac{1}{\zeta(3)} \log^3 n$.

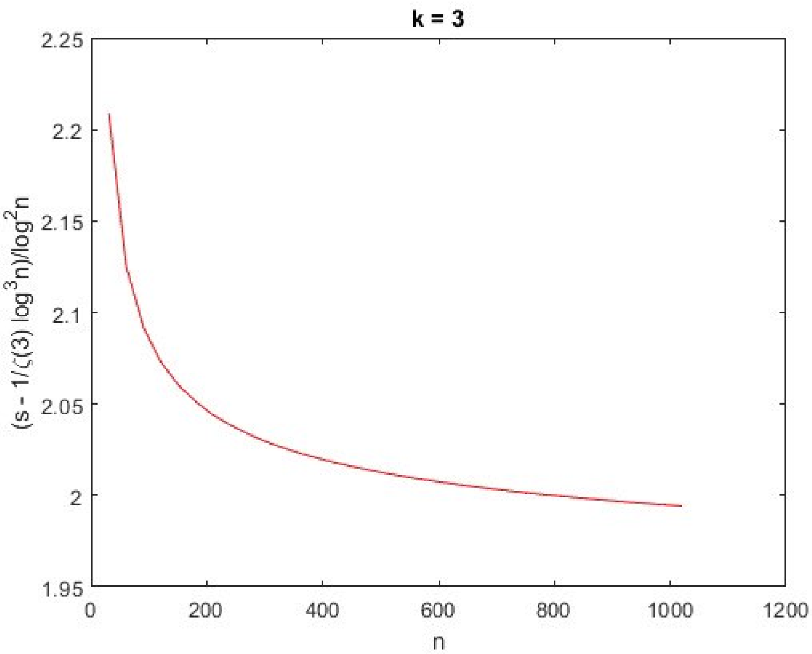


Figure 5.6: The constant C in $S_3(n) = \frac{1}{\zeta(3)} \log^3 n + C \log^2 n) + O(\log n)$ has an upper bound of 2.00 as suggested by the computational result.

7 References

[1] L. Euler (1744), *Commentarii academiae scientiarum Petropolitanae*, Volume 9, pp. 160-188.

[2] F. Mertens (1874), *Ein Beitrag zur analytischen Zahlentheorie, Journal für die reine und angewandte Mathematik 78*, pp. 46-62.

[3] E. Landau (1909), *Handbuch der Lehre von der Verteilung der Primzahlen*, reprinted Chelsea, New York (1953), XIII, §55, pp. 197-203.

[4] P. Dusart (2010), *Estimates of Some Functions Over Primes without R.H.*, https://arxiv.org/abs/1002.0442

[5] M. Ram Murty (2001), *Problems in Analytic Number Theory*, Volume 206, ISBN : 978-1-4757-3443-0.

[6] A. Walfisz (1963), *Weylsche Exponentialsummen in der neueren Zahlentheorie. Mathematische Forschungsberichte. Vol. 16.*

[7] J. E. Nymann (1972), *On the probability that k positive integers are relatively prime, Journal of Number Theory*, Volume 4, Issue 5, pp. 469-473, ISSN 0022-314X.

[8] V. Bergelson, F.K. Richter (2017), *On the density of coprime tuples of the form* $(n, \lfloor f_1(n) \rfloor, \cdots, \lfloor f_k(n) \rfloor)$*, where* $f_1, \cdots, f_k$ *are functions from a Hardy field*, Elsholtz, C., Grabner, P. (eds) Number Theory – Diophantine Problems, Uniform Distribution and Applications. Springer, Cham. https://doi.org/10.1007/978-3-319-55357-3_5