ELEMENTARY
ANALYTIC NUMBER THEORY

BY

MASUM BILLAL

Preface

While analytic number theory is a very broad subject and there are a great many books on this topic, there are not many books that are truly introductory. There are some that are introductory enough such as Apostol[1] but they usually depend on abstract algebra and complex analysis heavily. In contrast to that, it can be argued that both topics are entirely excluded from this book. Thus, the primary objective of this book is to discuss analytic number theory in the most *elementary* way possible. Before I explain what I mean by elementary, I will mention a few more details.

(1) I neither discuss class number of quadratic forms nor do I follow Landau[2] to prove that $L(1, \chi) \neq 0$ for real non-principal Dirichlet character χ. Therefore, one of the two proofs I present for Dirichlet's prime number theorem is a mixture of the approach taken by Landau[3] and Apostol.[4] Thus, we avoid the unnecessary bulky calculation presented by Landau[5] while still keeping the proof of Dirichlet's theorem completely elementary. The other proof is due to Selberg.[6] Both will be included in Chapter III.

(2) I will treat the reader to a topic I consider to be bittersweet–*Sieve Theory*. However, I will only discuss Brun's theorem on prime pairs and the idea behind Selberg's sieve; the reason being that I primarily intend to lay the groundwork for a solid foundation. One can consult Cojocaru and Murty[7] and Friedlander and Iwaniec[8] after reading this chapter. If I only discuss a lot of sieving techniques and prove a lot of theorems, that may cause a lack of sense in the mind of the reader as to why such methods are necessary and what leads one to think in such a way that allows us to prove such powerful results. Friedlander and Iwaniec[9] has an enormous discussion on the matter and it is mostly elementary but I think the text is a little difficult for a non-enthusiast.

[1]Tom Mike Apostol

1976 *Introduction to analytic number theory*, 1st ed., Undergraduate Texts in Mathematics, Springer New York, NY, DOI: 10.1007/978-1-4757-5579-4.

[2]Edmund Landau

1969 *Elementary number theory*, 1st ed., Chelsea, vol. 1, Part Two, Chapter III.

[3]Ibid., Part two, Chapter III, §3.

[4]Apostol, *Introduction to analytic number theory* cit., Chapter VI.

[5]Landau, *Elementary number theory* cit., Theorem 152.

[6]Atle Selberg

1949 "An elementary proof of Dirichlet's theorem about primes in an arithmetic progression", *The Annals of Mathematics*, 50, 2, pp. 297-304, DOI: 10.2307/1969454.

[7]Alina Carmen Cojocaru and Maruti Ram Pedaprolu Murty

2006 *An introduction to sieve methods and their applications*, Cambridge University Press.

[8]Jhn Friedlander and Henryk Iwaniec

2010 *Opera de cribro*, Colloquium Publications, American Mathematical Society, vol. 1.

[9]Ibid.

(3) The reason I include sieve theory at all in this book is that this is the strongest and most interesting area in all of number theory. I will explain further. The development in recent number theory has been a little slow in comparison to the previous century. Brun's theorem on twin primes and Chen's theorem on almost primes related to Goldbach conjecture are still two of the most spectacular results in all of mathematics. Yet both of them are really old and I do not know of any improvements over these results that are of any real significance.

(4) Brun's theorem is of incredible historical importance; being the starting point of sieve methods today despite being over 100 years old. A story goes that Erdős was asked what he thought the strongest theorem was in elementary number theory. His response was Brun's theorem on twin prime pairs. Indeed, one can see that the result of Viggo Brun on twin primes is still the most spectacular result regarding twin primes which requires no analysis or deep results. Friedlander and Iwaniec[10] named their chapter on Brun's theorem *Brun's Sieve–The Big Bang*. It just goes to show how beautiful this result really is and I believe this theorem is the most underrated result in all of mathematics. Another story goes that Landau had been sitting on Brun's paper for 8 years, mostly due to the use of difficult notations by Brun. Later, Landau dedicated an entire chapter in his *Elementaren Zahlentheorie*, volume I.1 of Landau[11] to Brun's theorem. I personally believe this story to be true, and that Landau did this out of guilt that he had deprived the mathematics world from such influential results for so long. The other result, Chen's theorem is based on Selberg's sieve that states that an even number greater than 2 is the sum of a prime and an *almost prime* (product of two primes). To my knowledge, this is still the best result available related to Goldbach's conjecture despite being $\sim$60 years old. As you can see, the most influential results related to the oldest problems in number theory are actually fruits of sieve methods and quite old. This goes to show how difficult it is to improve on sieve methods.

(5) I will discuss Turan's proof of a weaker version of a theorem of normal order by Ramanujan-Hardy which essentially contains *Turan's sieve* in it. I do not go into the sieving technique itself but I show this proof because I mention normal order early in the book.

(6) I will discuss two elementary proofs of the prime number theorem. Again, this begs the question: *exactly what do we consider to be elementary?* which will be answered below. Both proofs will be included in Chapter IV.

(7) I initially wanted to discuss basic complex analysis with connection to the convergence of Dirichlet series but later decided not to include it at all. It simply does

---

[10]Ibid., Chapter VI.
[11]Edmund Landau

1927 *Vorlesungen über Zahlentheorie*, Hirzel, vol. 1.

not go with the spirit of this book. Similarly, I did not follow Apostol[12] and discuss Dirichlet characters from a more general point of view using group theory.

(8) The reader may omit Chapter VI entirely given that it is more of an opinion of mine than an actual mathematical discussion. The reason behind including this chapter is that I believe if this dispute had not occurred, we might have had a few more influential discoveries like the elementary proof of the prime number theorem from the collaboration of Erdős and Selberg. It was a crime that we were deprived of further collaboration between these two mathematical giants just because some third party that did not even witness the incidents first hand poked their noses where they did not belong; which consequently drove a wedge between Erdős and Selberg. It will be clear why I feel so strongly on the matter in the respective chapter. Even though I say this is purely my opinion, proper references will be provided for the history along with some relevant information such as letters between parties involved. I will attach the letters in their original form purely because of historical reasons and also add the corresponding textual versions since they can be difficult to read occasionally. Thus, despite this chapter being a personal opinion, the reader will have some (if not all) necessary relevant evidences that are currently available to me so everyone can draw their own conclusions.

I will now explain what I mean by elementary. My initial thought on elementary is that a result is elementary if it only involves what we learn in grades 1-12 (that is, before undergraduate study starts since the number of grades may differ depending on what country the reader is from). In that sense, basic calculus such as differentiation or integration is elementary. I believe this is an opinion mathematicians will share in general. For example, Landau,[13] Ingham[14] also consider basic integration along with basic properties of zeta functions to be elementary. I should warn the reader that elementary does not imply simplicity. In reality, it is often the exact opposite. Very frequently what we can prove by the use of deep/analytic methods can also be proven by elementary means, but with much more difficulty and an even greater amount of effort. The best example to demonstrate this is the elementary proof of prime number theorem by Selberg and Erdős. Selberg stated in his paper of the elementary proof of the prime number theorem that the proof used only the *simplest properties of logarithm.* And yet it took humanity over 150 years to produce an elementary proof of this theorem and a joint effort of two of the biggest mathematical giants of twentieth century. More context and specific details will be provided on this matter in Chapter VI, The Selberg and Erdős Dispute which will shed light on why this was such a difficult task.

Masum Billal
4:03 AM, 13 May, 2022

---

[12]Apostol, *Introduction to analytic number theory* cit.
[13]Landau, *Elementary number theory* cit.
[14]Albert Edward Ingham

1932 *The distribution of prime numbers*, 1st ed., Cambridge University Press.

Prerequisites

As for the prerequisites of this book, a fair introduction to number theory is necessary. There are many great texts that cover the basics that would suffice for this purpose; so I am not specifying any in particular. However, I have to state that the theory barely matters here. What matters is how well someone can make sense of the theories. I will go into detail using a particular example. Usually most students learn how to calculate greatest common divisor and least common multiple by the time they are in grade 5 or 6. For example, usually in grade 4 or 5 they are taught how to compute the greatest common divisor using Euclid's division algorithm; where one keeps dividing by the smaller number until 0 is reached. Once 0 is found as remainder, the divisor is the greatest common divisor. I prioritize on students making sense of why this division works rather than just using this method as a technique and knowing that this works. In practice, most students are unable to make sense of why this works during their lifetime. The point here is that; there is a way to make sense of it that even a 5th grader can think of. And yet none of the students I have asked this question have ever been able to explain to me why this makes sense. The best answer I get uses prime factorization and computing greatest common divisor from there. If the reader has never thought about this before, it is recommended that the reader tries this before moving on with the book. This is to make the reader understand what is more important for understanding number theory. It does not matter how much someone knows, if they are unable to make sense of it. I think of it like a bottleneck; to the effect that regardless of the volume inside bottle, the output when pouring will still be limited by the size of the cork that prevents the contents from flowing out.

CONTENTS

CHAPTER I.

Arithmetic Functions.

CHAPTER II.

Bertrand to Tschebischeff.

CHAPTER III.

Two Elementary Proofs of Legendre-Dirichlet Prime Number Theorem.

CHAPTER IV.

Two Elementary Proofs of the Prime Number Theorem.

CHAPTER V.

Sieve Theory.

CHAPTER VI.

The Selberg and Erdős Dispute.

Notations

$\gcd(a, b)$ Greatest common divisor of a and b.

$\text{lcm}(a, b)$ Least common multiple of a and b.

$\varphi(n)$ Euler's totient function of n, $\varphi(n)$ is the number of positive integers not exceeding n which are relatively prime to n.

$J_k(n)$ Jordan function of n, the number of tuples $(a_1, \dots, a_k)$ such that $\gcd(a_1, \dots, a_k, n) = 1$ and $1 \leqq a_1, \dots, a_k \leqq n$.

$\tau_k(n)$ Generalized number of divisors of n, $\tau_k(n) = \sum_{d_1 \cdots d_k = n} 1$. For $k = 1$, $\tau_1(n) = \tau(n)$, number of divisors of n.

$\sigma_k(n)$ Generalized sum of divisors of n, $\sigma_k(n) = \sum_{d \mid n} d^k$. For $k = 1$, $\sigma_1(n) = \sigma(n)$, sum of divisors of n.

$\omega(n)$ Number of distinct prime divisors of n.

$\Omega(n)$ Number of total prime divisors of n.

$[x]$ Floor of x, greatest integer not exceeding x.

$I(n)$ Identity function, $I(n) = \left[\frac{1}{n}\right]$.

$\mu(n)$ Möbius function of n, $\mu(n) = (-1)^{\omega(n)}$ if n is square-free, otherwise $\mu(n) = 0$.

$\lambda(n)$ Liouville function of n, $\lambda(n) = (-1)^{\Omega(n)}$.

$\Lambda(n)$ Von Mangoldt Function of n. $\Lambda(n) = \log p$ if $n = p^e$ for some positive integer e, otherwise $\Lambda(n) = 0$.

$\vartheta(x)$ Tchebycheff function of the first kind.

$\psi(x)$ Tchebycheff function of the second kind.

$\zeta(s)$ Zeta function of the complex number s.

$H(x)$ Harmonic sum for x, $H(x) = \sum_{n \leq x} \frac{1}{x}$.

$\alpha * \beta$ Dirichlet convolution of two arithmetic functions α and β.

$\alpha \circ \beta$ General convolution of two arithmetic functions α and β.

$\alpha \cdot \beta$ Generalized convolution of two arithmetic functions α and β.

γ Euler-Mascheroni constant.

CHAPTER I.

ARITHMETIC FUNCTIONS

In this chapter, we will discuss some generalized arithmetic functions and their asymptotic behavior. By asymptotic behavior, we mean that we want to understand how a function $f(x)$ grows as x tends to infinity. A common way of analyzing growth of an arithmetic function f is to consider the order of an arithmetic function. Erdős

Order of Arithmetic Function. The order of an arithmetic function f is defined by the asymptotic $\lim_{x \to \infty} f(x)$. To understand the growth of f, we often analyze the asymptotic of partial summation

$$\lim_{x \to \infty} \sum_{n \leq x} f(n)$$

For example, the prime counting function is

$$\pi(x) = \sum_{n \leq x} C(n)$$

where $C(n)$ is the characteristic function of n, that is, $C(n) = 1$ if n is a prime otherwise $C(n) = 0$. One of the biggest questions we will try to answer is how $\lim_{x \to \infty} \pi(x)$ behaves.

Summatory Function. For an arithmetic function f, the *summatory function* of f is defined as

$$F(n) = \sum_{d \in S} f(d)$$

where S is some set possibly dependent on n. When S is the set of divisors of n, the number of divisor function $\tau(n)$ is the summatory function of the unit function $u(n) = 1$ and the sum of divisor function $\sigma(n)$ is the summatory function of the invariant function $f(n) = n$. Another summatory function is the partial summation

$$\sum_{n \leq x} f(n)$$

Associated with this is the average order of f.

Average Order. For an arithmetic function f,

$$\lim_{x \to \infty} \frac{\sum_{n \leq x} f(x)}{x}$$

is the *average order.* In this context, a very interesting way of analyzing growth is the *normal order* of f. The concept of normal numbers arises from Hardy and Ramanujan.[1]

Normal Order. Let f and F be arithmetic functions such that

(‡1) $$(1 - \epsilon)F(n) < f(n) < (1 + \epsilon)F(n)$$

holds for *almost* all $n \leq x$ as $x \to \infty$. Then we say that F is the *normal order* of f. A trivial(?) example of normal order is that almost all positive integers not exceeding x are composite if x is sufficiently large. We should probably elaborate on what we mean by *almost* here. One interpretation is that the number of positive integers not exceeding x which are prime is very small compared to x. Similarly, f is of order F means that the number of positive integers n not exceeding x which do not satisfy (‡1) is very small compared to x.

An interesting property in summatory functions is that

$$\sum_{i=1}^{n} F(i) = \sum_{i=1}^{n} \sum_{d|i} f(d)$$

$$= \sum_{i=1}^{n} \left[\frac{n}{i}\right] f(i)$$

Here, the last equation is true because there are $[n/i]$ multiples of i not exceeding n.

§1. Order of Some Arithmetic Functions

Recall that the number of divisor function

$$\tau(n) = \sum_{ab=n} 1$$

We can generalize this as follows.

---

[1]Godfrey Harold Hardy and Srinivasa Ramanujan
1917 "The normal number of prime factors of a number n", *Quarterly Journal of Mathematics*, 48, pp. 76-92.

Generalized Number of Divisors. The *generalized number of divisor* function is defined as

$$\tau_k(n) = \sum_{d_1 \cdots d_k = n} 1$$

So $\tau_k(n)$ is the number of ways to write n as a product of k positive integers. Similarly, we can take the sum of divisor function and generalize it.

Generalized Sum of Divisors. The *generalized sum of divisor* function can be defined as

$$\sigma_k(n) = \sum_{d|n} d^k$$

At this point, we should discuss some asymptotic notions.

Big O. Let f and g be two real or complex valued functions. We say that

$$f(x) = O(g(x))$$

if there is a positive real constant C such that

$$|f(x)| \leq Cg(x)$$

for all sufficiently large x. It is also written as $f(x) \ll g(x)$ or $g(x) \gg f(x)$. When we say g is an asymptotic estimate of f, we mean that

$$f(x) = g(x) + O(h(x))$$

for two functions g and h as $x \to \infty$. Here, h is the **error term** which obviously should be of lower magnitude than g. In particular, $f(x) = O(1)$ means that f is bounded above by some positive constant. Some trivial examples are $x^2 = O(x^3)$, $x + 1 = O(x)$ and $x^2 + 2x = O(x^2)$. We usually want $g(x)$ to be as small as possible to avoid triviality. A useful example is

$$\lfloor x \rfloor = x + O(1)$$

since $x = \lfloor x \rfloor + \{x\}$ and $0 \leq \{x\} < 1$.

Small O. Let f and g be two real or complex valued functions. Then the following two statements are equivalent

(‡2)
$$f(x) = o(g(x))$$

(‡3)
$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0$$

Some trivial examples are $1/x = o(1)$, $x = o(x^2)$ and $2x^2 \neq o(x^2)$. Landau[2] states that

the symbol O had been first used by Bachmann.[3] Hardy[4] uses the notations $\prec$ and $\succ$ respectively but they are no longer in practice. Hardy and Riesz[5] adopted the notations small o and big O and today these are the primary notations for this purpose.

It should be evident that having an estimate with respect to O asymptotic formulas is more desirable than o formulas. By nature, O formulas give us a better understanding and a specific estimate whereas o does not always say as much. Moreover, working with O is a lot easier than working with o. For example,

$$\sum O(f(x)) = O(\sum f(x))$$
$$\int O(f(x))dx = O(\int f(x)dx)$$

Or consider the possibility that we can very easily deal with constants that would otherwise pop up here and there unnecessarily. With the help of O,

$$O(1) + c = O(1)$$
$$O(cf(x)) = O(f(x))$$

and so on.

Equivalence. Let f and g be two real or complex valued functions. We say that they are *asymptotically equivalent* if

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1$$

and we denote it by $f \sim g$. So, we can say that g is an asymptotic formula for f. An example is $x^2 \sim x^2 + x$. Another example in connection with normal order is that f has normal order F if the number of n not satisfying (‡1) is $o(x)$. We can also say, the number of n satisfying (‡1) is $\sim x$. Note the following.

$$f \sim g \iff |f(x) - g(x)| = o(g(x))$$

We will use these symbols extensively throughout the book. It is of utmost importance that the reader gets well familiarized with these notions since they will be crucial in

[2]Edmund Landau

1909 *Handbuch der Lehre von der Verteilung der Primzahlen*, vol. 2, Page 883 (second volume is paged consecutively after first volume).

[3]Paul Gustav Heinrich Bachmann

1894 *Analytische zahlentheorie*, vol. 2, Page 401.

[4]Godfrey Harold Hardy

1910 *Orders of Infinity: The 'Infinitärcalcül' of Paul Du Bois-Reymond, Cambridge Tracts in Mathematics*, Cambridge University Press.

[5]Godfrey Harold Hardy and Marcel Riesz

1915 *The general theory of Dirichlet's series*, Cambridge University Press.

understanding much of this book. The primary motivation behind these asymptotic notions is to get an as precise as possible idea about the ***order of magnitude*** of a certain function. This is why we will be leaning more towards $x^2 + 2x = O(x^2)$ than $x^2 + 2x = O(x^3)$ even though both are mathematically correct. The reason is, even though $x^2 + 2x = O(x^3)$ is true, it is taking away a great portion of the accuracy to which we suppose $x^2 + 2x$ should be measured with. On the other hand, we easily see that we cannot have $x^2 + 2x = O(x^\epsilon)$ for $\epsilon < 2$. Under the same philosophy, we define the order of magnitude equivalence.

Definition. If f and g are functions such that both $f(x) \ll g(x)$ and $g(x) \ll f(x)$ hold, then we write $f \asymp g$ and say that f and g have the ***same order of magnitude.***

Now, we are interested in the order of general number of divisors and general sum of divisors. Let us define the cumulative sum of these functions.

$$S_k(x) = \sum_{n \leq x} \sigma_k(n)$$

$$T_k(x) = \sum_{n \leq x} \tau_k(n)$$

Notice the following.

$$S_k(x) = \sum_{n \leq x} \sum_{d \mid n} d^k$$

$$= \sum_{n \leq x} \left\lfloor \frac{x}{n} \right\rfloor n^k$$

$$= \sum_{n \leq x} \left(\frac{x}{n} + O(1) \right) n^k$$

$$= x \sum_{n \leq x} n^{k-1} + O\left(\sum_{n \leq x} n^k \right)$$

We can use this to establish an asymptotic for $T_k(x)$ if we can establish the asymptotic of $A_2(x)$. We will get to that in a moment. First, let us take care of the summation within the big O bracket. We have the trivial inequality that

$$\sum_{n \leq x} n^k \leq \sum_{n \leq x} x^k$$

$$= x^k \sum_{n \leq x} 1$$

$$= \lfloor x \rfloor x^k$$

$$= (x + O(1)) x^k$$

$$= x^{k+1} + O(x^k)$$

We have that $S_k(x) = x(x^k + O(x^{k-1})) + O(x^{k+1}) = O(x^{k+1})$. Although weak, we get an estimate this way. On this note, an interested reader can try and prove that

$$(n+1)^{k+1} - 1 = \sum_{i=0}^{k} \binom{k+1}{i} \mathfrak{S}(n, i)$$

where $\mathfrak{S}(x, k) = \sum_{n \leq x} n^k$. This is known as the **Pascal identity** (see Pascal,[6] for an English translation, see Knoebel et al.[7]). Lehmer[8] proves that

$$(\ddagger 4) \qquad\qquad \mathfrak{S}(x, k) = \frac{x^{k+1}}{k+1} + \Delta$$

where $|\Delta| \leqq x^k$. The reader may also be interested in MacMillan and Sondow.[9]

We shall try to estimate T in a similar fashion. First, see that

$$
\begin{aligned}
\tau_k(n) &= \sum_{d_1 \cdots d_k = n} 1 \\
&= \sum_{d_k | n} \sum_{d_1 \cdots d_{k-1} = n/d_k} 1 \\
&= \sum_{d|n} \tau_{k-1}\left(\frac{n}{d}\right)
\end{aligned}
$$

Note that the two sets $\{d : d \mid n\}$ and $\{n/d : d \mid n\}$ are actually the same. So, we get

$$\tau_k(n) = \sum_{d|n} \tau_{k-1}(d)$$

[6]Blaise Pascal

1964 “Sommation des puissances numériques”, *Oeuvres complètes, Jean Mesnard, ed., Desclée-Brouwer, Paris,* 3, pp. 341-367.

[7]Arthur Knoebel et al.

2007 “Sums of numerical powers”, in, *Mathematical Masterpieces: Further chronicles by the explorers,* Springer-Verlag, pp. 32-37.

[8]Derrick Norman Lehmer

1900 “Asymptotic evaluation of certain Totient Sums”, *American Journal of Mathematics*, 22, 4, pp. 293-335, DOI: 10.2307/2369728, Chapter II, Theorem 1.

[9]Kieren MacMillan and Jonathan Sondow

2011 “Proofs of power sum and binomial coefficient congruences via Pascal's identity”, *The American Mathematical Monthly*, 118, 6, pp. 549-551, DOI: 10.4169/amer.math.monthly.118.06.549.

Beumer[10] also considers the generalization $\tau_k(n)$ in this exact form. Using this for T,

$$
\begin{aligned}
T_k(x) &= \sum_{n \leq x} \tau_k(n) \\
&= \sum_{n \leq x} \sum_{d \mid n} \tau_{k-1}(d) \\
&= \sum_{n \leq x} \left[\frac{x}{n} \right] \tau_{k-1}(n) \\
&= \sum_{n \leq x} \left(\frac{x}{n} + O(1) \right) \tau_{k-1}(n) \\
&= x \sum_{n \leq x} \frac{\tau_{k-1}(n)}{n} + O\left(\sum_{n \leq x} \tau_{k-1}(n) \right)
\end{aligned}
$$

Thus, we have the recursive result

$$
T_k(x) = x \sum_{n \leq x} \frac{\tau_{k-1}(n)}{n} + O(T_{k-1}(x))
$$

It gets nontrivial how to proceed from here. Consider the *harmonic sum*

$$
H(x) = \sum_{n \leq x} \frac{1}{n}
$$

It does not seem easy to calculate H accurately, however, we can make a decent attempt to estimate H. The tool that is best suited for carrying out such an estimation is the *Abel partial summation formula.* Abel[11] states this formula which today is a cornerstone of analytic number theory.

Theorem 1 (Abel partial summation formula). *Let $\{a_n\}$ be a sequence of real numbers and f be a continuous differentiable function in $[y, x]$. If the partial sums of $\{a_n\}$ is*

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

are known, then

$$
\sum_{y < n \leq x} a_n f(n) = A(x)f(x) - A(y)f(y) - \int_y^x A(t)f'(t)dt
$$

---
[10]Martin Beumer
1962 "The arithmetical function $\tau_k(n)$", *The American Mathematical Monthly*, 69, 8, pp. 777-781, DOI: 10.2307/2310778, (§8).
[11]Niels Henrik Abel
1826 "Untersuchungen über die Reihe: $1 + (m/1)x + m \cdot (m-1)/(1 \cdot 2) \cdots x^2 + m \cdot (m-1) \cdot (m-2)/(1 \cdot 2 \cdot 3) \cdots x^3 + \ldots$", *Journal für Math.*, 1, pp. 311-339, DOI: 10.1515/9783112347386-030.

In particular, if f is an arithmetic function,

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

Proof. □

It is not straightforward to realize how such a formula can be as influential as we are describing it to be. Notice that, the formula essentially converts a discreet sum into an integral, which occasionally may be calculable. If the integral is not calculable, we may be able to estimate its value sometimes. We should mention that Ramanujan[12] also uses a method that can only be described as the partial summation formula. It is unclear if Ramanujan simply knew about this. He essentially derives the partial summation formula while trying to express a sum of the form

$$\sum_{p \leq x} \phi(p)$$

with respect to $\pi(x)$, $\phi(x)$ and an integral where $\pi(x)$ is the number of primes not exceeding x. A consequence of Abel partial summation formula is the celebrated *Euler's summation formula.*

Theorem 2 (Euler's summation formula). *Let f be a continuous differentiable function in* $[y, x]$. *Then*

$$\sum_{y < n \leq x} f(n) = \int_y^x f(t)dt + \int_y^x \{t\}f'(t)dt + \{y\}f(y) - \{x\}f(x)$$

where $\{t\} = t - [t]$ *is the fractional part of* t.

Proof. □

As an application of Euler's summation formula, we can derive a result similar to (‡4) taking $f(n) = n^k$ for $k \geq 0$.

$$\begin{aligned}
\mathfrak{S}_k(x) &= \sum_{n \leq x} n^k \\
&= \int_1^x t^k dt + k\int_1^x t^{k-1}(t - [t])dt + 1 - (x - [x])x^k \\
&= \frac{x^{k+1}}{k+1} - \frac{1}{k+1} + O\left(k\int_1^x t^{k-1}dt\right) + O(x^k) \\
&= \frac{x^{k+1}}{k+1} + O(x^k)
\end{aligned}$$

---

[12]Srinivasa Ramanujan

1927 "Highly Composite Numbers", in, *Collected papers of Srinivasa Ramanujan,* ed. by Godfrey Harold Hardy et al., Cambridge University Press, pp. 78-128, Page 83, §4.

Setting $a_n = \tau_{k-1}(n)$ and $f(n) = 1/n$ in Abel partial summation formula, we get

$$\sum_{n \leq x} \frac{\tau_{k-1}(n)}{n} = \frac{T_{k-1}(x)}{x} - \int_1^x -\frac{T_{k-1}(t)}{t^2} dt$$

Thus, we have a result where we can inductively get to the final expression. First, let us see the case $k = 2$.

$$\sum_{n \leq x} \tau(n) = \sum_{n \leq x} \left[\frac{x}{n}\right]$$

Clearly, this is just the number of pairs (a, b) such that $ab \leq x$. We can divide the pairs in two classes. In the first class, $1 \leq a \leq \sqrt{x}$ and in the second one, $a > \sqrt{x}$. In the first case, for a fixed a, there are $[x/a]$ possible choices for a valid value of b. So, the number of pairs in the first case is

$$\sum_{a \leq \sqrt{x}} \left[\frac{x}{a}\right]$$

In the second case, since $a > \sqrt{x}$ and $b \leq x/a$, we must have $b \leq \sqrt{x}$. For a fixed b, there are $[x/b] - \sqrt{x}$ choices for a valid value of a, the choices namely are

$$[x] + 1, \ldots, \left[\frac{x}{b}\right]$$

Then the number of pairs in this case is

$$\sum_{b \leq \sqrt{x}} \left[\frac{x}{b}\right] - [\sqrt{x}]$$

Thus, the total number of such pairs is

(‡5)
$$\sum_{a \leq \sqrt{x}} \left[\frac{x}{a}\right] + \sum_{b \leq \sqrt{x}} \left(\left[\frac{x}{b}\right] - [\sqrt{x}]\right) = 2 \sum_{n \leq \sqrt{x}} \left[\frac{x}{n}\right] - [x]^2$$

For getting past this sum, we have to deal with the sum

$$\sum_{n \leq \sqrt{x}} [x/n] = \sum_{n \leq \sqrt{x}} \left(\frac{x}{n} + O(1)\right)$$

$$= x \sum_{n \leq \sqrt{x}} \frac{1}{n} + O(\sqrt{x})$$

$$= xH(\sqrt{x}) + O(\sqrt{x})$$

Setting $a_n = 1$ and $f(n) = 1/n$ in Abel partial summation formula, we get

$$H(x) = \frac{A(x)}{x} - \int_1^x -\frac{A(t)}{t^2} dt$$

Here, $A(x) = \lfloor x \rfloor = x + O(1)$. Using this,

$$H(x) = 1 + O\left(\frac{1}{x}\right) + \int_1^x \left(\frac{1}{t} + \frac{O(1)}{t^2}\right) dt$$

$$= 1 + O\left(\frac{1}{x}\right) + \int_1^x \frac{1}{t} dt + O\left(\int_1^x \frac{1}{t^2} dt\right)$$

$$= 1 + O\left(\frac{1}{x}\right) + \log x + O\left(1 - \frac{1}{x}\right)$$

Thus, we have the following result.

Theorem 3 (Divergence of Harmonic Sum). *For $x \geqq 1$,*

$$H(x) = \log x + C + O\left(\frac{1}{x}\right)$$

where C is a constant.

We get a more precise formulation of $H(x)$ by considering the limit $x \to \infty$ which removes $O(1/x)$ from the expression since this limit would be 0.

Theorem 4. *There is a constant γ such that*

$$\gamma = \lim_{x \to \infty} (H(x) - \log x)$$

This constant γ is now known as ***Euler's constant*** or ***Euler-Mascheroni's constant***, although, neither Euler nor Mascheroni used the notation γ for this constant. Euler[13] (republished in Euler[14]) used C and O in his original paper. Mascheroni[15] used A and a. Today it is not known whether γ is even irrational. For now, we will not require the use

---

[13]Leonhard Euler

1740   "De Progressionibus Harmonicis Observationes", *Commentarii academiae scientiarum Petropolitanae*, 7, pp. 150-161.

[14]Leonhard Euler

2020b   "E-43: De Progressionibus Harmonicis Observationes", *Spectrum*, pp. 133-141, DOI: 10.1090/spec/098/23.

[15]Lorenzo Mascheroni

1790   *Adnotationes ad calculum Integralem Euleri*, Galeatii.

of γ, so we will use Divergence of Harmonic Sum. Applying this, we have

$$\sum_{n \leq \sqrt{x}} \left[\frac{x}{n}\right] = xH(\sqrt{x}) + O(\sqrt{x})$$

$$= x\left(C + \log \sqrt{x} + O\left(\frac{1}{\sqrt{x}}\right)\right) + O(\sqrt{x})$$

$$= \frac{1}{2}x \log x + Cx + O\left(\frac{x}{\sqrt{x}}\right) + O(\sqrt{x})$$

$$= \frac{1}{2}x \log x + O(x)$$

We can now use this to get

$$\sum_{n \leq x}' \tau(n) = 2 \sum_{n \leq \sqrt{x}} [x/n] - \left[\sqrt{x}\right]^2$$

$$= x \log x + O(x)$$

Thus, we get the following result.

$$\frac{\sum_{n \leq x} \tau(n)}{x} = \log x + O(1)$$

Dirichlet[16] actually proves the more precise result given below.

Theorem 5 (Dirichlet's average order of τ theorem).

$$\frac{\sum_{n \leq x} \tau(n)}{x} = \log x + 2\gamma - 1 + O\left(\frac{1}{\sqrt{x}}\right)$$

where γ is the Euler-Mascheroni constant.

Then Dirichlet's theorem on τ can be restated as *the average order of τ is $O(\log x)$*. Ramanujan[17] points out in his paper that the error term $O(1/\sqrt{x})$ in Dirichlet's theorem can be improved to $O(x^{-2/3+\epsilon})$ or $O(x^{-2/3} \log x)$ as Landau[18] shows.

---

[16]Johann Peter Gustav Lejeune Dirichlet

1897b "Über Die Bestimmung Der Mittleren Werthe", in, *G. Lejeune Dirichlet's Werke*, ed. by Leopold Kronecker and László Fuchs, Druck Und Verlag Von Georg Reimer., vol. 2, pp. 49-66, DOI: 10.1017/cbo9781139237345.007.

[17]Ramanujan, "Highly Composite Numbers" cit.

[18]Edmund Landau

1912a "Über die Anzahl der Gitterpunkte in geweissen Bereichen", *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 19, pp. 687-772, Page 689.

We can now get back to estimating T. Using Abel partial summation formula, we were able to deduce

$$T_k(x) = O(T_{k-1}(x)) + x\int_1^x \frac{T_{k-1}(t)}{t^2} dt$$

Using Dirichlet's average order of τ theorem, $T(x) = x\log x + O(x)$, so

$$T_3(x) = O(T(x)) + x\int_1^x \frac{T(t)}{t^2} dt$$

$$= O(x\log x) + x\int_1^x \frac{\log t + O(1)}{t} dt$$

$$= O(x\log x) + x\int_1^x \frac{\log t}{t} dt + xO\left(\int_1^x \frac{1}{t} dt\right)$$

$$= O(x\log x) + x\int_1^x \frac{\log t}{t} dt$$

Using integration by parts,

$$\int \frac{\log t}{t} dt = \log t \int \frac{1}{t} - \int \left(\frac{1}{t}\int \frac{1}{t} dt\right) dt$$

$$= \log^2 t - \int \frac{\log t}{t} dt$$

Thus, we get

$$\int_1^x \frac{\log t}{t} dt = \frac{1}{2}\log^2 x$$

which in turn gives

$$T_3(x) = \frac{1}{2}x\log^2 x + O(x\log x)$$

We leave it as an exercise for the reader to prove the following (from what we have already developed, induction is one way to go about it).

Theorem 6. *Let k be a positive integer. Then*

$$T_k(x) = \frac{1}{(k-1)!}x\log^{k-1} x + O\left(x\log^{k-2} x\right)$$

The reason we do not write $T_k(x)$ as $O\left(x\log^{k-1}x\right)$ directly is because in this case, we already know the constant multiplier of $x\log^{k-1}x$ which is not ugly. Usually, we write $O(f(x))$ when we do not know what the constant multiplier of $f(x)$ is or when it gets too big to keep track of. Landau[19] states a sharper result.

$$T_k(x) = x\left(\sum_{m=0}^{k-1} b_m \log^m x\right) + O\left(x^{1-\frac{1}{k}}\right) + O\left(x^{1-\frac{1}{k}}\log^{k-2}x\right)$$

Let us now turn our attention to improving the asymptotic of $S_k(x)$.

$$\begin{aligned}
S_k(x) &= \sum_{n\leq x}\sum_{d|n} d^k \\
&= \sum_{n\leq x}\sum_{m\leq x/n} m^k \\
&= \sum_{n\leq x} \mathfrak{S}_k\left(\frac{x}{n}\right) \\
&= \sum_{n\leq x} \frac{x^{k+1}}{(k+1)n^{k+1}} + O\left(\frac{x^k}{n^k}\right) \\
&= \frac{x^{k+1}}{k+1}\sum_{n\leq x}\frac{1}{n^{k+1}} + O\left(x^k\sum_{n\leq x}\frac{1}{n^k}\right)
\end{aligned}$$

Here, we can see that the function

$$\sum_{n\leq x}\frac{1}{n^k}$$

occurs repeatedly. It is in fact, the partial sum of the famous Euler's *zeta function.*

Zeta Function.   For a complex number s, the zeta function $\zeta(s)$ is defined as

$$\zeta(s) = \sum_{n\geq 1}\frac{1}{n^s}$$

We will discuss zeta function in details in Section *§2..* For now, let us establish a result similar to Dirichlet's average order of τ theorem for partial sums of ζ. Setting $f(n) := n^{-s}$

[19]Edmund Landau

1912b "Über eine idealtheoretische funktion", *Transactions of the American Mathematical Society*, 13, 1, pp. 1-21, DOI: 10.1090/s0002-9947-1912-1500901-6, Page 2.

and $a_n = 1$ in Abel partial summation formula, $A(x) = \lfloor x \rfloor = x + O(1)$ and

$$\sum_{n \leq x} \frac{1}{n^s} = \lfloor x \rfloor x^{-s} - \int_1^x (t + O(1)) f'(t) dt$$

$$= x^{1-s} + O(x^{-s}) + s \int_1^x t^{-s} dt + O\left(s \int_1^x t^{-s-1} dt \right)$$

$$= x^{1-s} + \frac{s}{1-s}(x^{1-s} - 1) + O\left(\int_1^x t^{-s-1} dt \right)$$

$$= \frac{x^{1-s}}{1-s} + C + O(x^{-s})$$

Similar to γ, we can take $x \to \infty$ and get the following result.

Theorem 7. *Let s be a positive real number other than 1. Then*

$$\sum_{n \leq x} \frac{1}{n^s} = \frac{x^{1-s}}{1-s} + C + O(x^{-s})$$

where C is a constant similar to Euler-Mascheroni constant dependent on s and

$$C = \lim_{x \to \infty} \left(\sum_{n \leq x} \frac{1}{n^s} - \frac{x^{1-s}}{1-s} \right)$$

Furthermore, if $0 < s < 1$, then $C = \zeta(s)$ since $x^{1-s} \to 0$.

We can now get back to estimating $S_k(x)$.

$$S_k(x) = \frac{x^{k+1}}{k+1} \sum_{n \leq x} \frac{1}{n^{k+1}} + O\left(x^k \sum_{n \leq x} \frac{1}{n^k} \right)$$

$$= \frac{x^{k+1}}{k+1}\left(\frac{x^{-k}}{-k} + \zeta(k+1) + O(x^{-k-1}) \right) + O\left(x^k \left(\frac{x^{1-k}}{1-k} + \zeta(k) + O(x^{-k}) \right) \right)$$

$$= \frac{x}{-k(k+1)} + \frac{x^{k+1}}{k+1}\zeta(k+1) + O(x^{k+1-k-1}) + \left(\frac{x}{1-k} + x^k\zeta(k) + O(1) \right)$$

$$= \frac{x^{k+1}}{k+1}\zeta(k+1) + O(x) + O(1) + O(x + x^k)$$

From this, we finally get the following.

Theorem 8. *Let k be a positive integer. Then*

$$S_k(x) = \frac{x^{k+1}}{k+1}\zeta(k+1) + O(x^{\max(1,k)})$$

We leave the case when k is a negative integer as an exercise. Next, we consider a generalization of the Euler's totient function $\varphi(n)$.

$$\varphi(x, a) = \sum_{\substack{n \leq x \\ \gcd(n,a)=1}} 1$$

For a positive integer n, $\varphi(n) = \varphi(n, n)$ and **Jordan function** is a generalization of φ.

Jordan Function. Let n and k be positive integers. Then the Jordan function $J_k(n)$ is the number of k tuples of positive integers not exceeding n that are relatively prime to n.

$$J_k(n) = \sum_{\substack{1 \leq a_1, \ldots, a_k \leq n \\ \gcd(a_1, \ldots, a_k, n)=1}} 1$$

Lehmer[20] used the notation $\varphi_k(n)$ but today $J_k(n)$ is used more often. Jordan[21] first discussed this function and Lehmer[22] developed some asymptotic results. Jordan totient function is interesting not only because it is a generalization of Euler's totient function but also because it has many interesting properties. For example, similar to φ, we can show that

$$J_k(n) = \prod_{p^e \| n} p^{k(e-1)}(p - 1)$$

$$J_k(n^m) = n^{k(m-1)} J_k(n)$$

Lehmer[23] proves the following which he calls the **fundamental theorem.**

($\ddagger 6$) $$\qquad\qquad J_k(mn) = J_k(n) \prod_{p^e \| m} \left(p^{ke} - p^{k(e-1)} \lambda(n, p) \right)$$

where $\lambda(n, p) = 0$ if $p \mid n$ otherwise $\lambda(n, p) = 1$. We leave the proof of this result and the following to the reader.

($\ddagger 7$) $$\qquad\qquad \sum_{d|n} J_k(d) = n^k$$

Like $\sigma_k(n)$, $J_k(n)$ is also related to the sum $\mathfrak{S}(x, k)$. But we do not derive the order of $J_k(n)$ yet.

---

[20]Lehmer, "Asymptotic evaluation of certain Totient Sums" cit.
[21]Camille Jordan
1870 *Traiteé des substitutions et des équations algébriques,* Gauthier-Villars, Paris, Page $95 - 97$.
[22]Lehmer, "Asymptotic evaluation of certain Totient Sums" cit.
[23]Ibid., Theorem VI.

§2. Dirichlet Series and Dirichlet Convolution

We encountered ζ when we tried to develop an asymptotic for $S_k(x)$. The function ζ has quite a rich history. Today ζ is mostly called Riemann's zeta function, however, Euler is the first one to investigate this function. Euler started working on ζ around 1730. During that period, the value of $\zeta(2)$ was unknown and of high interest among prominent mathematicians. Ayoub[24] is a very good read on this subject. Euler's first contribution in this matter is Euler[25] where he proves that $\zeta(2) \approx 1.644934$. The paper was first presented to the St. Petersburg Academy on March 5, 1731 and republished in Euler.[26] Euler[27] (republished in Euler[28]) proves the following fundamental result which essentially gives a new proof of infinitude of primes.

Theorem 9 (Euler Product Identity). *Let s be a positive integer. Then*

$$\zeta(s) = \prod_p \frac{p^s}{p^s - 1}$$

where p extends over all primes.

One of the results in Euler[29] is the following which we shall prove later.

$$\sum_{n \leq x} \frac{1}{p} \sim \log \sum_{n \leq x} \frac{1}{x}$$

[24]Raymond Ayoub

1974 "Euler and the zeta function", *The American Mathematical Monthly*, 81, 10, pp. 1067-1086, DOI: 10.2307/2319041.

[25]Leonhard Euler

1738 "De summatione innumerabilium progressionum", *Commentarii academiae scientiarum Petropolitanae*, 5, pp. 91-105.

[26]Leonhard Euler

2020a "E-20: De summatione innumerabilium Progressionum", *Spectrum*, pp. 52-64, DOI: 10.1090/spec/098/10.

[27]Leonhard Euler

1744 "Variae Observationes circa series infinitas", *Commentarii academiae scientiarum Petropolitanae*, 9, pp. 160-188.

[28]Leonhard Euler

2020c "E-72: Variae Observationes circa series Infinitas", *Spectrum*, pp. 249-260, DOI: 10.1090/spec/098/41.

[29]Euler, "Variae Observationes circa series infinitas" cit.

Here, $\sim$ is the asymptotic equivalence we have already defined. Even though Euler is the main architect behind the development of ζ, Riemann[30] is the first one to consider ζ for complex s instead of real s only. By tradition, we write $s = \sigma + it$ where $\sigma = \Re(s)$ is the real part of s and $t = \Im(s)$ is the imaginary part of s.

Dirichlet Series.   For a complex number s, a **Dirichlet series** is a series of the form

$$\mathfrak{D}_a(s) = \sum_{n \geq 1} \frac{a(n)}{n^s}$$

So, ζ is a special case of $\mathfrak{D}$ when $a(n) = 1$ for all n. Hardy and Riesz[31] considers the following as *general Dirichlet series*

(‡8)
$$\sum_{n \geq 1} a_n e^{-\lambda_n s}$$

where (λ_n) is a strictly increasing sequence of real numbers that tend to infinity. Following this, Hardy and Riesz[32] calls $\mathfrak{D}$ the *ordinary Dirichlet series* when $\lambda_n = \log n$. Dirichlet[33] considers real values of s and proves a number of important theorems. As Hardy states, Jensen[34] discusses the first theorems where s is complex involving the nature of convergence of ‡8. Cahen[35] makes the *first attempt to construct a systematic theory of the function $\mathfrak{D}_f(s)$ although much of the analysis which it contains is open to serious criticism, has served—and possibly just for that reason—as the starting point of most of the later researches in the subject.*

Euler Product.   Let s be a complex number and f be a bounded multiplicative function. Then **Euler product** is a special case of Dirichlet series that can be written as

$$\prod_p \sum_{i \geq 1}' \frac{f(p^i)}{p^{is}}$$

[30]Bernhard Riemann

1859   "Ueber die anzahl der primzahlen unter einer gegebenen grösse", *Monatsberichte der Berliner Akademie* (Nov. 1859), pp. 136-144, DOI: 10.1017/cbo9781139568050.008.

[31]Hardy and Riesz, *The general theory of Dirichlet's series* cit., §1, Page 1.

[32]Ibid.

[33]Johann Peter Gustav Lejeune Dirichlet

1879   *Vorlesungen Über Zahlentheorie*, ed. by R. Dedekind, Cambridge University Press.

[34]Johan Ludwig William Valdemar Jensen

1884   "OM RÆKKERS KONVERGENS", *Tidsskrift for mathematik*, 5th ser., 2, pp. 63-72, ISSN: 09092528, 24460737, http://www.jstor.org/stable/24540057; Johan Ludwig William Valdemar Jensen

1888   "Sur une généralisation d'un théorème de Cauchy", *Comptes Rendus* (Mar. 1888).

[35]Eugène Cahen

1894   "Sur la fonction $\zeta(s)$ de Riemann et sur des fonctions analogues", fr, *Annales scientifiques de l'École Normale Supérieure*, 11, pp. 75-164, DOI: 10.24033/asens.401.

where p extends over all primes. From the fundamental theorem of arithmetic,

$$\sum_{n \geq 1} \frac{f(n)}{n^s} = \prod_p \sum_{i \geq 1} \frac{f(p^i)}{p^{is}}$$

If f is completely multiplicative, then the sum inside the product becomes a geometric series and we have

$$\sum_{n \geq 1} \frac{f(n)}{n^s} = \prod_p \frac{1}{1 - \dfrac{f(p)}{p^s}}$$

Consider the Dirichlet series for two arithmetic functions f and g.

$$\mathfrak{D}_f(s) = \sum_{n \geq 1} \frac{f(n)}{n^s}$$

$$\mathfrak{D}_g(s) = \sum_{n \geq 1} \frac{g(n)}{n^s}$$

Then we have

$$\mathfrak{D}_f(s)\mathfrak{D}_g(s) = \sum_{n \geq 1} \frac{f(n)}{n^s} \sum_{n \geq 1} \frac{g(n)}{n^s}$$

Now, imagine we want to write this product as another Dirichlet series. Then it would be of the form

$$\mathfrak{D}_h(s) = \sum_{n \geq 1} \frac{h(n)}{n^s}$$

The coefficients $h(n)$ of $\mathfrak{D}_h(s)$ is determined as follows.

$$h(n) = \sum_{de = n} f(d)g(e)$$

After a little observation, it seems quite obvious that this is indeed correct. In fact, this is what we call Dirichlet convolution today.

Dirichlet Convolution.   For two arithmetic functions f and g, the **Dirichlet product** or **Dirichlet convolution** of f and g is defined as

$$f * g = \sum_{d \mid n} f(d)g\left(\frac{n}{d}\right)$$

Theorem 10. *Let f and g be multiplicative arithmetic functions. Then $f * g$ is also multiplicative.*

Proof. □

Theorem 11 (Associativity of Dirichlet Convolution). *Dirichlet convolution is associative. That is, if f, g and h are arithmetic functions, then*

$$(f * g) * h = f * (g * h)$$

Proof. □

An interesting function associated with Dirichlet convolution and summatory functions is the **Möbius function** μ, defined in Möbius.[36]

$$\mu(n) = \begin{cases} 0 & \text{if } p^2 \mid n \text{ for some prime } p \\ (-1)^{\omega(n)} & \text{otherwise} \end{cases}$$

where $\omega(n)$ is the number of distinct prime divisors of n. On the other hand, $\Omega(n)$ is the total number of prime divisors of n. So, $\omega(12) = 2$ whereas $\Omega(12) = 3$.

Theorem 12 (Möbius Inversion). *Let f be an arithmetic function and F be the summatory function*

$$F(n) = \sum_{d\mid n} f(d)$$

Then

$$f(n) = \sum_{d\mid n} \mu(d) F\left(\frac{n}{d}\right)$$

Proof. It it known that $*$ is associative. We have $f * f^{-1} = I$ where

$$I(n) = \left\lfloor \frac{1}{n} \right\rfloor$$

Note that,

$$\sum_{d\mid n} \mu(d) = I(n)$$

We can write it as $\mu * u = I$ where $u(n) = 1$ is the unit function. Then $\mu^{-1} = u$ and $u^{-1} = \mu$. Now, we can write $F(n) = \sum_{d|n} f(d)$ as $F = f * u$. Multiplying both sides to the left by μ,

$$
\begin{aligned}
F * \mu &= (f * u) * \mu \\
&= f * (u * \mu) \\
&= f * I \\
&= f
\end{aligned}
$$

since $f * I = f(n)$. Expanding this, we have the result.   □

Following Cojocaru and Murty,[37] let us define **dual convolution**.

Divisor Closed Set.   A set of positive integers S is a **divisor closed set** if $d \mid n$, then $d \in$ S holds for all $n \in$ S.

Dual Convolution.   Let f and g be arithmetic functions. Then the dual convolution of f and g is the arithmetic function h defined as

$$
h(n) = \sum_{\substack{n|d \\ d \in D}} f(d) g\left(\frac{d}{n}\right)
$$

where D is a divisor closed set.

Theorem 13 (Dual Möbius Inversion). *Let f be an arithmetic function and F be the summatory function*

$$
F(n) = \sum_{\substack{n|d \\ d \in D}} f(d)
$$

where D *is a divisor closed set. Then*

$$
f(n) = \sum_{\substack{n|d \\ d \in D}} \mu\left(\frac{d}{n}\right) f(d)
$$

Proof. This is not as difficult as it looks. We mainly need to look at D_n, the set of divisors of n for $n \in$ D. If $m \mid n$, then $D_m \in D_n$. Let $M(n) = \max\{m \in D : n \mid m\}$ for $n \in$ D, $N(n) = M(n)/n$ and $P(n) = \prod_{p|N(n)} p$. Since

$$
\begin{aligned}
F(n) &= \sum_{\substack{n|d \\ d \in D}} f(d) \\
&= \sum_{k|N(n)} f(nk)
\end{aligned}
$$

───────────────

[37]Cojocaru and Murty, *An introduction to sieve methods and their applications* cit., Page 4, Theorem 1.2.3.

For a prime p, if $\nu_p(k) > 1$, then $D_{nk} \in D_{np}$, so we don't need to consider any of $F(nk)$ separately for $nk \in D$. We only need to consider the set of sets

$$\{D_{nq} : q \mid P(n)\}$$

Note that for distinct $q, r \in D_{P(n)}$, $D_{nq} \cap D_{nr} = D_{nqr}$. Thus, we can easily see that

$$\sum_{n \mid d} f(d) \mu\left(\frac{d}{n}\right) = \sum_{q \mid P(n)} f(nq) \mu(q)$$

From this, it is pretty obvious that unless $q = 1$, all the other terms cancel out. Indeed, for $q \mid P(n), f(nq)$ appears $\binom{\omega(P(n))}{\omega(q)}$ times in the sum $\sum_{q \mid P(n)} f(nq) \mu(q)$ and

$$\binom{m}{0} + \binom{m}{2} + \dots = \binom{m}{1} + \binom{m}{3} + \dots$$

So, running q through all divisors of $P(n)$, the conclusion follows. $\qquad\square$

While discussing inversion, we should also mention Dirichlet inverse.

Dirichlet Inverse. Let f be an arithmetic function such that $f(1) \neq 0$. Then the ***Dirichlet inverse*** of f is a function g such that $f * g = I$ where I is the ***identity function***

$$I(n) = \left\lfloor \frac{1}{n} \right\rfloor$$

This inverse g can be expressed recursively.

$$g(1) = \frac{1}{f(1)}$$

$$g(n) = -\frac{1}{f(1)} \sum_{\substack{d \mid n \\ d < n}} f\left(\frac{n}{d}\right) g(d)$$

Haukkanen[38] proves the following closed formula to find the Dirichlet inverse of an arithmetic function f which we do not prove here.

Theorem 14. *Let f be an arithmetic function such that $f(1) = 1$. Then the Dirichlet inverse of f is*

$$f^{-1}(n) = \sum_{k=1}^{\Omega(n)} (-1)^k \sum_{\substack{d_1 \cdots d_k = n \\ d_1, \dots, d_k > 1}} f(d_1) \cdots f(d_k)$$

[38] Pentti Haukkanen

2000   "Expressions for the Dirichlet Inverse of an Arithmetical Function", *Notes on Number Theory and Discrete Mathematics, ISSN 1310-5132 Volume 6, 2000, Number 4, Pages 118—124*, 6, 4, pp. 118-124, DOI: https://nntdm.net/volume-06-2000/number-4/118-124/, Theorem 2.2.

We leave the following as exercise.

1. If f is a multiplicative arithmetic function, then the Dirichlet inverse f^{-1} is also multiplicative.

2. If f and $f * g$ are multiplicative functions, then g is also multiplicative.

3. $\sum_{d|n} \mu(d) = I(n)$.

§3. General Convolution and Dirichlet Hyperbola Method

In this chapter, we will first discuss Dirichlet convolution and a generalization. Then we will discuss a variation and another generalization both of which are very useful in many cases.

We proved before that

$$\sum_{n \leq x} \tau(n) = 2 \sum_{n \leq \sqrt{x}} \left[\frac{x}{n}\right] - \left[\sqrt{x}\right]^2$$

In a similar manner, we can also prove the following.

$$\sum_{n \leq x} \sigma(n) = \frac{1}{2}\left(\sum_{n \leq \sqrt{x}} \left[\frac{x}{n}\right] + \sum_{n \leq \sqrt{x}} (2n + 1)\left[\frac{x}{n}\right] - \left[\sqrt{x}\right]^2 - \left[\sqrt{x}\right]^3 \right)$$

Note that, in both cases, we are able to express the partial sum of a multiplicative function up to x in terms of a combination of some partial sums of some other functions up to $\sqrt{x}$. The generalization of this method is known as the ***Dirichlet hyperbola method.***

Theorem 15 (Dirichlet Hyperbola Method). *Let f and g be arithmetic functions. If h is the Dirichlet convolution of f and g, then*

$$\sum_{n \leq x} h(n) = \sum_{n \leq a} f(n)G\left(\frac{x}{n}\right) + \sum_{n \leq b} g(n)F\left(\frac{x}{n}\right) - F(a)G(b)$$

where F and G are the partial sums of f and g respectively.

$$F(x) = \sum_{n \leq x} f(n)$$
$$G(x) = \sum_{n \leq x} g(n)$$

Specially when $a = b,$

$$\sum_{de \leq x} f(d)g(e) = \sum_{n \leq \sqrt{x}} \left(f(n)G\left(\frac{x}{n}\right) + g(n)F\left(\frac{x}{n}\right) \right) - F(\sqrt{x})G(\sqrt{x})$$

Next, we will discuss generalizations of Dirichlet convolution. Let f and g be arithmetic functions such that $g(x) = 0$ if $0 < x < 1$. Then the **general convolution** of f and g is

$$f \circ g(x) = \sum_{n \leq x} f(n)g\left(\frac{x}{n}\right)$$

We can easily prove the following.

Theorem 16 (General convolution theorem). *Let f, g and h be arithmetic functions. Then*

$$(f * g) \circ h = f \circ (g \circ h)$$

From this, we can also get the general Möbius inversion formula.

Theorem 17. *Let f, g be an arithmetic functions and f^{-1} be the Dirichlet inverse of f. If*

$$G(x) = \sum_{n \leq x} f(n)g\left(\frac{x}{n}\right)$$

then

$$g(x) = \sum_{n \leq x} f^{-1}(n)G\left(\frac{x}{n}\right)$$

§4. A Variation of Generalized Convolution

We will now consider a slight variation of generalized convolution.

$$(f \diamond g)(x) = \sum_{n \leq x} f(n)g\left(\left[\frac{x}{n}\right]\right)$$

Theorem 18. *Let f and g be arithmetic functions. Then*

$$(f \diamond g)(x) = \sum_{n \leq \sqrt{x}} g(n)\left(F\left(\left[\frac{x}{n}\right]\right) - F\left(\left[\frac{x}{n+1}\right]\right) \right) + \sum_{n \leq x/([\sqrt{x}+1])} f(n)g\left(\left[\frac{x}{n}\right]\right)$$

We can also write it as

$$(f \diamond g)(x) = \sum_{n \le \sqrt{x}} g(n)\left(F\left(\left[\frac{x}{n}\right]\right) - F\left(\left[\frac{x}{n+1}\right]\right)\right) + f(n)g\left(\left[\frac{x}{n}\right]\right) - \mathcal{B}(x)$$

where

$$\mathcal{B}(x) = \begin{cases} f([\sqrt{x}])g\left(\left[\frac{x}{[\sqrt{x}]}\right]\right) & \text{if } x < [\sqrt{x}]([\sqrt{x}]+1) \\ 0 & \text{otherwise} \end{cases}$$

Proof. Consider the integers

$$\left[\frac{x}{1}\right], \dots, \left[\frac{x}{[x]}\right]$$

All $n \le x$ does not appear in this list, only $n \le \sqrt{x}$ and numbers of the form $[x/n]$ for $n \le \sqrt{x}$ appear on this list. In fact, at most $2[\sqrt{n}]$ distinct values appear in this list. $\square$

For the rest of this section, consider $\diamond$ for arbitrary f and g,

$$F(x) = \sum_{n \le x} f(n)$$
$$= O(x^\xi)$$
$$g(x) = [x]^k$$

for a constant ξ and a fixed positive integer. Then we have

$$(f \diamond g)(x) = \sum_{n \le x} f(n)\left[\frac{x}{n}\right]^k$$

$$= \sum_{n \le \sqrt{x}} n^k\left(F\left(\left[\frac{x}{n}\right]\right) - F\left(\left[\frac{x}{n+1}\right]\right)\right) + f(n)\left[\frac{x}{n}\right]^k - \mathcal{B}(x)$$

$$= \sum_{n \le \sqrt{x}} n^k O\left(\left[\frac{x}{n}\right]^\xi\right) + f(n)\left[\frac{x}{n}\right]^k - \mathcal{B}(x)$$

$$= x^\xi O\left(\sum_{n \le \sqrt{x}} \frac{n^k}{n^\xi}\right) + \sum_{n \le \sqrt{x}} f(n)\left(\left(\frac{x}{n}\right) + O(1)\right)^k - \mathcal{B}(x)$$

$$= x^\xi O\left(\sum_{n \le \sqrt{x}} \frac{n^k}{n^\xi}\right) + \sum_{n \le \sqrt{x}} f(n)\left(\left(\frac{x}{n}\right)^k + O\left(\left(\frac{x}{n}\right)^{k-1}\right)\right) - \mathcal{B}(x)$$

$$= x^\xi O\left(\sum_{n \le \sqrt{x}} \frac{n^k}{n^\xi}\right) + x^k \sum_{n \le \sqrt{x}} \frac{f(n)}{n^k} + O\left(x^{k-1} \sum_{n \le \sqrt{x}} \frac{f(n)}{n^{k-1}}\right) - \mathcal{B}(x)$$

Now we need to focus on the following two sums.

$$\mathfrak{M}_s(x) = \sum_{n \leq x} \frac{n^s}{n^\xi}$$

$$\mathfrak{G}_s(x) = \sum_{n \leq x} \frac{f(n)}{n^s}$$

where $s \geq 1$. Then

(‡9) $$(f \diamond g)(x) = x^\xi O(\mathfrak{M}_k(\sqrt{x})) + x^k \mathfrak{G}_k(\sqrt{x}) + O(x^{k-1} \mathfrak{G}_{k-1}(\sqrt{x})) - \mathfrak{B}(x)$$

We can use Theorem 7 for computing $\mathfrak{M}$. If $s + 1 < \xi$,

(‡10)
$$\mathfrak{M}_s(x) = \sum_{n \leq x} \frac{1}{n^{\xi-s}}$$
$$= \frac{x^{1+s-\xi}}{1+s-\xi} + \zeta(\xi - s) + O(x^{s-\xi})$$

If $s + 1 = \xi$, from Divergence of Harmonic Sum,

$$\mathfrak{M}_s(x) = \log x + C + O\left(\frac{1}{x}\right)$$

The case $\xi - 1 < s < \xi$ is possible if and only if ξ is not an integer and $s = [\xi]$ which can be taken care of in the same manner as (‡10). We can now assume $s \geq \xi$. In this case, $s - \xi \geq 0$ and from Theorem 8

$$\mathfrak{M}_s(x) = \sum_{n \leq x} n^{s-\xi}$$
$$= \frac{x^{s-\xi+1}}{s-\xi+1} + O(x^{s-\xi})$$

For handling $\mathfrak{G}$, we will use Abel partial summation formula.

$$\mathfrak{G}_s(x) = \frac{F(x)}{x^s} + s\int_1^x F(t)t^{-s-1}dt$$
$$= O(x^{\xi-s}) + sO\left(\int_1^x t^{\xi-s-1}dt\right)$$

Thus, we have

$$\mathfrak{M}_s(x), \mathfrak{G}_s(x) = \begin{cases} \log x + C + O\left(\dfrac{1}{x}\right), O(x) & \text{if } s+1 = \xi \\[2mm] \dfrac{x^{1+s-\xi}}{1+s-\xi} + \zeta(\xi - s) + O(x^{s-\xi}), O(x^{\xi-s}) & \text{if } s < \xi \text{ and } s+1 \neq \xi \\[2mm] x + O(1), O(\log x) & \text{if } s = \xi \\[2mm] \dfrac{x^{s-\xi+1}}{s-\xi+1} + O(x^{s-\xi}), O(x^{\xi-s}) & \text{if } s > \xi \end{cases}$$

Plugging these back in ($\ddagger$9), we get the following result.

Theorem 19. *Let f and g be arithmetic functions such that $F(x) = \sum_{n \leq x} f(n) = O(x^\xi), g(x) = [x]^k$. Then*

$$(f \diamond g)(x) = \begin{cases} O(x^{k+1} \log x) & \textbf{\textit{if }} k+1 = \xi \\[1mm] O\left(x^{\frac{k+\xi+1}{2}} + x^\xi \zeta(\xi - k)\right) & \textbf{\textit{if }} k < \xi, k+1 \neq \xi \\[1mm] O(x^{k+\frac{1}{2}}) & \textbf{\textit{if }} k = \xi \\[1mm] O(x^{\frac{k+\xi+1}{2}}) & \textbf{\textit{if }} k > \xi \end{cases}$$

§5. Generalization of General Convolution

Let $\mathbf{x} = (x_1, \ldots, x_k)$ and $\mathbf{a} = (a_1, \ldots, a_n)$ be vectors of positive real numbers. $\{\sqrt[\mathbf{a}]{\mathbf{x}}\}$ denotes the largest positive integer n for which $n^{a_i} \leq x_i$ for some $1 \leq i \leq k$. That is,

$$\max\{\sqrt[\mathbf{a}]{\mathbf{x}}\} = \max\{[\sqrt[a_1]{x_1}], \ldots, [\sqrt[a_k]{x_k}]\}$$

For a positive integer n, let $n^{\mathbf{a}} \leq \mathbf{x}$ denote that $n \leq \max\{\sqrt[\mathbf{a}]{\mathbf{x}}\}$.

Let f be a real or complex valued function defined in k variables. For a vector of positive real numbers $\mathbf{a}$, let $\mathbf{x}/\mathbf{a}$ denote the vector $(x_1/a_1, \ldots, x_k/a_1)$, $[\mathbf{x}/\mathbf{a}]$ denote the vector $([x_1/a_1], \ldots, [x_k/a_k])$ and

$$f(\mathbf{x}) = f(x_1, \ldots, x_k)$$
$$f\left(\frac{\mathbf{x}}{n^{\mathbf{a}}}\right) = f\left(\frac{x_1}{n^{a_1}}, \ldots, \frac{x_k}{n^{a_k}}\right)$$
$$f\left(\left[\frac{\mathbf{x}}{n^{\mathbf{a}}}\right]\right) = f\left(\left[\frac{x_1}{n^{a_1}}\right], \ldots, \left[\frac{x_k}{n^{a_k}}\right]\right)$$

Generalized Convolution.   Let the generalized convolution of an arithmetic function α and a function f defined for k real numbers and a positive integer a be

(‡11)
$$(\alpha \bullet f)(\mathbf{x}, \mathbf{a}) = \sum_{n^{\mathbf{a}} \leq \mathbf{x}} \alpha(n) f\left(\frac{\mathbf{x}}{n^{\mathbf{a}}}\right)$$

We have the next theorem about the associativity of $\bullet$ convolution.

Theorem 20 (Associativity of Generalized Convolution). *Let $\mathbf{x}$ be a vector of k positive real numbers, α, β be arithmetic functions, a be a fixed positive integer and $f(x_1, \dots, x_k)$ be a real or complex valued multivariate function. Then*

$$(\alpha \bullet (\beta \bullet f))(\mathbf{x}, \mathbf{a}) = ((\alpha * \beta) \bullet f)(\mathbf{x}, \mathbf{a})$$

where $f * g$ is the usual Dirichlet convolution of arithmetic functions f and g.

Proof. From the definition,

$$(\beta \bullet f)(\mathbf{x}, \mathbf{a}) = \sum_{m^{\mathbf{a}} \leq \mathbf{x}} \beta(m)\left(\frac{x_1}{m^{a_1}}, \dots, \frac{x_k}{m^{a_k}}\right)$$

$$(\alpha \bullet (\beta \bullet f))(\mathbf{x}, \mathbf{a}) = \sum_{n^{\mathbf{a}} \leq \mathbf{x}} \alpha(n)\left((\beta \bullet f)\left(\frac{\mathbf{x}}{n^{\mathbf{a}}}, \mathbf{a}\right)\right)$$

$$= \sum_{n^{\mathbf{a}} \leq \mathbf{x}} \alpha(n)\left((\beta \bullet f)\left(\frac{x_1}{n^{a_1}}, \dots, \frac{x_k}{n^{a_k}}\right)\right)$$

$$= \sum_{n^{\mathbf{a}} \leq \mathbf{x}} \alpha(n) \sum_{m^{a} \leq \mathbf{x}/n^{\mathbf{a}}} \beta(m) f\left(\frac{x_1}{m^{a}n^{a}}, \dots, \frac{x_k}{m^{a}n^{a}}\right)$$

We can collect the m and n together and write

$$(\alpha \bullet (\beta \bullet f))(\mathbf{x}, \mathbf{a}) = \sum_{(mn)^{\mathbf{a}} \leq \mathbf{x}} \alpha(n)\beta(m) f\left(\frac{x_1}{m^{a}n^{a}}, \dots, \frac{x_k}{m^{a_k}n^{a_k}}\right)$$

$$= \sum_{n^{\mathbf{a}} \leq \mathbf{x}} \left(\sum_{d|n} \alpha(d)\beta\left(\frac{n}{d}\right)\right) f\left(\frac{x_1}{n^{a_1}}, \dots, \frac{x_k}{n^{a_k}}\right)$$

$$= \sum_{n^{\mathbf{a}} \leq \mathbf{x}} (\alpha * \beta) f\left(\frac{\mathbf{x}}{n^{\mathbf{a}}}\right)$$

$$= (\alpha * \beta) \bullet f(\mathbf{x}, \mathbf{a})$$

□

Theorem 21 (Inversion of Generalized Convolution). *Let α be an arithmetic function and f be a real or complex valued multivariate function. If*

$$g(\mathbf{x}, \mathbf{a}) = \sum_{n^{\mathbf{a}} \leq \mathbf{x}} \alpha(n) f(\mathbf{x}, \mathbf{a})$$

and α^{-1} is the Dirichlet inverse of α, then

$$f(\mathbf{x}, \mathbf{a}) = \sum_{n^{\mathbf{a}} \leq \mathbf{x}} \alpha^{-1}(n) g\left(\frac{\mathbf{x}}{n^{\mathbf{a}}}\right)$$

Proof. First, we see that

$$\begin{aligned}
(I \bullet f)(\mathbf{x}, \mathbf{a}) &= \sum_{n^{\mathbf{a}} \leq \mathbf{x}} I(n) f(\mathbf{x}, \mathbf{a}) \\
&= I(1) f(\mathbf{x}, \mathbf{a}) + \sum_{\substack{n^{\mathbf{a}} \leq \mathbf{x} \\ n > 1}} I(n) f(\mathbf{x}, \mathbf{a}) \\
&= f(\mathbf{x}, \mathbf{a})
\end{aligned}$$

Since $g = \alpha \bullet f$, we will use Associativity of Generalized Convolution on α^{-1} and g. We have

$$(\alpha^{-1} \bullet (\alpha \bullet f))(\mathbf{x}, \mathbf{a}) = ((\alpha^{-1} * \alpha) \bullet f)(\mathbf{x}, \mathbf{a})$$

From the definition of Dirichlet inverse, $\alpha^{-1} * \alpha = I$. So, we have

$$\begin{aligned}
(\alpha^{-1} \bullet g)(\mathbf{x}, \mathbf{a}) &= (\alpha^{-1} \bullet (\alpha \bullet f))(\mathbf{x}, \mathbf{a}) \\
&= ((\alpha^{-1} * \alpha) \bullet f)(\mathbf{x}, \mathbf{a}) \\
&= (I \bullet f)(\mathbf{x}, \mathbf{a}) \\
&= f(\mathbf{x}, \mathbf{a})
\end{aligned}$$

Thus, we have the theorem. $\qquad\square$

If we set $\mathbf{a} = (1), k = 1$ and $\mathbf{x} = (x)$ for a real number x in Associativity of Generalized Convolution and Inversion of Generalized Convolution, we have the usual General convolution theorem and Inversion of Generalized Convolution.

Theorem 22. *Let f and g be arithmetic functions and $h = f * g$. If*

$$F(\mathbf{x}, \mathbf{a}) = \sum_{n^{\mathbf{a}} \leq \mathbf{x}} f(n)$$

$$G(\mathbf{x}, \mathbf{a}) = \sum_{n^{\mathbf{a}} \leq \mathbf{x}} g(n)$$

$$H(\mathbf{x}, \mathbf{a}) = \sum_{n^{\mathbf{a}} \leq \mathbf{x}} h(n)$$

then we have

$$\begin{aligned}
H(\mathbf{x}, \mathbf{a}) &= \sum_{n^{\mathbf{a}} \leq \mathbf{x}} f(n) \left(\frac{\mathbf{x}}{n^{\mathbf{a}}}, \mathbf{a}\right) \\
&= \sum_{n^{\mathbf{a}} \leq \mathbf{x}} g(n) F\left(\frac{\mathbf{x}}{n^{\mathbf{a}}}, \mathbf{a}\right)
\end{aligned}$$

Proof. We can write H as follows.

$$
\begin{aligned}
H(\mathbf{x}, \mathbf{a}) &= \sum_{n^{\mathbf{a}} \leq (\mathbf{x})} h(n) \\
&= \sum_{(de)^{\mathbf{a}} \leq \mathbf{x}} f(d) g(e) \\
&= \sum_{d^{\mathbf{a}} \leq \mathbf{x}} f(d) \sum_{e^{\mathbf{a}} \leq \mathbf{x}/d^{\mathbf{a}}} g(e) \\
&= \sum_{d^{\mathbf{a}} \leq \mathbf{x}} f(d) G\left(\frac{\mathbf{x}}{d^{\mathbf{a}}}, \mathbf{a}\right)
\end{aligned}
$$

We can prove the other part similarly by fixing e and letting d run through for f instead of g. $\qquad\square$

As a corollary, we have the following theorem.

Theorem 23. *Let f be an arithmetic function. If*

$$
F(\mathbf{x}, \mathbf{a}) = \sum_{n^{\mathbf{a}} \leq \mathbf{x}} f(n)
$$

then we have

$$
\begin{aligned}
\sum_{n \leq \sqrt[\mathbf{a}]{\mathbf{x}}} \sum_{d \mid n} f(d) &= \sum_{n \leq \sqrt[\mathbf{a}]{\mathbf{x}}} \left[\frac{\sqrt[\mathbf{a}]{\mathbf{x}}}{n}\right] f(n) \\
&= \sum_{n^{\mathbf{a}} \leq \mathbf{x}} F\left(\frac{\sqrt[\mathbf{a}]{\mathbf{x}}}{n}\right)
\end{aligned}
$$

We will now see some applications of generalized convolution $\bullet$. Let s be a fixed positive integer and for the rest of the section, f be defined as

$$
f(\mathbf{x}) = \prod_{i=1}^{k} [x_i]
$$

Define the function u as

$$
u(n) = n^s
$$

From (‡7),

$$
\sum_{d \mid n} J(d) = n^s
$$

Using Möbius Inversion, we also get that

(‡12) $$\qquad\qquad\qquad \mu * u = J$$

Let $F(\mathbf{x})$ be the number of vectors of positive integers $(a_1, \dots, a_k)$ such that $1 \le a_i \le x_i$ and $\gcd(a_1, \dots, a_k) = 1$. Then we have

$$F(\mathbf{x}) = (\mu \bullet f)(\mathbf{x}, \mathbf{1})$$

The total number of vectors such that $1 \le a_i \le x_i$ is $x_1 \cdots x_k$. Consider an arbitrary vector $(a_1, \dots, a_k)$. If $g = \gcd(a_1, \dots, a_k) > 1$, then every a_i has to be divisible by g. Then the number of such vectors is

$$t(g) = \left(\frac{\mathbf{a}}{g}\right)$$
$$= \left\lceil \frac{a_1}{g} \right\rceil \cdots \left\lceil \frac{a_k}{g} \right\rceil$$

We can see that the $t(p)$ vectors which has all elements divisible by p also has all vectors which are divisible by a multiple of p. So, if g is composite, and has r prime factors, every vector of the $t(g)$ vectors is also divisible by any of those r prime factors. Using a simple principle of inclusion and exclusion, we see that the number of vectors divisible by g has the sign $\mu(g)$. So, the total number of vectors where they have a common factor other than 1 is

$$\sum_{2 \le g \le \min(\mathbf{x})} \mu(g) \left\lceil \frac{x_1}{g} \right\rceil \cdots \left\lceil \frac{x_k}{g} \right\rceil$$

Then the number of vectors where $\gcd(a_1, \dots, a_k) = 1$ is

$$x_1 \cdots x_k + \left(\sum_{2 \le g \le \min(\mathbf{x})} \mu(g) \left\lceil \frac{x_1}{g} \right\rceil \cdots \left\lceil \frac{x_k}{g} \right\rceil \right) = \sum_{n \le \min(\mathbf{x})} \mu(n) f(\mathbf{x}, \mathbf{1})$$

Thus, we have the result. As a consequence of this result, we can prove the next result using the fact that the number of non-decreasing sequences $(a_1, \dots, a_k)$ such that $1 \le a_i \le a_{i+1} \le n$ is $\binom{n+k-1}{k}$.

Let $B(n, k)$ be the number of vectors of non-decreasing sequences $(a_1, \dots, a_k)$ such that $1 \le a_1 \le \dots \le a_k \le n$ and $\gcd(a_1, \dots, a_k) = 1$. If for a positive integer m, $\mathbf{m} = \underbrace{(m, \dots, m)}_{k \text{ times}}$ and

$$f(\mathbf{m}) = \binom{m+k-1}{k}$$

then we have

$$B(n, k) = (\mu \bullet f)(\mathbf{n}, \mathbf{1})$$

Next, let S be the sum

$$S(\mathbf{x}) = \sum_{1 \leq a_i \leq x_i} g(\mathbf{a})^s$$

where $g(\mathbf{a}) = \gcd(a_1, \ldots, a_k)$ for the vector of positive integers $\mathbf{a} = (a_1, \ldots, a_k)$. Then we have

(‡13)
$$S(\mathbf{x}) = \sum_{n \leq \mathbf{x}} J_s(n) \prod_{i=1}^{k} \left\lceil \frac{x_i}{n} \right\rceil$$

(‡14)
$$= \sum_{n \leq \mathbf{x}} \mu(n) \left(\sum_{i \leq \mathbf{x}/n} i^s \prod_{j=1}^{k} \left\lceil \frac{x_j}{ni} \right\rceil \right)$$

(‡13) follows from Associativity of Generalized Convolution and (‡12).

$$(J_s \bullet f)(\mathbf{x}, 1) = (\mu \bullet (u \bullet f))(\mathbf{x}, 1)$$

So, we will only prove (‡14). Consider the vector $(a_1, \ldots, a_k)$ and $g = \gcd(a_1, \ldots, a_k)$. Letting $a_i = gb_i$, we have that $\gcd(b_1, \ldots, b_k) = 1$. The number of such vectors is $(\mu \bullet f)(\mathbf{x}, 1)$. Each of these vectors contribute g^s to the sum, so for a particular g, the contribution of g in the sum is

$$g^s(\mu \bullet f)(\mathbf{x}, 1)$$

Then by the principle of inclusion and exclusion, we have that

$$S(\mathbf{x}) = \sum_{n \leq \mathbf{x}} n^s(\mu \bullet f)(\mathbf{x}, 1)$$
$$= (u \bullet (\mu \bullet f))(\mathbf{x}, 1)$$

Remark. We could prove this result without using $\bullet$ convolution as well. For example, in the case $s = 1$, if $d \mid g$ and $d < g$, then g has already appeared in the vectors of d. Thus, we cannot consider any d that shares a common factor with g. $n \leq g$ will contribute a new sum to the vectors only if $\gcd(n, g) = 1$. So, the total sum of $g(\mathbf{a})$ with $\gcd(a_1, \ldots, a_k) = g$ is $\varphi(g)$. Generalizing this for arbitrary s, we can easily see that the contributed sum for g is

$$J_s(g) \sum_{n \leq \min(\mathbf{x})/g} f(\mathbf{x}/n)$$

CHAPTER II.

BERTRAND TO TSCHEBISCHEFF

We said before that *almost* all natural numbers are composite. A major objective of this book is to discuss how often the primes occur. The same question has bugged mathematicians for a centuries. It was Gauss who first observed that the change in the distribution of primes in every interval $[x, x + 1000]$ was around $1/\log x$. Thus, the rough estimate

$$\pi(x) = \int\limits_{2}^{x} \frac{1}{\log t} dt$$

was made which is now known as *logarithmic integral*. Gauss conjectured (see Landau[1]) around 1792 or 1793 that

$$\lim_{x \to \infty} \frac{\pi(x)}{\frac{x}{\log x}} = 1$$

Tschebischeff[2] is the first one to make any substantial progress on the matter. In fact, Tschebischeff was close to proving the prime number theorem himself. Tschebischeff[3] first proved that if $\pi(x)$ was of order $\frac{x}{\log^N x}$ as $x \to \infty$, then $N = 1$. Consequently, he also proved that if the limit

$$\lim_{\xi \to \infty} \frac{\xi}{\log \xi}$$

exists, then it is 1. The only problem was to actually prove that this limit indeed exists.

---

[1]Edmund Landau

1911 "Handbuch der Lehre von der Verteilung der Primzahlen", *Monatshefte für Mathematik und Physik*, 22, 1 (Dec. 1911), DOI: 10.1007/bf01742852, Page 37.

[2]Pafnutï Lvovitch Tschebischeff

1852b "Sur la totalité des nombres premiers inférieurs à une limite donnée", *Journal de Mathématiques Pures et Appliquées*, 17, 1, pp. 341-365.

[3]Pafnutï Lvovitch Tschebischeff

1852a "Mémoire sur les nombres premiers", *Journal de Mathématiques Pures et Appliquées*, 17, 1, pp. 366-390.

Theorem 24. *Let ξ be a positive real number. Then there are constants a and A such that*

$$a\frac{\xi}{\log \xi} \leqq \pi(\xi) \leqq A\frac{\xi}{\log \xi}$$

Actually, Tschebischeff gave a specific statement that

$$a < \frac{\pi(\xi)}{\frac{\xi}{\log \xi}} < \frac{6a}{5}$$

holds for large enough x where

$$a = \frac{\log 2}{2} + \frac{\log 3}{3} + \frac{\log 5}{5} - \frac{\log 30}{30}$$

The following proof is inspired by Landau;[4] which is actually a translation of the first section of the first volume of Landau.[5]

Proof. For any $x \geqq 0$,

$$[x] - 2\left[\frac{x}{2}\right] \leqq 1$$

since

$$[x] - 2\left[\frac{x}{2}\right] < x - 2\left(\frac{x}{2} - 1\right)$$
$$= 2$$

Let $n \geqq 2$. For every $p \leqq 2n$, let r denote the largest positive integer such that $p^r \leqq 2n$ (which is $[\log 2n/\log p]$). We will first show that

(‡1)
$$\prod_{n < p \leqq 2n} p \mid \binom{2n}{n}$$

(‡2)
$$\binom{2n}{n} \mid \prod_{p \leqq 2n} p^r$$

where the products run through the primes only. For a prime p,

$$\nu_p\left(\binom{2n}{n}\right) = \nu_p((2n)!) - 2\nu_p(n!)$$

---

[4]Landau, *Elementary number theory* cit., Theorem 112.
[5]Landau, *Vorlesungen über Zahlentheorie* cit.

Since $p > n$, $\nu_p(n!) = 0$ and $\nu_p\left(\binom{2n}{n}\right) = \nu_p((2n)!) \geq \nu_p\left(\prod_{n < p \leq 2n} p\right)$. On the other hand, for $p \leq 2n$,

$$\nu_p\left(\binom{2n}{n}\right) = \nu_p((2n)!) - 2\nu_p(n!)$$
$$= \sum_{i \geq 1}\left[\frac{2n}{p^i}\right] - 2\left[\frac{n}{p^i}\right]$$

Since $[2x] - 2[x] \leq 1$ for any $x \geq 0$ and $\nu_{p^i}((2n)!) = 0$ for $i > r$, we have

$$\sum_{i \geq 1}\left[\frac{2n}{p^i}\right] - 2\left[\frac{n}{p^i}\right] \leq \sum_{1 \leq i \leq r} 1$$

Thus, for every prime p,

$$p^{\nu_p\left(\binom{2n}{n}\right)} \mid p^r$$

and evidently

$$\binom{2n}{n} \mid \prod_{p \leq 2n} p^r$$

Now, there are $\pi(2n) - \pi(n)$ primes in the interval $(n, 2n)$ each of which are greater than n. So,

$$n^{\pi(2n) - \pi(n)} \leq \prod_{n < p \leq 2n} p$$
$$\leq \binom{2n}{n}$$
$$\leq \prod_{n < p \leq 2n} p^{[\log_p 2n]}$$
$$\leq (2n)^{\pi(2n)}$$

Taking logarithm,

$$(\pi(2n) - \pi(n)) \log n \leq \pi(2n) \log 2n$$

Next, we have that

$$\binom{2n}{n} \leq \sum_{i=0}^{2n} \binom{2n}{i}$$
$$= 2^{2n}$$
$$\binom{2n}{n} = \frac{(n+1)\cdots(2n)}{1\cdots n}$$
$$\geq \prod_{i=1}^{n} \frac{n+i}{i}$$
$$= 2^{n}$$

giving

$$(\pi(2n) - \pi(n))\log n \leq \log\binom{2n}{n} \leq \pi(2n)\log 2n$$

Thus, we have

$$(\pi(2n) - \pi(n))\log n \leq \log 2^{2n}$$

and also

$$\pi(2n)\log 2n \geq \log 2^{n}$$
$$\pi(2n) \geq c_1 \frac{2n}{\log 2n}$$

for a positive constant c_1. For $\xi \geq 2$, setting $\eta = [\xi/2]$ and $\xi = 2\eta + r$ with $0 \leq r < 2$, we have $\xi \leq 3\eta$ and $\log 2\eta \leq \log \xi$.

$$\pi(\xi) \geq \pi(2\eta)$$
$$\geq c\frac{2\eta}{\log 2\eta}$$
$$= \frac{2c}{3}\frac{3\eta}{\log 2\eta}$$
$$\geq a\frac{\xi}{\log \xi}$$

for a positive constant a. On the other hand, since $x < 2 + 2[\frac{x}{2}]$,

$$\pi(x) - \pi\left(\frac{x}{2}\right) = \pi(x) - \pi\left(\left[\frac{x}{2}\right]\right)$$
$$\leqq 2 + \pi\left(2\left[\frac{x}{2}\right]\right) - \pi\left(\left[\frac{x}{2}\right]\right)$$
$$\leqq 2 + 2\log 2 \frac{\left[\frac{x}{2}\right]}{\log\left[\frac{x}{2}\right]}$$
$$< d\frac{x}{\log x}$$

for a constant d. Using $\pi\left(\frac{x}{2}\right) \leqq \frac{x}{2}$,

$$\log x\pi(x) - \log \frac{x}{2}\pi\left(\frac{x}{2}\right) = \log x\left(\pi(x) - \pi\left(\frac{x}{2}\right)\right) + \log 2 \cdot \pi\left(\frac{x}{2}\right)$$
$$< \log x \cdot d_1\frac{x}{\log x} + \frac{x}{2}$$
$$< d_2 x$$

If $2^{m+1} \leqq 2^{v+1}\xi < 2^{v+2}$ (that is, $v + 1 = [\log_2 \xi]$),

$$\log \frac{\xi}{2^m}\pi\left(\frac{\xi}{2^m}\right) - \log \frac{\xi}{2^{m+1}}\pi\left(\frac{\xi}{2^{m+1}}\right) < d_2\frac{\xi}{2^m}$$

Summing over this for $0 \leqq m \leqq v$,

$$\log \xi\pi(\xi) = \sum_{m=0}^{v}\left(\log \frac{\xi}{2^m}\pi\left(\frac{\xi}{2^m}\right) - \log \frac{\xi}{2^{m+1}}\pi\left(\frac{x}{2^{m+1}}\right)\right)$$
$$< d_2\xi \sum_{m=0}^{v}\frac{1}{2^m}$$
$$= A\xi$$

$\square$

A corollary is the following.

Theorem 25. *For a positive integer $n > 1$,*

$$bn\log n < p_n < Bn\log n$$

The proof is left to the reader as an exercise. Using this, we can prove Euler's divergence of sum of reciprocals of prime theorem. Since

$$\frac{1}{p_n} > \frac{1}{Bn\log n}$$

we only need to show that

$$\sum_{n \geqq 1} \frac{1}{n \log n}$$

diverges. This is also left to the reader as an exercise. Euler[6] proved this first although his method was a bit questionable.

Theorem 26 (Divergence of sum of reciprocals of primes). *The sum*

$$\sum_p \frac{1}{p} = \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \dots$$

where the sum is taken over all primes diverges.

This proof is again inspired by Landau.[7]

Proof. From Divergence of Harmonic Sum, we already know that

$$\sum_{n \geqq 1} \frac{1}{n} = \prod_p \frac{1}{1 - \frac{1}{p}}$$

and that this sum diverges. Now,

$$\log \left(\sum_{n \geqq 1} \frac{1}{n} \right) = \sum_p \left(- \log \left(1 - \frac{1}{p} \right) \right)$$

Setting $\eta := 1/p$ and using

$$\log (1 - x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots$$

[6]Leonhard Euler

1737 "Variae Observationes circa series Infinitas", *Commentarii Academiae Scientiarum Petropolitanae*, 9, pp. 160-188, DOI: 10.1090/spec/098/41.
    [7]Landau, *Elementary number theory* cit., Theorem 114.

we have

$$\log\left(\sum_{n\geqq 1}\frac{1}{n}\right) = \sum_{p}\left(\eta + \frac{\eta^2}{2} + \frac{\eta^3}{3} + \ldots\right)$$

$$< \sum_{p}\left(\eta + \eta^2 + \ldots\right)$$

$$= \sum_{p}\frac{\eta}{1-\eta}$$

$$< 2\sum_{p}\eta$$

$$= 2\sum_{p}\frac{1}{p}$$

If $\sum_{p}\frac{1}{p}$ does not diverge, then

$$\log\left(\sum_{n\geqq 1}\frac{1}{n}\right) < C$$

for a constant C. Thus,

$$\sum_{n\geqq 1}\frac{1}{n} < e^C$$

and does not converge either. This is impossible. So, the original sum must diverge. $\square$

Mertens[8] actually proved that

$$\sum_{p\leqq x}\frac{1}{p} = \log\log x$$

This is the first formal proof since Euler's method wasn't exactly clean. Euler[9] uses

$$\log\left(\frac{1}{1-x}\right) = \sum_{n\geqq 1}\frac{x^n}{n}$$

and sets $x := 1$ to conclude

$$\sum_{n\geqq 1}\frac{1}{n} = \infty$$

---

[8] Franz Mertens

1874 "Ein Beitrag Zur analytischen zahlentheorie", *Journal für die reine und angewandte Mathematik Band 78*, 78, pp. 46-62, DOI: 10.1515/9783112389843-002.

[9] Leonhard Euler

1748 *Introductio in Analysin Infinitorum*, vol. 1, Page 228.

This is indeed true, however, Euler's statement is vague. Today it is known that

$$\sum_{p \leqq x} \frac{1}{p} \geqq \log\log(x+1) - \frac{\pi^2}{6}$$

CHAPTER III.

TWO ELEMENTARY PROOFS OF LEGENDRE-DIRICHLET PRIME NUMBER THEOREM

§1. Dirichlet Characters

We will carry some concepts from elementary number theory. Legendre[1] defined the *Legendre symbol* which plays a pivotal role in the theory of quadratic residue. Jacobi[2] generalized Legendre symbol which is now known as *Jacobi symbol.* Kronecker[3] generalized this to *Kronecker symbol* which is unfortunately less known today. This is because

---

[1]Adrien Marie Legendre

1798 *Essai sur la théorie des nombres,* 2nd ed., Duprat, DOI: 10.1017/cbo9780511693199.020, pp. 186.

[2]Carl Gustav Jacob Jacobi

1846 "Über die Kreisteilung und ihre Anwendung auf die Zahlentheorie", *Bericht Ak. Wiss. Berlin,* 30, pp. 127-136, DOI: 10.1515/crll.1846.30.166.

[3]Leopold Kronecker

1885 "Zur Theorie der elliptischen Funktionen", *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin,* pp. 761-784, pp. 770.

Dirichlet[4] introduced (also see Dirichlet[5]) had already introduced a generalization of Kronecker symbol. For this section, let k be a fixed positive integer.

Dirichlet Character. An arithmetic function χ is called a ***character*** $(\bmod\ k)$ if

    (I) $\chi(a) = 0$ if $\gcd(a, k) > 1$.

    (II) $\chi(1) \neq 0$.

    (III) $\chi(ab) = \chi(a)\chi(b)$ for all positive integers a, b.

    (IV) $\chi(a) = \chi(b)$ if $a \equiv b \pmod{k}$.

Note that Legendre symbol $\left(\frac{a}{p}\right)$ $(\bmod\ p)$ for prime p, Jacobi symbol $\left(\frac{a}{n}\right)$ and Kronecker symbol $\left(\frac{a}{n}\right)$ are all characters $(\bmod\ p)$, $(\bmod\ n)$ and $(\bmod\ n)$ where n is square-free and $n \equiv 0, 1 \pmod{4}$ respectively. Another example of χ is $\chi(n) = 0$ for $n \equiv 0, 2 \pmod{4}$, $\chi(n) = 1$ for $n \equiv 1 \pmod{4}$ and $\chi(n) = -1$ for $n \equiv 3 \pmod{4}$. Obviously, χ is completely multiplicative; hence $\chi(1) = 1$.

Theorem 27. *If* $\gcd(n, k) = 1$, *then* $\chi(n)^k = 1$. *So* $\chi(n)$ *is actually a complex number which is an n-th root of unity and* $|\chi(n)| = 1$.

Proof. From Euler's theorem,

$$a^{\varphi(k)} \equiv 1 \pmod{k}$$

So, $\chi(n)^k = \chi(n^k) = \chi(1) = 1$. $\square$

Theorem 28.

§2. Non-vanishing of L-Series At $s = 1$

[4]Johann Peter Gustav Lejeune Dirichlet

1837   "Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält", *Abhandlungen der Königlich Preussischen Akademie der Wissenschaften*, pp. 45-81, DOI: 10.1017/cbo9781139237321.012.

[5]Johann Peter Gustav Lejeune Dirichlet

1897a   "Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält", in, *G. Lejeune Dirichlet's Werke*, ed. by Leopold Kronecker and László Fuchs, Druck Und Verlag Von Georg Reimer., vol. 1, pp. 313-342, DOI: 10.1017/cbo9781139237321.012.

§3. First Proof

§4. Proof by Selberg

CHAPTER IV.

TWO ELEMENTARY PROOFS OF THE PRIME NUMBER THEOREM

The two proofs in this chapter essentially use the same ideas. There is a lot of history until this point in mathematics. We will cover some of it in Chapter VI. For now, we will ignore the history and focus only on the proof.

A MODEST INTRODUCTION TO SIEVE THEORY

A composite positive integer n has at least one prime factor not exceeding $\sqrt{x}$. Thus, the number of primes in the interval $[\sqrt{x}, x]$ is

$$\pi(x) - \pi(\sqrt{x}) + 1 = [x] - \sum_{p \leq x}\left[\frac{x}{p}\right] + \sum_{p_1 < p_2 \leq x}\left[\frac{x}{p_1 p_2}\right] - \sum_{p_1 < p_2 < p_3 \leq x}\left[\frac{x}{p_1 p_2 p_3}\right] + \ldots$$

$$= \sum_{n \leq x} \mu(n)\left[\frac{x}{n}\right]$$

Now, $[x] = x + O(1)$, so

$$\pi(x) - \pi(\sqrt{x}) + 1 = x \sum_{\substack{n \leq x \\ \varrho(n) \leq \sqrt{x}}} \frac{\mu(n)}{n} + O\left(\sum_{\substack{n \leq x \\ \varrho(n) \leq \sqrt{x}}} \mu(n)\right)$$

$$= x \prod_{p \leq \sqrt{x}}\left(1 - \frac{1}{p}\right) + O(2^{\pi(\sqrt{x})})$$

The last line is true since there are $\pi(\sqrt{x})$ primes not exceeding $\sqrt{x}$ and $|\mu(n)| = 1$ for all square-free $n \leq x$ such that $\varrho(n) \leq \sqrt{x}$. However, this is not particularly useful so we want to improve on this. This is the starting point for sieves. Let us generalize this concept first. Consider a set of integers A. $\mathfrak{M}, \mathfrak{G}$

THE SELBERG AND ERDŐS DISPUTE

GLOSSARY

Erdős He is Erdős 1

ARTICLES

Niels Henrik Abel
   1826 "Untersuchungen über die Reihe: $1 + (m/1)x + m \cdot (m-1)/(1 \cdot 2) \cdots x^2 + m \cdot (m-1) \cdot (m-2)/(1 \cdot 2 \cdot 3) \cdots x^3 + ...$", *Journal für Math.*, 1, pp. 311-339, DOI: 10.1515/9783112347386-030.

Raymond Ayoub
   1974 "Euler and the zeta function", *The American Mathematical Monthly*, 81, 10, pp. 1067-1086, DOI: 10.2307/2319041.

Martin Beumer
   1962 "The arithmetical function $\tau_k(n)$", *The American Mathematical Monthly*, 69, 8, pp. 777-781, DOI: 10.2307/2310778.

Eugène Cahen
   1894 "Sur la fonction $\zeta(s)$ de Riemann et sur des fonctions analogues", fr, *Annales scientifiques de l'École Normale Supérieure*, 11, pp. 75-164, DOI: 10.24033/asens.401.

Johann Peter Gustav Lejeune Dirichlet
   1837 "Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält", *Abhandlungen der Königlich Preussischen Akademie der Wissenschaften*, pp. 45-81, DOI: 10.1017/cbo9781139237321.012.

Leonhard Euler
   1737 "Variae Observationes circa series Infinitas", *Commentarii Academiae Scientiarum Petropolitanae*, 9, pp. 160-188, DOI: 10.1090/spec/098/41.
   1738 "De summatione innumerabilium progressionum", *Commentarii academiae scientiarum Petropolitanae*, 5, pp. 91-105.
   1740 "De Progressionibus Harmonicis Observationes", *Commentarii academiae scientiarum Petropolitanae*, 7, pp. 150-161.

Leonhard Euler

> 1744 "Variae Observationes circa series infinitas", *Commentarii academiae scientiarum Petropolitanae*, 9, pp. 160-188.
>
> 2020a "E-20: De summatione innumerabilium Progressionum", *Spectrum*, pp. 52-64, DOI: 10.1090/spec/098/10.
>
> 2020b "E-43: De Progressionibus Harmonicis Observationes", *Spectrum*, pp. 133-141, DOI: 10.1090/spec/098/23.
>
> 2020c "E-72: Variae Observationes circa series Infinitas", *Spectrum*, pp. 249-260, DOI: 10.1090/spec/098/41.

Godfrey Harold Hardy and Srinivasa Ramanujan

> 1917 "The normal number of prime factors of a number n", *Quarterly Journal of Mathematics*, 48, pp. 76-92.

Pentti Haukkanen

> 2000 "Expressions for the Dirichlet Inverse of an Arithmetical Function", *Notes on Number Theory and Discrete Mathematics, ISSN 1310-5132 Volume 6, 2000, Number 4, Pages 118—124*, 6, 4, pp. 118-124, DOI: https://nntdm. net/volume-06-2000/number-4/118-124/.

Carl Gustav Jacob Jacobi

> 1846 "Über die Kreisteilung und ihre Anwendung auf die Zahlentheorie", *Bericht Ak. Wiss. Berlin*, 30, pp. 127-136, DOI: 10.1515/crll.1846.30.166.

Johan Ludwig William Valdemar Jensen

> 1884 "OM RÆKKERS KONVERGENS", *Tidsskrift for mathematik*, 5th ser., 2, pp. 63-72, ISSN: 09092528, 24460737, http://www.jstor.org/stable/245400 57.
>
> 1888 "Sur une généralisation d'un théorème de Cauchy", *Comptes Rendus* (Mar. 1888).

Leopold Kronecker

> 1885 "Zur Theorie der elliptischen Funktionen", *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, pp. 761-784.

Edmund Landau

> 1911 "Handbuch der Lehre von der Verteilung der Primzahlen", *Monatshefte für Mathematik und Physik*, 22, 1 (Dec. 1911), DOI: 10.1007/bf01742852.
>
> 1912a "Über die Anzahl der Gitterpunkte in geweissen Bereichen", *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 19, pp. 687-772.
>
> 1912b "Über eine idealtheoretische funktion", *Transactions of the American Mathematical Society*, 13, 1, pp. 1-21, DOI: 10.1090/s0002-9947-1912-1500901-6.

Derrick Norman Lehmer

  1900 "Asymptotic evaluation of certain Totient Sums", *American Journal of Mathematics*, 22, 4, pp. 293-335, DOI: 10.2307/2369728.

Kieren MacMillan and Jonathan Sondow

  2011 "Proofs of power sum and binomial coefficient congruences via Pascal's identity", *The American Mathematical Monthly*, 118, 6, pp. 549-551, DOI: 10.4169/amer.math.monthly.118.06.549.

Franz Mertens

  1874 "Ein Beitrag Zur analytischen zahlentheorie", *Journal für die reine und angewandte Mathematik Band 78*, 78, pp. 46-62, DOI: 10.1515/9783112389843-002.

August Ferdinand Möbius

  1832 "Über eine besondere art von Umkehrung der Reihen." *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 9, pp. 105-123, DOI: 10.1515/crll.1832.9.105.

Blaise Pascal

  1964 "Sommation des puissances numériques", *Oeuvres complètes, Jean Mesnard, ed., Desclée-Brouwer, Paris*, 3, pp. 341-367.

Bernhard Riemann

  1859 "Ueber die anzahl der primzahlen unter einer gegebenen grösse", *Monatsberichte der Berliner Akademie* (Nov. 1859), pp. 136-144, DOI: 10.1017/cbo9781139568050.008.

Atle Selberg

  1949 "An elementary proof of Dirichlet's theorem about primes in an arithmetic progression", *The Annals of Mathematics*, 50, 2, pp. 297-304, DOI: 10.2307/1969454.

Pafnutï Lvovitch Tschebischeff

  1852a "Mémoire sur les nombres premiers", *Journal de Mathématiques Pures et Appliquées*, 17, 1, pp. 366-390.

  1852b "Sur la totalité des nombres premiers inférieurs à une limite donnée", *Journal de Mathématiques Pures et Appliquées*, 17, 1, pp. 341-365.

BOOKS

Tom Mike Apostol
    1976   *Introduction to analytic number theory*, 1st ed., Undergraduate Texts in Mathematics, Springer New York, NY, DOI: 10.1007/978-1-4757-5579-4.

Paul Gustav Heinrich Bachmann
    1894   *Analytische zahlentheorie*, vol. 2.

Alina Carmen Cojocaru and Maruti Ram Pedaprolu Murty
    2006   *An introduction to sieve methods and their applications*, Cambridge University Press.

Johann Peter Gustav Lejeune Dirichlet
    1879   *Vorlesungen Über Zahlentheorie*, ed. by R. Dedekind, Cambridge University Press.
    1897a   "Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält", in, *G. Lejeune Dirichlet's Werke*, ed. by Leopold Kronecker and László Fuchs, Druck Und Verlag Von Georg Reimer., vol. 1, pp. 313-342, DOI: 10.1017/cbo9781139237321.012.
    1897b   "Über Die Bestimmung Der Mittleren Werthe", in, *G. Lejeune Dirichlet's Werke*, ed. by Leopold Kronecker and László Fuchs, Druck Und Verlag Von Georg Reimer., vol. 2, pp. 49-66, DOI: 10.1017/cbo9781139237345.007.

Leonhard Euler
    1748   *Introductio in Analysin Infinitorum*, vol. 1.

Jhn Friedlander and Henryk Iwaniec
    2010   *Opera de cribro*, Colloquium Publications, American Mathematical Society, vol. 1.

Godfrey Harold Hardy
    1910   *Orders of Infinity: The 'Infinitärcalcül' of Paul Du Bois-Reymond, Cambridge Tracts in Mathematics*, Cambridge University Press.

Godfrey Harold Hardy and Marcel Riesz

    1915   *The general theory of Dirichlet's series*, Cambridge University Press.

Albert Edward Ingham

    1932   *The distribution of prime numbers*, 1st ed., Cambridge University Press.

Camille Jordan

    1870   *Traiteé des substitutions et des équations algébriques*, Gauthier-Villars, Paris.

Arthur Knoebel, Reinhard Laubenbacher, Jerry Lodder, and David Pengelley

    2007   "Sums of numerical powers", in, *Mathematical Masterpieces: Further chronicles by the explorers*, Springer-Verlag, pp. 32-37.

Edmund Landau

    1909   *Handbuch der Lehre von der Verteilung der Primzahlen*, vol. 2.

    1927   *Vorlesungen über Zahlentheorie*, Hirzel, vol. 1.

    1969   *Elementary number theory*, 1st ed., Chelsea, vol. 1.

Adrien Marie Legendre

    1798   *Essai sur la théorie des nombres*, 2nd ed., Duprat, DOI: 10.1017/cbo9780511693199.020.

Lorenzo Mascheroni

    1790   *Adnotationes ad calculum Integralem Euleri*, Galeatii.

Srinivasa Ramanujan

    1927   "Highly Composite Numbers", in, *Collected papers of Srinivasa Ramanujan*, ed. by Godfrey Harold Hardy, Peruvemba Venkatesvara Seshu Aiyar, and Bertram Martin Wilson, Cambridge University Press, pp. 78-128.

INDEX